Fraunhofer
FOKUS

Fraunhofer Institute for Open Communication Systems



ISBN: 978-3-00-038715-9

Title: EuroITV 2012 – Adjunct Proceedings
Editors: Stefan Arbanowski, Stephan Steglich, Hendrik Knoche, Jan Hess
Date: 20120630
Publisher: Fraunhofer Institute for Open Communication Systems, FOKUS

Chairs' Welcome

It is our great pleasure to welcome you to the *2012 European Interactive TV conference – EuroITV'12.* This year's conference marks the 10th anniversary of our community. The conference continues its tradition of being the premier forum for researchers and practitioners in the area of interactive TV and video to share their insights on systems and enabling technologies, human computer interaction as well as media and economic studies.

EuroITV'12 provides researchers and practitioners a unique interdisciplinary venue to share and learn from each others' perspectives. This year, the theme of the conference - *Bridging People, Places and Platforms* - aimed at eliciting contributions that cover the increasingly diverse and heterogeneous infrastructure, services, applications and concepts that make up the experience of interacting with audio-visual content. We hope you will find them interesting and help you identify new directions for future research and development.

The call for papers attracted submissions from Asia, Australia, Europe, and the Americas. The program committee accepted 32 workshop papers, 13 posters, 6 contributions from PHD students (doctoral consortium), 13 demos, 9 contributions for iTV in industry and 15 contributions for the grand challenge. These papers cover a variety of topics, including *cross-platform experiences, leveraging social networks, novel gesture based interaction techniques, accessibility, media consumption, content production and delivery optimization*. In addition, the program includes keynote speeches by Janet Murray (Georgia Tech) on Emerging Storytelling Structures and Olga Khroustaleva & Tom Broxton (YouTube) on supporting the ecosystem of YouTube. We hope that these proceedings will serve as a valuable reference for researchers and practitioners in the area of interactive multimedia consumption.

Putting together *EuroITV'12* was a genuine team effort spanning six continents and many time zones. We first thank the authors for providing the content of the program. We are also grateful to the program committee, who worked very hard to review papers and provide feedback for authors, and to the track chairs for their dedicated work and meta-reviews. Finally, we thank our sponsor Fraunhofer, our supporter DFG (Deutsche Forschungsgemeinschaft, German Research Foundation), our (in-)cooperation partners the ACM SIGs (SIGCHI, SIGWEB, SIGMM) and interactiondesign.org, and our media partners informitv and Beuth University Berlin.

We hope that you will find this program interesting and thought provoking and that the conference will provide you with a valuable opportunity to share ideas with other researchers and practitioners from institutions and companies from around the world.

Stefan Arbanowski & Stephan Steglich
EuroITV'12 General Chairs
Fraunhofer FOKUS, Berlin, Germany

Hendrik Knoche[1] & Jan Hess[2]
EuroITV'12 Program Chairs
[1] EPFL, Switzerland
[2] University of Siegen, Germany

Table of Contents

Keynotes		7
	Transcending Transmedia: Emerging Story Telling Structures for the Emerging Convergence Platforms	8
	Supporting an Ecosystem: From the Biting Baby to the Old Spice Man	9
Demos		10
	iNeighbour TV: A social TV application for senior citizens	11
	GUIDE – Personalized Multimodal TV Interaction	13
	StoryCrate: Tangible Rush Editing for Location Based Production	15
	Connected & Social Shared Smart TV on HbbTV	17
	Interactive Movie Demonstration – Three Rules	19
	Gameinsam – A playful application fostering distributed family interaction on TV	21
	Antiques Interactive	23
	Enabling cross device media experience	25
	Enhancing Media Consumption in the Living Room: Combining Smartphone based applications with the 'Companion Box'	27
	A Continuous Interaction Principle for Interactive Television	29
	Story-Map: iPad Companion for Long Form TV Narratives	31
	WordPress as a Generator of HbbTV Applications	33
	SentiTVChat: Real-time monitoring of Social-TV feedback	35
Doctoral Consortium		37
	Towards a Generic Method for Creating Interactivity in Virtual Studio Sets without Programming	38
	Providing Optimal and Involving Video Game Control	42
	The remediation of television consumption: Public audiences and private places	46
	The Role of Smartphones in Mass Participation TV	50
	Transmedia Narrative in Brazilian Fiction Television	54
	The 'State of Art' at Online Video Issue [notes for the 'Audiovisual Rhetoric 2.0' and its application under development]	58
Posters		62
	Defining the modern viewing environment: Do you know your living room?	63
	A Study of 3D Images and Human Factors: Comparison of 2D and 3D Conceptions by Measuring Brainwaves	67
	Subjective Assessment of a User-controlled Interface for a TV Recommender System	76
	MosaicUI: Interactive media navigation using grid-based video	80
	My P2P Live TV Station	84
	TV for a Billion Interactive Screens: Cloud-based Head-ends for Media Processing and Application Delivery to Mobile Devices	88
	Social TV System That establishes new relationship among sports game audience on TV Receiver	92
	Leveraging the Spatial Information of Viewers for Interactive Television Systems	96
	Creating and sharing personal TV channels in iTV	100
	Interactive TV: Interaction and Control in Second-screen TV Consumption	104

	Elderly viewers identification: designing a decision matrix	108
	Tex-TV meets Twitter – what does it mean to the TV watching experience?	112
	Contextualising the Innovation: How New Technology can Work in the Service of a Television Format	116
iTV in Industry		120
	From meaning to sense	121
	TV Recommender System Field Trial Using Dynamic Collaborative Filtering	123
	Interactive TV leads to customer satisfaction	127
	MyScreens: How audio-visual content will be used across new and traditional media in the future – and how Smart TV can address these trends.	128
	Nine hypotheses for a sustainable TV platform based on HbbTV presented on the base of a on a Best-of example	129
	Lean –back vs. lean –forward: Do television consumers really want to be increasingly in control?	130
	NDS Surfaces	131
	Personalisation of Networked Video	132
	Market Insights from the UK	133
Tutorials		134
	Foundations of interactive multimedia content consumption	135
	Designing Gestural Interfaces for Future Home Entertainment Environment	136
Workshops		138
WS I	**Third International Workshop on Future Television: Making Television more integrated and interactive**	139
	ImTV: Towards an Immersive TV experience	140
	Growing Documentary: Creating a Computer Supported Collaborative Story Telling Environment	152
	Semi-Automatic Video Analysis for Linking Television to the Web	154
	Second-Screen Use in the Home: an Ethnographic Study	162
	Challenges for Multi-Screen TV Apps	174
	Texmix: An automatically generated news navigation portal	186
	Towards the Construction of an Intelligent Platform for Interactive TV	191
	A study of interaction modalities for a TV based interactive multimedia system	197
WS II	**Third Euro ITV Workshop on Interactive Digital TV in Emergent Economies**	202
	Context Sensitive Adaptation of Television Interface	203
	A Distance Education System for Internet Broadcast Using Tutorial Multiplexing	210
	Analyzing Digital TV through the Concepts of Innovation	215
	The Evolution of Television in Brazil: a Literature Review	220
	Using a Voting Classification Algorithm with Neural Networks and Collaborative Filtering in Movie Recommendations for Interactive Digital TV	224
	The Uses and Expectations for Interactive TV in India	229
	Reconfigurable Systems for Digital TV Environment	233
WS III	**Third Workshop on Quality of Experience for Multimedia Content Sharing**	237
	Quantitative Performance Evaluation Of the Emerging HEVC/H.265 Video Codec	238
	Motion saliency for spatial pooling of objective video quality metrics	242

	Profiling User Perception of Free Viewpoint Video Objects in Video Communication	246
	Subjective evaluation of Video DANN Watermarking under bitrate conservation constraints	250
	Subjective experiment dataset for joint development of hybrid video quality measurement algorithms	254
	Instance Selection Techniques for Subjective Quality of Experience Evaluation	258
	Lessons Learned during Real-life QoE Assessment	262
	Measuring the Quality of Long Duration AV Content – Analysis of Test Subject / Time Interval Dependencies	266
WS IV	**UP-TO-US: User-Centric Personalized TV ubiquitOus and secUre Services**	270
	SIP-Based Context-Aware Mobility for IPTV IMS Services	271
	MTLS: A Multiplexing TLS/DTLS based VPN Architecture for Secure Communications in Wired and Wireless Networks	281
	Optimization of Quality of Experience Through File Duplication in Video Sharing Servers	286
	Personalized TV Service through Employing Context-Awareness in IPTV NGN Architecture	292
	Quality of Experience for Audio-Visual Services	299
	A Triplex-layer Based P2P Service Exposure Model in Convergent Environment	306
	TV widgets: Interactive applications to personalize TV's	315
	Automation of learning materials: from web based nomadic scenarios to mobile scenarios	321
	Quality Control, Caching and DNS – Industry Challenges for Global CDNs	323
Grand Challenge		327
	EuroITV Competition Grand Challenge 2010-2012	328
Organizing Committee		331

Keynotes

Transcending Transmedia: Emerging Story Telling Structures for the Emerging Convergence Platforms

Janet Murray, Experimental Television Lab, Georgia Tech, USA



Although the current paradigm for expanded participatory storytellling is the "transmedia" exploitation of the same storyworld on multiple platforms, it can be more productive to think of the digital medium as a single platform, combining all the functionalities we now associate with networked computers, game consoles, and conventionally delivered episodic television. This emerging future medium may have multiple synchronized screens (such as a tablet operating in conjunction with one or more larger displays) and multiple modalities of interaction (such as synchronized distant viewing and gestural input). This talk will focus on what television has traditionally done so well — immerse us in fictional worlds with episodic drama — and ask how the experience of these compelling fictional worlds may change as we move beyond "transmedia" and learn how best to exploit the affordances of an increasingly integrated digital entertainment medium.

Supporting an Ecosystem: From the Biting Baby to the Old Spice Man

Olga Khroustaleva & Tom Broxton, YouTube User Experience





As YouTube evolves we look at how content creators thrive within the ecosystem, what motivates them, and how they get the best results. We also investigate how patterns of media consumption change in an engagement-driven environment, built upon a mix of premium and user-generated content, and fragmented across platforms and devices.

These changes in creator and viewer expectations prompt video advertisers to rethink what effectiveness means in this new space, where a biting baby gets more attention than a high-cost music video. How do lessons learned through years of TV and display advertising and research translate to the paradigm of active social engagement?

What can media producers – advertisers and content owners – do to ensure that their work succeeds in these rapidly changing conditions?

Demos

A social TV application for senior citizens – iNeighbour TV

Jorge Abreu
CETAC.MEDIA
University of Aveiro
3810-193 Aveiro
Portugal
jfa@ua.pt

Pedro Almeida
CETAC.MEDIA
University of Aveiro
3810-193 Aveiro
Portugal
almeida@ua.pt

ABSTRACT
The iNeighbour TV system aims to promote health care and social interaction among senior citizens, their relatives and caregivers. The TV set was the device chosen to mediate all the action, since it is a friendly device and one with which the elderly are used to interact. This system was implemented in an IPTV infrastructure with the collaboration of a Portuguese operator. It includes a set of features that include: medication reminder, monitoring system, caregiver support, events planning, audio calls and a set of tools to promote community service. All these features are seamlessly integrated in TV (in overlay with TV content). The authors consider that this system, already evaluated in a Field Trial, will bring innovation in the support of senior citizens.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Prototyping.

General Terms
Design, Experimentation, Human Factors.

Keywords
Elderly, social iTV, community, IPTV, health care, demo

1. INTRODUCTION
The demographic and social changes seen, especially, in Europe confirm the tendency for the ageing of the population. This has carried the attention of researchers and the industry because technology can act as an important tool to support the ageing problems. In Portugal according to the Portuguese National Institute of Statistics (INE) the number of people over 65 years old (1.901.153) is now higher than children under 14 (1.616.617) [1]. Although these numbers may vary from country to country, this is a problem that governments from developed and some developing countries will have to deal within the coming years. The ageing of the population carries a rising concern as problems such as loneliness and mobility issues may get deeper, especially in non-rural areas where senior citizens are less acquainted with nearby neighbours. For this reason, society must find ways to control the rise of costs and ensure the quality of life of the elderly. In response to this, the European Community proposed an Action Plan "Ageing well in an Information Society" [2] with the following objectives: i) Ageing well at work; ii) Ageing well in the community; iii) Ageing well at home.

In the context of the home, social tools equipped with features regarding the presence and needs of users in similar social circumstances may take an important role to achieve higher levels of comfort, companionship and social interaction between senior users. Some projects are already targeting their efforts in this domain. T-Asisto [3] is a Spanish project that integrates tele-assistance services with television, using digital terrestrial television technology (DTT). This system integrates the common set-top-box (STB) used by DTT with a tele-assistance terminal that receives alerts from various types of sensors scattered in the user's house. There are other projects on the wellness area (funded by the European Commission) [4] that also consider the usage of the TV set as a central device to provide support to elderly users. However, an integrated solution working seamlessly with an existing TV service was yet to be developed.

In this context, we designed and developed an interactive TV application (iNeighbour TV) targeted to senior citizens that integrates seamlessly with TV reception. The development of the application is financially supported by the FCT (Foundation for Science and Technology) and its main objectives are to contribute to the improvement of the quality of life of the elderly and to minimize the impact of an ageing population on developed societies aiming a virtual extension of the neighbourhood concept. It is also intended that the system is able to allow the identification and interaction of individuals based on: i) Common interests; ii) Geographical proximity; iii) Kinship's relations – with its inherent companionship, vigilance and proximal communication benefits.

2. iNeighbour TV FEATURES
iNeighbour TV is mainly an enhanced Social TV application built with Microsoft Mediaroom IPTV technology. It intends to contribute to supporting senior citizens' social relationships and improving their quality of life. Taking advantage of its communication and monitoring systems, it assumes a health care role by providing useful tools for both users and caregivers.
The system, a fully functional application, was longitudinally evaluated on commercial STBs by means of a Field Trial that allowed the team to predict the project's impact in the quality of life of senior citizens.
Following a user centred design approach and an in-depth analysis of the target audience specific needs, the research team implemented a set of features that aim to fulfil the identified requirements. These features allowed the definition of an application organized in six major areas: i) community; ii) health; iii) leisure; iv) information; v) wall and; vi) communication. Each of the first five modules has a set of sub-areas assigned to its own module while communication aggregates a set of features that are always available in the application and can be used in different situations to enhance other existent features.

The **Community area** is focused in seclusion and loneliness issues addressed in the introduction. Its goal is to facilitate the establishment of new relations and strengthen existing ones through social interactions mediated by television. This module is divided in three subareas: i) friends; ii) profile; iii) search.

In the *Friends* subarea the user can see who of his friends is

online; know what they are watching (with the option to jump to the same TV channel); and start an audio call (from his TV set to the TV set of the buddy to whom he wants to speak).



Fig. 1. iNeighbour TV - "Community" area

In the *profile* sub-area, the user settings can be edited including the definition of interests and skills. This profile information is special relevant when other people want to search for unknown iNeighbour TV users that correspond to a common interest or a required skill, e.g. search for other users that also like movies or that have a special talent in gardening (they do this in the *search* sub-area). The system can also use the location of the STBs running the iNeighbour TV application to sort the results by proximity easing the process of finding "Neighbours".

The **Health area** is one of iNeighbour TV's most important modules. This module supports the following complementary subareas: i) medication control; ii) appointments and; iii) management of medical prescriptions.

The sub-area of *medication control* is a key feature of the iNeighbour TV. Health problems usually are felt more intensely in this stage of life, often leading to a medication dependence on an everyday basis. This can be a bigger problem when combined with memory loss issues also reported in this life period. For this reason, the prototype includes a medication reminder system. This feature is centred on a medication agenda; on the automatic trigger of reminders (displayed over the TV image) and; on the delivery of complementary e-mails or SMS in all occasions that the user does not acknowledge the reminders sent to his TV set. The sub-area *appointments* allow the user to check a health schedule and get notifications or alerts about upcoming appointments.

Complementary to the features in the Health area, other attributes transversal to the whole application also address health situations. Senior citizens often require permanent assistance or surveillance from caregivers (either relatives or health professionals). By providing alert and monitoring features, accessible to a caregiver that is indicated and authorized by the elder user, the team believes this feature may reduce the need for constant proximal contact of the caregiver with the elder user. To ensure this support the system is able to keep track of the user's television viewing habits and detect when a significant variation occurs. This deviation combined with occurrences of not taking medications might trigger a warning, alerting the caregiver through TV overlaid messages, e-mail or SMS. The caregiver may also use a mobile APP allowing him to track the state of each person in care.

As mentioned before, the elderly spend a lot of time watching television. Even if it seems a paradox, TV can be used as a tool to promote a healthier life style. The **Leisure area** includes features to encourage seniors to leave the front of the TV or even the house to socialize and to do physical exercise. The user has access to three areas in this module: *Events* - that enables the user, with the aid of a wizard, to create a "rendezvous"and invite participants, to a specific location at a particular date and time; *Calendar* - allowing a quick overview of what the user has to and might do and; *Community (service)* - where the user can search and apply for community service or voluntary work offers that matches his skills and work history.

The **Information area** provides useful information to the daily routine, like *weather* reports or *pharmacies on duty* nearby the user's location. This feature is interrelated with the 'leisure area' being able to, as an example, suggest the creation of a new outdoor event if the weather forecast is favourable.

The **Wall area** is similar to a social network wall that streams status or events updates from the user's contacts.

Finally, the **Communication module** provides support to several features of other modules. It supports TV based text communication and, considering the elders potential problems with vision accuracy [5], SMS reading and creation. This allows users of the iNeighbour TV to read and reply to text messages redirected from their mobile phone. Nevertheless, audio calls are also supported in the system.

3. A SOCIAL TV DEMO

The iNeighbour TV system aims to contribute in a field of expertise that, due to a global ageing population phenomenon, has become an important field of research and development. Although the application was developed using a commercial STB a demo may be provided using the IPTV framework simulator connected to a TV screen and interacting via a standard remote control. This allows experiencing the features of iNeighbour TV with limited technical requirements. Additionally, an interactive video explaining all the areas is also provided.

4. ACKNOWLEDGMENTS

The research leading to these results has received funding from FEDER (through COMPETE) and National Funding (through FCT) under grant agreement no. PTDC/CCI-COM/100824/2008.

5. REFERENCES

[1] Instituto Nacional de Estatistica. População Residente . Retrieved September 5, 2010 from: http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0000611&contexto=pi&selTab=tab0

[2] União Europeia (2008), IP/08/994 Envelhecer Bem, Comissão Europeia liberta 600 milhões de euros para o desenvolvimento de novas soluções digitais destinadas aos idosos europeus. Press Release, June (2008)

[3] Net2u (2010), T-Asisto, Desarrollo de una plataforma de servicios interactivos para la teleasistencia sociala través de televisión digital terrestre. Net2u. Retrieved September 6, 2010 from: http://t-asisto.net2u.es/servicios.html

[4] Boulos, K. et al. (2009). Connectivity for Healthcare and Well-Being Management: Examples from Six European Projects. International Journal of Environmental Research and Public Health. 2009; 6(7): 1947-1971

[5] Carmichael, A. (1999). Style Guide for the Design of Interactive Television Services for Elderly Viewers. Independent Television Commission, Kings Worthy Court, Winchester.

GUIDE - Personalized Multimodal TV Interaction

Carlos Duarte
University of Lisbon
DI, FCUL, Edifício C6
Lisbon, Portugal
carlos.duarte@di.fc.ul.pt

José Coelho
University of Lisbon
DI, FCUL, Edifício C6
Lisbon, Portugal
jcoelho@di.fc.ul.pt

Christoph Jung
Fraunhofer IGD
Fraunhoferstr. 5
Darmstadt, Germany
christoph.jung@igd.fraunhofer.de

ABSTRACT

New interaction devices (e.g. Microsoft Kinect) are making their way into the living room once dominated by TV sets. These devices, by allowing natural modes of interaction (speech and gestures) make user's individual interaction patterns (often resulting from their sensor and motor abilities, but also their past experiences) an important factor in the interaction experience. However, developers of TV based applications still do not have the expertise to take advantage of multimodal interaction and individual abilities. The GUIDE project introduces an open source software framework which is capable to enable applications with multimodal interaction, adapted to the user's characteristics. The adaptation to the user's need is performed based on a user profile, which can be generated by the framework through a sequence of interactive tests. The profiles are user-specific and context-independent, and allow UI adaptation across services. This software layer sits between interaction devices and applications, and reduces changes to established application development procedures. This demonstration presents a working prototype of the GUIDE framework, for Web based applications. Two applications are included: an initialization procedure for creating the user's profile; and a video conferencing application between two GUIDE enabled platforms. The demo highlights the adaptations achieved and resulting benefits for end-users and TV services stakeholders.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Design, Human Factors

Keywords

Multimodal interaction, adaptation, TV, GUIDE

1. INTRODUCTION

During the past years, digital TV as a media consumption platform has increasingly turned from a simple receiver and presenter of broadcast signals to an interactive and personalized media terminal, with access to traditional broadcast as well as internet-based services. At the same time, new interaction devices have made their way into the living room, mainly through game consoles. TV can make use of these to improve its users' experience, both through leveraging multimodal interaction and by personalizing itself (adapting) to its users. However, TV application developers do not possess the expertise required to explore such settings, albeit some recent announcements were made in this direction, but not encompassing all these points in one solution.

The aim of the European project GUIDE[1] is to fill the gaps in expertise identified above. This is realized through a comprehensive approach for the development and dissemination of multimodal user interfaces capable to intelligently adapt to the individual needs and preferences of users.

2. THE GUIDE FRAMEWORK

2.1 Learning about the User

The goals of the GUIDE project are met through the use of multi modalities and adaptation. To benefit from adaptation from the earlier stages of interaction, we designed, as the first step for a new user, the User Initialization Application (UIA). The UIA serves two purposes: (1) introduce to the user the novel interaction possibilities that are available; (2) collect information about preferences and about sensor and motor abilities. The information collected by the UIA is then forwarded to a User Model, where the User Profile is then created (based on knowledge gained with a survey and trials with over 75 users in three different countries [2]). Adaptation performance is thus increased since the beginning. Furthermore, the User Profile is then continuously improved through the monitoring of user's interactions.

2.2 Interaction Devices

Users can interact with TV applications using natural interaction modalities (speech, pointing and gestures) as well as the device they are used to, the remote control (RC). GUIDE's RC is endowed with a gyroscopic sensor, which means it can be used to control an on-screen pointer, in addition to standard functions. Additionally, a tablet interface can also be used to control the on-screen pointer, recognize gestures on its surface or perform selections. Output is based on the TV set. Besides rendering the TV based applications, GUIDE also uses a Virtual Character for output. This human like character is employed to assist the user in several tasks, and takes an active role in situations where inputs have not been recognized or need to be disambiguated. The tablet interface can also be used for output in more than one way: to replicate the output being rendered on the TV set or to complement it, (e.g by providing additional content).

[1] www.guide-project.eu

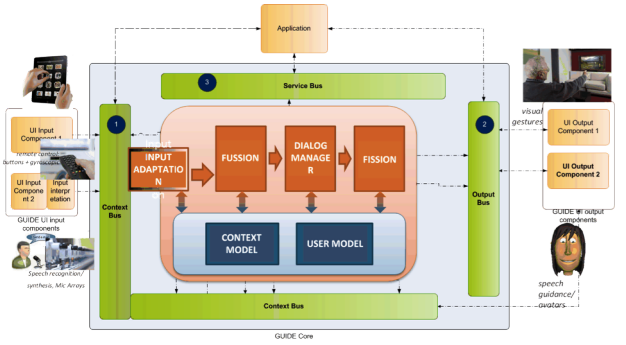


Figure 1: GUIDE's framework architecture.

2.3 Architecture

Figure 1 presents the framework's architecture. This multimodal system includes adaptation mechanisms in several components. All pointing related input channels are processed in input adaptation [1]. Through the use of several algorithms (e.g. gravity wells) it is capable of assisting in pointing tasks by attracting the pointer to targets or removing jitter. Additionally, it estimates the most likely target the user is pointing at. This information is used by multimodal fusion [4], combined with all the other input sources. Fusion is also adaptive, making use of information about the user and the interaction context. This allows to change modality weights to match the user and context properties. Multimodal fission [3] also takes advantage of the user and context characteristics to adapt the rendering of the application's content. Two levels of adaptation exhist: augmentation, where the visual output is augmented with content in other output modalities; adjustment, where the visual output is adapted to the user's abilities and preferences and to the interaction context.

2.4 Interfacing with Applications

This framework is independent of the application's execution environment (e.g. a Web browser for Web applications). A set of API's have been defined to allow exchanging information between framework, input and output devices, and applications. In what concerns applications, the interface should be capable to translate application's states to UIML, the standard used to represent interfaces inside the GUIDE framework. Currently, an interface as already been implemented for Web based applications: the Web browser interface (WBI). The WBI can operate as a browser plug-in, and, among other features, makes available a JavaScript API capable of: (1) parsing a (dynamic) page's HTML and CSS code, converting it to UIML; (2) receiving the adaptations that should be made to the Web page and perform the corresponding changes at the DOM level. Supported by this mechanism, an application developer does not need to change anything in the development process to be able to benefit from the use of the GUIDE framework. To assist the interface adaptation process, the Web application developer can augment applications through the use of a set of WAI-ARIA tags, which impart semantic meaning to the application's code.

3. THE DEMO

This demonstration comprises the above mentioned GUIDE framework, applied to TV based applications running on a Web browser. The initial user profiling process is highlighted, with the execution of the UIA and the subsequent generation of the user profile (the user profile can be consulted at any time, being shown in a standard proposal format being currently prepared in the scope of the VUMS[2] cluster). After completion of the initialization procedure a video-conferencing application is available for interaction. The application has been developed to demonstrate UI adaptations in a meaningful service environment. The demo illustrates: (1) multimodal fusion of user input data (speech, remote control, gestures); (2) GUI adaptations; (3) free-hand cursor data filtering.

4. CONCLUSION

With this demo we show how the GUIDE open source framework is capable to manage a complex multimodal user interface, decoupling it from application logic. This allows for service personalization to become more efficient and convenient for interested stakeholders (end-users and industrial), through the generation of user profiles and subsequent automatic configuration of the UI.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°248893.

6. ADDITIONAL AUTHORS

Pascal Hamisu (Fraunhofer IGD, email: `pascal.hamisu@igd.fraunhofer.de`), Pat Langdon (University of Cambridge, email: `pml24@eng.cam.ac.uk`), Pradipta Biswas (University of Cambdrige, email: `pb400@eng.cam.ac.uk`), Gregor Heinrich (vSonix, email: `gregor.heinrich@vsonix.com`, Luís Almeida (Centro de Computação Gráfica, email: `Luis.Almeida@ccg.pt`, Daniel Costa (University of Lisbon, email: `thewisher@lasige.di.fc.ul.pt`) and Pedro Feiteira (University of Lisbon, email: `pfeiteira@di.fc.ul.pt`).

7. REFERENCES

[1] P. Biswas, P. Robinson, and P. Langdon. Designing inclusive interfaces through user modeling and simulation. *International Journal of Human-Computer Interaction*, 28(1):1–33, 2012.

[2] J. Coelho, C. Duarte, P. Biswas, and P. Langdon. Developing accessible tv applications. In *ASSETS'11*, pages 131–138. ACM, 2011.

[3] D. Costa and C. Duarte. Adapting multimodal fission to user's abilities. In *UAHCI'11*, pages 347–356. Springer-Verlag, 2011.

[4] P. Feiteira and C. Duarte. An evaluation framework for assessing and optimizing multimodal fusion engines performance. In *IIHCI'12*, 2012.

[2] www.veritas-project.eu/vums/

StoryCrate: Tangible Rush Editing for Location Based Production

Tom Bartindale
Newcastle University
tom@bartindale.com

Alia Sheik
BBC Research & Development
Alia.Sheikh@bbc.co.uk

Patrick Olivier
Newcastle University
p.l.olivier@ncl.ac.uk

ABSTRACT
TV and film is an industry dedicated to creative practice, involving many different skilled practitioners from a variety of disciplines working together to produce a quality product. Production houses are now under more pressure than ever to produce this content with smaller budgets and fewer crew members. Various technologies are in development for digitizing the end-to-end workflow, but little work has been done at the initial stages of production, primarily the highly creative area of location based filming. We present a tangible, tabletop storyboarding tool for on-location shooting which provides a variety of tools for meta-data collection, rush editing and clip playback, which we used to investigate digital workflow facilitation in a situated production environment. We deployed this prototype in a real-world small production scenario in collaboration with the BBC and Newcastle University.

Categories and Subject Descriptors
H5.3. [Information interfaces and presentation]: Group and Organization Interfaces

General Terms
Design, Experimentation, Human Factors.

Keywords
Broadcast, tangible interface, collaborative work, inter-disciplinary, prototype, storyboarding, editing

1. INTRODUCTION
This project aimed to bridge the gap between in-house BBC products that were being developed for digitizing the production process, and new interaction techniques and technologies. Our overarching research goal was to integrate the media production process and its multiple technological components, whilst driving existing production staff to produce better content. To enable our understanding of this domain we developed a technology prototype designed to facilitate collaborative production in a broadcast scenario, which we deployed and evaluated during a professional film shoot.

Storyboarding is a creative skill that most practitioners know about, but few have used in practice, especially for live video production. As we wished to provide both an awareness indicator for the current shoot state, but also a method of creating a representation of the end product that the viewer would experience, bringing back the traditional storyboard was the obvious choice. Storyboards are also easily represented on physically large devices, and are recognizable from a distance (i.e. across as busy set)



Figure 1. StoryCrate Prototype.

Our prototype implementation has many discrete elements of functionality, but they all revolve around creating better quality content as a shoot outcome. By giving more ownership and awareness of produced content to the crew who have been hired for their experience and skill, we maximize a production's creative assets, whilst bringing creative decision forward through the production process, as in Figure 2.



Figure 2. The broadcast production workflow.

In order to guide our design process and also design and implement a user study, we chose to categorize creativity in terms of practical activities which could both be facilitated and observed.

- Exploring Alternatives
- Changing Roles
- Linking to the Unexpected
- Externalization of Actions
- Group Communication
- Random Access

2. STORYCRATE HARDWARE
After an extensive design process, based around iterative prototyping and hardware implementation, the prototype system was ready for deployment.

StoryCrate is an interactive table consisting of a computer with two rear-projected displays behind a horizontal surface creating a 60" x 25" high-resolution display, with two LCD monitors mounted



vertically behind it. Shaped plastic tiles used to control the system are optically tracked by two PlayStation 3 Eye Cameras in infrared through the horizontal surface using fiducial markers and the

Figure 3. On Location with StoryCrate.

reacTIVision [3] tracking engine. The entire device and all associated hardware is housed in a 1.5m long flight case on castors, with power and Ethernet inputs on the exterior, and was built to be robust and easily transportable to shoot locations. StoryCrate is written in Microsoft .NET 4, utilizing animation and media playback features built into Windows Presentation Foundation.

On StoryCrate, this is represented as a linear timeline, where each media item is represented as a thumbnail of the media on the display. Almost the entire display is filled with this shared representation of filming state, providing users with a single focal point for keeping track of group progress. During the shoot, a take, or clip immediately appears on the device shortly after it is filmed, and can be manipulated and previewed on StoryCrate as a thumbnail. The interface is based on a multi-track video editor, with time horizontally, and multiple takes of a shot vertically. All interaction with the device is performed by manipulating plastic tiles or 'tangibles' on the surface of the display, rather than touch or mouse input. This integral feature in the design, combined with limiting the number of tangibles, enforces and encourages communication between con-current users by driving verbal communication about task intention.

StoryCrate provides discrete functional elements for the following tasks, where each task is independent of another, allowing for complete flexibility in how users choose to operate it:

- adding textual metadata to clips;
- playback of both the timeline and individual clips;
- selecting in- and out-points on clips;
- moving, deleting, inserting and editing clips on the timeline;
- adding new hand-drawn storyboard content.

3. ON LOCATION

Rogers "Why it's worth the hassle", comments that Ubiquitous computing is difficult to evaluate due to context of use, and that traditional lab studies fail to capture the complexities and richness of a domain [2]. Consequently, beyond obvious usability issues, lab-based studies are unlikely to be able to predict how a real production team will use, and adapt to, the new technology when it is deployed "in the wild". Consequently, we chose to deploy StoryCrate on a live film shoot to evaluate how a real crew would use specific aspects of its functionality, and see the impact it had on their workflow.

Deploying a prototype for real world use involves creating a robust system, both in terms of the software and its mechanical properties. Although high fidelity prototyping has been shown to be an effective approach it is not as widely deployed in interaction design as, say, agile programming for systems design. StoryCrate"s discrete elements of functionality were based on activities performed as part of a traditional workflow. However, three use cases reflect our expectations about how StoryCrate could potentially improve creativity during a shoot, and in our study we paid particular attention to observing whether aspects of these emerged.

- Clip Review
- Context Explanation
- Logging

Although on-going, our initial analysis from this deployment demonstrates that although presented with a multi-user, collaborative interface, crew members chose to delegate a member of the team to maintain the interface, subsequently using the display to explain directors vision at key intervals.

4. PROPOSED DEMO INTERACTION

Our demonstration is based loosely around an on-location film shoot at the conference. Users can easily walk up and interact with the physicality and tangibles within the system, and will be drawn into viewing existing footage that has been shot previously, but are also encouraged to shoot their own "mini interview". This interview will be pre-storyboarded with three simple shots, which they are encouraged to fill in by filming, editing and marking up the shots themselves or in collaboration with us or other users. Throughout this process we will give a running commentary on the technology, design choices and results of the system in real world scenarios.

Users can experience the entire range of interaction, just as if they were on a real film shoot, being able to:

- Log Meta data for each shot in the system.
- Edit clips together and make editorial decisions.
- Generate custom storyboards, visual cues and notes.
- Review footage from previous shoots, including their own.
- Create a rush edit of their footage and play it back.

We envisage a dwell time of between 1 and 5 minutes per person, with every user not investigating all functions of the device.

Extensive analysis of the design process, crew learning stages and final deployment have been produced, and we would love to have the opportunity to discuss these findings with visiting users, using the demonstration as a unique way of describing the results.

5. REFERENCES

1. Hornecker, E. Understanding the benefits of graspable interfaces for cooperative use. *Cooperative Systems Design*, 71 (2002).
2. Rogers, Y., Connelly, K., Tedesco, L., et al. Why it's worth the hassle: the value of in-situ studies when designing In *Proc. Ubicomp*. (2007), 336-353.
3. Kaltenbrunner, M. and Bencina, R. reacTIVision: a computer-vision framework for table-based tangible interaction. In *Proc. TEI*, (2007), 69 - 74.

Demo: Connected & Social Shared Smart TV on HbbTV

Robert Strzebkowski
Beuth University of Technology Berlin
Luxemburger Str. 10
13353 Berlin
+49-30-4504-5212
robertst@beuth-hochschule.de

Roman Bartoli
Beuth University of Technology Berlin
Luxemburger Str. 10
13353 Berlin
+49-30-4504-2282
roman.bartoli@gmail.com

Sven Spielvogel
Beuth University of Technology Berlin
Luxemburger Str. 10
13353 Berlin
+49-30-4504-2282
spielvogel@beuth-hochschule.de

ABSTRACT
In this demo we will present connected Smart TV scenario with the synchronous usage of mobile second screens – two Tablets - based on the HbbTV standard. Content and application splitting as well as triggering are the main features. The user are able to choose, which content will be present on primary or on secondary screen. The HbbTV 'Red Button' application is triggering on broadcast segments synchronized additional information as well as interactive app on their certain states. Concurrent collaborative activity - joint painting – will be presented as well as the self built DVB/HbbTV Playout, chain and their functionality.

Categories and Subject Descriptors
B.4.1 [**Input/Output And Data Communication**]: Data Communications Devices - *Transmitters*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems - *Distributed applications* D.3.0 [**Programming Languages**]: General – *Standards* -, Language Classifications - *Object-oriented languages, Design lamguages, Extensible languages;* J.7 [**Computer Applications**]: Arts and Humanities – *Fine arts* -, Computers in other Systems – *Consumer products*.

General Terms
Algorithms, Design, Experimentation, Human Factors, Standardization, Languages, Verification.

Keywords
HbbTV, Smart TV, Connected TV, Distributed TV, Interactive TV, Tablet Device, HTML 5, Content Splitting

1. INTRODUCTION
There is an evident growth of the Smart TV / Hybrid TV infrastructure [1], where TV devices are connected to the Internet and are able to perform online interactive TV applications. At the time there are in Europe two major ways to provide and get interactive TV applications in the scope of Smart TV: on the one side with the App-based approach - realized with slightly proprietary programming and frameworks and not synchronized with the running TV program. The well known representative of this approach are NetTV by Philips and Internet@TV by Samsung. On the other side there is the quite new pan-European initiative and standard for interactive and hybrid TV applications, the 'Hybrid Broadcast Broadband TV – HbbTV' [www.hbbtv.org]. In this approach the main idea is to establish a connection between the running broadcast/TV program – either as a live broadcast or as a TV / Video-On-Demand - and a content-related interactive HbbTV application with additionally content to the TV program provided online. The main TV devices manufacturers fortunately implement increasingly both technical systems in one TV device for the EU market.

Independent from the technically approach there are new challenges for TV program editors and producers, for TV-App developers as well as for the TV user to concept, produce, to engage and to use the interactive TV programs and applications. Beside of the challenges there are also new usability and technically problems.

Regarding the usability we have to state that the input device TV remote control – in quite the same functionality known for about forty years – is an insufficient device to navigate through interactive TV applications. Therefore has build companies like BskyB special prepared remote control for their interactive services to foster the 'Viewser' by using those services. There are new developments, which implement for example the accelerometer technique like the Wii control or the gesture recognition like Microsoft's Kinect. The problem is still not solved.

The other problem is the presentation of interactive applications and additional content on the same screen during watching a TV program/broadcast in company with one or some more buddies. Just watching the EPG could lead sometimes to a small familiar disaster.

The vast growing of mobile devices with multimedia capabilities is offering an interesting support and solution for the mentioned problems. Especially Tablet devices provide a powerful navigation, interaction and presentation possibilities. Studies show, that the owner of Tablet and Smartphones devices are using them in average up to 70% during the watching of TV [2]. There are already some proprietary TV eco systems like Apple, Philips, Samsung and Sony which have started to deploy different mobile devices including Smartphones all above for navigation purposes the EPG, the media databases or App bibliotheca but also to stream media on and to the different devices. There is also a increasing number of independent developer and provider of so called 'second screen' applications with the focus on social communication and just-in-time advertising like 'Couchfunk', 'TV Dinner', 'Viggle', 'WiO', or 'SecondScreen Networks'. There is A 'second screen' experiment at the BBC 2011 has shown an interesting potential and enormous interest to provide and use additional content to a broadcast on a second screen device. 92% of the 300 'viewser' reported an understanding benefit and 70% of them were encouraged to follow the Internet presence of the certain BBC program. [3]

2. HbbTV and the second screen scenarios

In the context of our research and development project 'Connected HbbTV' we examine the technically as well as possibilities concerning interactivity and usability issues of connected Hybrid / Smart TV with mobile devices based on the new pan-European open standard for interactive Television - HbbTV.

The 'second screen' approach within the HbbTV framework is quite new and by far until now not really wide explored. As official supporter [4] of this standard and the initiative we are very interested in the testing of the current technological capabilities and boundaries as well as in the further development of HbbTV. Hence we experiment e. g. with some HTML5 elements within the CE-HTML framework, to explore the future potentials for the next development steps of HbbTV (2.0).

HbbTV has been continuing developing. There are still many technologically challenges like use of mobile devices, content shifting and presentation/streaming between the primary and secondary screen, explicit/user triggered user tracking and bookmarking during the channel switching which will be explored and solved in the next time. Some of them we explore in our demo scenario.

3. The Demo Case – the Application Scenario

In the demo we are going to present and explore following interactivity cases and technologies:

- Signaling of Red Button HbbTV application also on mobile devices based on push notification services
- Triggering the application state/mode on the mobile device (second screen) trough the HbbTV signal and HbbTV application - manually as well as automatically
- Content splitting and shifting bidirectional between first and second screen – manually as well as automatically
- 'Application splitting' between the first and the second screen
- Fast transforming of HbbTV – CE-HTML - applications to android tablet devices with gesture navigation
- Synchronizing between certain film scenes and the additional HbbTV content with stream events
- Simultaneously and collaborative working of two users on two tablet devices on a common picture with bitmap merging
- User recognition and bookmark logging during the channel switching
- Use of HTML5 elements within the framework of CE-HTML to explore the 'next generation' HbbTV
- Self-built and inexpensive DVB/HbbTV Playout chain

4. The Demo Case – the Technology behind

Our research project is based on a self-assembled DVB/HbbTV Playout Chain based on the Open Source Software OpenCaster for generating and manipulating the MPEG-2 Transport streams, the DVB-S2 modulation card from Dektec and an appropriate HbbTV compatible TV device.

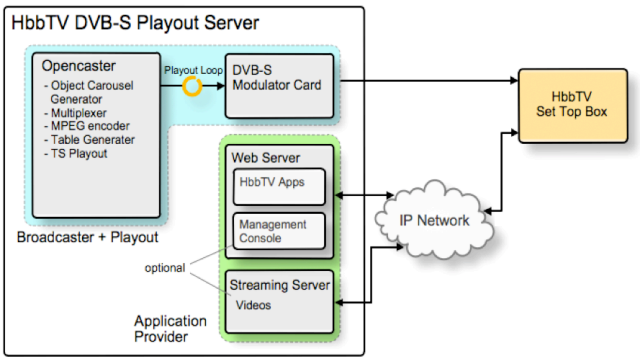


Figure 1. HbbTV DVB-S Playout Server and Chain.

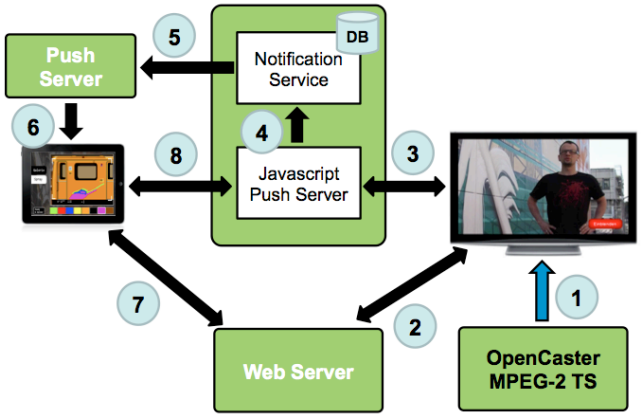


Figure 2. HbbTV signaling on the mobile device

The figure 1 above is showing the DVB/HbbTV Playout System and figure 2 the messages sequence to signaling the HbbTV event on the mobile device.

5. REFERENCES

[1] Deloitte Consulting 2011. *Smart-TV Geschäftsmodelle Internationale Perspektive*. http://www.bitkom.org/de/veranstaltungen/66455_67365.aspx, (24.02.2012)

[2] NielsenWire, 2011. *In the U.S., Tablets are TV Buddies while eReaders Make Great Bedfellows*, http://blog.nielsen.com/nielsenwire/?p=27702 (01.03.2012)

[3] Jones, T. 2011. *Designing for second screens: The Autumnwatch Companion*, http://www.bbc.co.uk/blogs/researchanddevelopment/2011/04/the-autumnwatch-companion---de.shtml (12.02.2012)

[4] http://hbbtv.org/pages/hbbtv_consortium/supporters.php

EuroITV 2012 Submission: Interactive Movie Demonstration Die Hobrechts, Drei Regeln (Three Rules)

Christoph Brosius
Managing Director
Hobrechtstraße 65
12047 Berlin, Germany
(+49) 030 / 629 012 32
christoph@diehobrechts.de

Daniel Helbig
Game Designer
Hobrechtstraße 65
12047 Berlin, Germany
(+49) 030 / 629 012 32
daniel@diehobrechts.de

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Prototyping.

General Terms
Design, Experimentation, Human Factors.

Keywords
Interactive Movie

1. THE COMPANY

Die Hobrechts is a Berlin-based agency for Game Design and Game Thinking founded in 2011. We support our customers with concepts and designs for entertainment and educational products. We also share our knowledge in professional trainings at private and public schools.

Our core competencies are Game and Interface Design, Web and 3D Programming as well as agile project management. The team has a broad experience in game development, reaching from Online, Retail, Console and Mobile Games to Serious and Alternate Reality Games.

2. THE PRODUCT

Drei Regeln (Three Rules) is a prototype for an interactive short movie, written, designed and produced by Die Hobrechts. In the movie the player can alter the storyline, change between points of views and influence the actions of the protagonists without any delay, idle time or interruption.

The goal is to provide a new technique for the narrative of interactive film and the future of IP-TV. The interactions are displayed in an easy to use split screen interface including a timer. As a result the film never stops or waits for input.

Originally shot in German, the movie is also available with English subtitles and interface.

2.1 Facts

Platforms	TV, Browser, MacOS, Windows, Android
Tech	Flash
Genre	Interactive short movie
Running Time	14-17 min depending on choices
Original Voice	German
Subtitles	English

2.2 Features

In certain moments of the movie the player is offered a limited timeframe to make a decision, which will alter how the movie will unfold. If the player doesn't decide, the system will do it for him, to ensure that the dramatic and pacing of the movie is never interrupted and also to allow the whole movie to be watched without using the interactive aspects of the product. The player's decisions can lead to one of three results:

Influence the story: Affect the actions of the characters. The consequences have an effect on the whole progress of the movie.



Figure 1, Influence the plot

Change the perspective: See an alternative view of the plot or its characters without changing the story.



Figure 2, Change Perspective

See additional material: See content you would not have seen otherwise and extend the total playtime (e.g. flash backs).



Figure 3, Additional Material

The web version offers a result screen which shows the players choices and the ending seen. He can compare results with other players and share his experience via facebook.



Figure 4, Result Screen

2.3 Platforms

Drei Regeln runs in a usual Browser environment on PC and Mac, Android Smartphones and TV screens in fullscreen mode.



Figure 5, TV/Browser/Android

2.4 Scope

There are four different choices to make, leading to a total of four different endings. The following flowchart displays the structure.

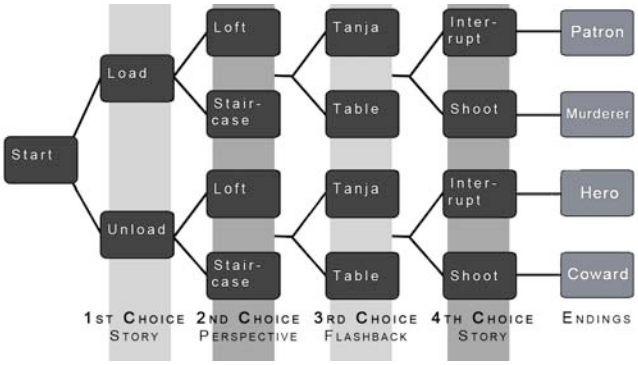


Figure 6, Flowchart

2.5 Story Synopsis

Chris and Max are on their way to make the first drug deal of their life. A deal that will change their lives forever, at least they hope so. The meeting location is set and the rules are clear, but knowing the rules and following them are two separate things, as both of them will have to learn the hard way.

Gameinsam - A playful application fostering distributed family interaction on TV

Katja Herrmanny Steffen Budweg Matthias Klauser Anna Kötteritzsch
University of Duisburg-Essen
Interactive Systems & Interaction Design
47057 Duisburg
+49-203-3792276
{katja.herrmanny, steffen.budweg, matthias.klauser, anna.koetteritzsch}@uni-due.de

ABSTRACT
In this paper, we describe the concept and the prototype implementation of an interactive TV application which aims to enrich everyday TV-watching with playful interaction components, in particular to meet the demand of elderly people to keep in touch with their peers and family members, even if living apart.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information System**]

K.8 [**Personal Computing: Games**]

J.4 [**Social and Behavioral Sciences: Psychology**]

General Terms
Design, Human Factors, Theory

Keywords
Interactive television, iTV, AAL, elderly, family, game, playful interaction, Social TV, Second Screen, FoSIBLE

1. INTRODUCTION

1.1 The Social Context of TV
Watching TV in the context of a multi-person household is a social situation, which has been shown in many investigations [6][9]. Sociable TV-watching therefore is a connecting point for the family members and can act as a "ticket to talk" [7]. In this context there are two basic roles TV takes [6]: 1. an internal social function (the situation of watching TV with the family) 2. an external social function (e.g. TV programs as topics of conversation).

To foster this external social function Sokoler and Sanchez Svensson suggest a presence mechanism called "PresenceRemote" while "keeping the original TV watching activity as intact as possible" [8].

Following Hochmuth [4], Briggs describes another phenomenon in the context with of popular TV quiz shows such as "Who wants to be a millionaire?". He defines „shoutability" as the need to make a guess and verbalize it when a question is asked during the show. This effect does not depend on the age or on the social state of the audience.

In the following chapters we describe how the social situation of watching TV and the phenomenon of shoutability are integrated in the design of a playful application concept and prototype to foster distributed family interaction on TV.

1.2 Family Context
Multi-person households - at least with more than two generations - are no more common today. Moreover, there is an increase of isolation of elderly people due to family breakdowns and as a result a feeling of loneliness [3]. On the other hand a strong fixation on the family by elderly people can be observed [2]. Ducheneaut et al. [1] claim that "sociability is becoming more and more distributed in this context as technology enables diverse remote interactions". This shows the urgent demand of solutions creating social situations for elderly and their families, which are easy to integrate in the daily life of all participants.

1.3 Elderly and Technology
As many interactive technologies are hard to handle or present barriers to getting in touch with those technologies for elderly [6], our approach builds on using existing well-known technologies, such as the personal TV with a standard remote control.

2. CONCEPT
To re-integrate elderly people in family life and to support connectedness, we have created an application to foster interaction with peer groups and family members. It is based on the idea that TV is a social connection point of the family.

We picked up the typical social aspects of watching TV together and the game-like use of TV broadcasts. Our starting point is the TV program as a part of the playful interaction. The application offers the opportunity of "shared shoutability", allowing each user to watch the program at home and share his or her guesses about the TV program (e.g. answers in a quiz show) using the standard remote control. He or she also sees which answers the other participants chose. Family members together achieve joint high scores offering a collaborative playful interaction.

Our concept is realizable with many TV genres, such as quiz shows, entertainment shows (e.g. "Who will win the game?"), casting shows (e.g. "Will the candidate be in the recall?") crime thriller (e.g. "Who is the murderer?", "Is the alibi true?"), or sports (e.g. "Who will win with how many points").

2.1 Application Scenario and Implementation
We implemented "Gameinsam" as a widget for the Samsung Smart TV (internet-enabled HbbTV) by using the Samsung Smart TV SDK 2.5.1. For the demonstration, we integrated a pre-recorded video of "Wer wird Millionär?" (the German version of "Who wants to be a millionaire?") instead of a live TV signal in the prototype.

After starting the "Gameinsam" app the user is presented the integrated interface consisting of the TV program (e.g. "Who wants to get a millionaire?") and the interaction opportunities. These include a buddy list (with all other participants who are logged in and watching the same program at that moment), others users answers (or a question mark if no answer is logged in yet), the user's own answers, a feedback element to remind the user to make an input, if the question is active and there is no input yet, and the family score (see Figure 1). Answers can be given by using the standard remote control. When playing "Who wants to get a millionaire?", there are always up to four available answers which can be chosen from with the color-coded remote control buttons. An answer can be corrected several times, as long as the question is active. When the solution is given in the program, the question is set inactive, the correct answers are colored green, the others red and the family score is updated. When there is a new question in the TV program, the interaction is set on active again, so that the users can make their input. All the answers will be considered in the family score, which is given in an overall percentage. Every participant can join or leave "Gameinsam" at any time during the program. After the end of the program the users see an end screen with the former high score, the current score and the new high score.

In summary, the application aims at providing high flexibility to meet different use contexts.

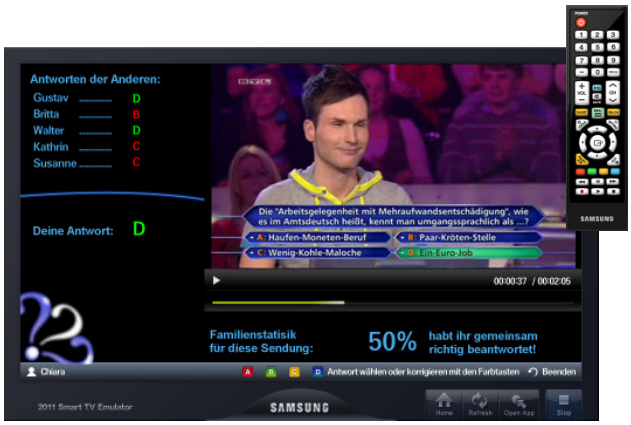


Figure 1. Interface of "Gameinsam".

3. Outlook & Conclusion

From our point of view the strength of our approach to playful Social TV does not only lie in its social components, but also in its synchronicity. Although asynchronous TV solutions exists, TV is still commonly used synchronously. This can be used to establish an interaction concept which is both: close to real interaction situations and easy to integrate in people's everyday life.

The above-mentioned internal and external functions of television are both included in "Gameinsam": We allow direct synchronous interaction and establish a common ground for later conversations.

Although our application contains playful components, it reaches beyond stand-alone game-approaches by targeting the common situation of watching TV together - at the same time, but in different places - to foster social interaction in today's often distributed families and households.

4. ACKNOWLEDGMENTS

Parts of the work presented here have been conducted within the FoSIBLE project which is funded by the European Ambient Assisted Living (AAL) Joint Program together with BMBF, ANR and FFG. The authors especially thank Maike Schäfer and Sascha Vogt for their support during the development of "Gameinsam".

5. REFERENCES

[1] Ducheneaut, N., Moore, R.J., Oehlberg, L., Thornton, J.D., and Nickell, E. 2008. Social TV: Designing for distributed, sociable television viewing. In *International Journal of Human-Computer Interaction*, *24* (2), 136-154.

[2] Harriehausen, C. 2009. *Wohnen im Alter Diagnose: Soziale Isolation*. Retrieved November 05, 2011, from: http://www.faz.net/aktuell/wirtschaft/immobilien/wohnen/wohnen-im-alter-diagnose-soziale-isolation-1768547.html

[3] Help The Aged. 2008. Isolation and loneliness. Retrieved March 28, 2012 from http://www.ageuk.org.uk/documents/en-gb/for-professionals/communities-and-inclusion/id6919_isolation_and_loneliness_2008_pro.pdf?dtrk=true

[4] Hochmuth, Teresa. *'Weltfernsehen' - die internationale Vermarktung der Quizshow 'Wer wird Millionaïr?'*. GRIN Verlag GmbH, München, 2008.

[5] Kriglstein, S., Wallner, G. 2005. HOMIE: An Artificial Companion for Elderly People, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, OR, USA, April 2-7, 2005). CHI'05, 2094-2098.

[6] Morrison, M. 2001. A look at mass and computer mediated technologies: understanding the roles of television and computers in the home. *Journal of Broadcasting and Electronic Media, 45*(1), 135- 161.

[7] Sacks, H. *Lectures on Conversation: Volumes I and II.* Blackwell, Oxford, 1992.

[8] Sokoler, T. and Sanchez Svensson, M.. 2008. PresenceRemote: Embracing Ambiguity in the Design of Social TV for Senior Citizens. In *Proceedings of the 6th European conference on Changing Television Environments* (EUROITV '08), Manfred Tscheligi, Marianna Obrist, and Artur Lugmayr (Eds.). Springer-Verlag, Berlin, Heidelberg, 158-162. DOI=10.1007/978-3-540-69478-6_20 http://dx.doi.org/10.1007/978-3-540-69478-6_20

[9] White, E. S. 1986. Interpersonal bias in television and interactive media. In *Inter/Media: Interpersonal Communication in a Media World*, G. Gumpert, and R. Cathcart, Eds. Oxford University Press. Oxford, 110-120

Antiques Interactive

Lotte Belice Baltussen
Netherlands Institute for Sound and Vision
lbbaltussen@beeldengeluid.nl

Mieke H.R. Leyssen
CWI
Mieke.Leyssen@cwi.nl

Jacco van Ossenbruggen
CWI
Jacco.van.Ossenbruggen@cwi.nl

Johan Oomen
Netherlands Institute for Sound and Vision
joomen@beeldengeluid.nl

Jaap Blom
Netherlands Institute for Sound and Vision
jblom@beeldengeluid.nl

Pieter van Leeuwen
Noterik BV
p.vanleeuwen@noterik.nl

Lynda Hardman
CWI
Lynda.Hardman@cwi.nl

ABSTRACT

We demonstrate the potential of automatically linking content from television broadcasts in the context of enriching the experience of users watching the broadcast. The demo focusses on (1) providing smooth user interface that allows users to look up web content and other audiovisual material that is directly related to the television content and (2) providing means for social interaction.

Categories and Subject Descriptors

H.5.2 [Information Interfaces And Presentation]: User Interfaces---*Interaction styles, Graphical user interfaces*; H.1.2 [Information Systems]: User/Machine Systems---*Human factors*; H.5.1 [Information Interfaces And Presentation]: Multimedia Information Systems---*Video*

General Terms

Design, Human Factors, Standardization.

Keywords

Connected TV, semantic multimedia, media annotation, interactive television, user interfaces, content analysis.

1. INTRODUCTION

The vision of the European project Television Linked To The Web (LinkedTV[1]) is to automatically enrich content to provide users with a richer interactive experience while watching television. The goal of this application is to demonstrate how television interfaces might look like in the future. We focus on the key user interface challenges that result from rich hyperlinked content: the need to unobtrusively present the interactive elements and to combine navigation with the play out of the main audiovisual stream. We will sketch the functionality of the application using a scenario based on a recent episode of the Dutch version of the well-known BBC programme Antiques Roadshow. The original episode is available online.[2]

2. DEMO SCENARIO

In this section, we describe the scenario on which the demo is based.

2.1 Introducing Rita – the persona

Rita is an administrative assistant at the Art History department of the University of Amsterdam. She didn't study art herself, but spends a lot of her free time on museum visits, creative courses and reading about art. One of her favourite programmes is the Antiques Roadshow (Dutch title: Tussen Kunst & Kitsch) from the Dutch public broadcaster AVRO[3]. Rita likes to watch the Antiques Roadshow because, on the one hand, she learns more about art history, and, on the other hand, because she thinks it's fun to guess how much the objects people bring in are worth. She's also interested in the locations where the programme is recorded, as this usually takes place in a historically interesting location, such as a museum or a cultural institute.

2.2 Rita watches the Antiques Roadshow

Rita is watching the latest episode of the Roadshow. The show's host, Nelleke van der Krogt, gives an introduction to the programme. Rita sees the show has been recorded in the Hermitage Museum in Amsterdam. She always wanted to visit the museum as well as finding out what the link is between the Amsterdam Hermitage and the Hermitage in St. Petersburg. She

[1] www.linkedtv.eu

[2] http://cultuurgids.avro.nl/front/detailtkk.html?item=8237850

[3] http://avro.nl/

sees a shot of the outside of the museum and notices that it was originally a home for old women from the 17th century. Intriguing! Rita wants to know more about the Hermitage location's history and see images of how the building used to look. After expressing her need for more information, a bar appears on her screen with additional background material about the museum and the building in which it is located. While Rita is browsing, the programme continues in a smaller part of her screen.

After the show introduced the Hermitage, a bit of its history and current and future exhibitions, the objects brought in by the participants are evaluated by the experts. One person has brought in a golden, filigree box from France in which people stored a sponge with vinegar they could sniff to stay awake during long church sermons. Inside the box, the Chi Ro symbol has been incorporated. Rita has heard of it, but doesn't really know much about its provenance and history. Again, Rita uses the remote to access information about the Chi Ro symbol on Wikipedia and to explore a similar object, a golden box with the same symbol, found on the Europeana portal. Since she doesn't want to miss the expert's opinion, Rita pauses the programme only to resume it after exploring the Europeana content.

The final person on the show (a woman in her 70s) has brought in a painting that has the signature 'Jan Sluijters'. This is in fact a famous Dutch painter, so she wants to make sure that it is indeed his. The expert - Willem de Winter - confirms that it is genuine. He states that the painting depicts a street scene in Paris, and that it was made in 1906. Rita thinks the painting is beautiful, and wants to learn more about Sluijters and his work. She learns that he experimented with various styles that were typical for the era: including fauvism, cubism and expressionism. She'd like to see a general overview of the differences of these styles and the leaders of the respective movements.

During the show Rita could mark interesting fragments by pressing the "tag" button on her remote control. While tagging she continued watching the show but afterwards these marked fragments are used to generate a personalized extended information show based on the topics Rita has marked as interesting. She can watch this related / extended content directly after the show on her television or decide to have this playlist saved so she can view it later. This is not only limited to her television but could also be a desktop, second screen or smartphone, as long as these are linked together. She's able to share this information on social networks, allowing her friends to see highlights related to the episode.

3. TECHNICAL DETAILS

The demo application shows the potential of automatically enriching television content for an enriched end-user experience. The chosen scenario allows a wide variety of enrichment techniques to be deployed: techniques to link AV content to Wikipedia articles, named entity recognition and linking of person, location and art style names, feature detection techniques to link close ups of art objects to visually similar objects in the Europeana data set, metadata based linking, etc.

For this demonstration the front-end is built for web browsers, using HTML5 and JavaScript technology for the implementation of the interactive user interface. This front-end works on top of the existing WebTV platform[4] used for the project, an XML based service-oriented platform where audiovisual content is stored, processed into different formats and qualities and made accessible through a RESTful web service. It's capable of storing and manipulating the audiovisual content, metadata and fragment [3] based annotations of the enriched broadcast.

In addition, the demo shows how the linked content can be unobtrusively integrated into a simple but aesthetically attractive TV interface that can be used both during and after the original broadcast, and thus forms a potential to make archived content more attractive

4. EVALUATION AND FUTURE WORK

The demo is released in May 2012 and will be made available to a selected group of potential users. Subsequent evaluations will be carried out within the context of LinkedTV. This will focus on usability of adding additional layer of information of TV broadcasts, interaction patterns. Based on the outcomes, we will work in a second version. The ambition of this second version is deployment on a real-life setting.

In building the demo application, we found that automatic linking is not perfect and requires moderation by editors of the programme. To alleviate the amount of editing work, we want to investigate the possibility of using social media and crowd sourcing in order to involve users in supplying additional data about specific items in the show. Using the effort of the crowd, we aim to improve and correct the available (context) data and also explore ways to visualize the user's perspective on the material. In this process we aim to maximize the quantity and quality of the content and aim to minimize the amount of moderation that is needed to correct the automated and user generated input. Lastly, subsequent versions will investigate possible ways of involving and engaging users, for instance by creating games or giving the opportunity to make users experts on certain topics.

5. ACKNOWLEDGMENTS

LinkedTV is funded by the European Commission through the 7th Framework Programme (FP7-287911).

6. REFERENCES

[1] Troncy, R., Mannens, E., Pfeiffer, S., and Deursen, D. van. 2012. *Media Fragments URI 1.0 (basic). W3C Proposed Recommendation*. Available at: http://www.w3.org/TR/media-frags/.

AUTHOR'S ADDRESS DETAILS

Netherlands Institute for Sound and Vision
Mediapark, Sumatralaan 45. Hilversum, The Netherlands

CWI
Science Park 123. Amsterdam, The Netherlands

Noterik BV
Prins Hendrikkade 120. Amsterdam, The Netherlands

[4] www.noterik.nl/index.php/id/5/item/43/&dummy=1333187451827

Enabling cross-device media experience

Martin Lasak[1], André Paul[1]
[1] Fraunhofer Institute FOKUS; {martin.lasak, andre.paul}@fokus.fraunhofer.de

ABSTRACT
In this demonstration we show how a cross-platform web technology based application framework on top of a secure communication infrastructure may be used to enhance media consumption. By illustrating the access to data and functionality across a user's disparate personal devices, each consolidated within a personal zone, we argue that novel use cases are feasible and viable with lowered development costs. In particular, we utilize a set of defined application programming interfaces to control television sets and use smartphones to provide and render media content within an overlay network. As controlling televisions or using their data is rather unknown in web applications we try to unveil the potential for many scenarios comprising these connected devices.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and Presentation**]: User Interfaces – *User-centered design.*

General Terms
Design, Human Factors, Experimentation

Keywords
Cross-device, service discovery, media sharing, remote control, interoperability, Web technologies

1. Introduction
For the realization of a new cross-device media experience we have build a demo application by using webinos [1]. webinos is an open source platform, designed for the converged world where applications and services run distributed across many devices, locally or remote or sometimes in the cloud. The webinos technology has been built on top of state of the art HTML5 standardization effort, widget and device API and many other standards. To handle the cross-device challenges, innovation has been created in the following key concepts and fields. The *personal zone* is a proposed method of binding devices and services to individuals, and for that individual to declare their identity. With *remoting and discovery* capabilities devices are given a way to broadcast their services, applications to discover these services, and a protocol for invoking these services. Implementing an *overlay network* provides a virtualized network overlaying physical networks to allow devices to communicate optimally, over the Internet if required, or over local bearer technologies if that is more appropriate.

2. Background
In recent years we have seen many devices and gadgets appearing on the market with the capability to generate or render media content. Although, some vendors provide proprietary synchronization or sharing features in their device portfolios, unified secure and network agnostic access schemes are not widely deployed. From the user experience point of view manual transfer of media files or synchronization with cloud storage solutions is practiced lowering the attractiveness of handling media or making media inaccessible in some situations. Cross-device and cross-domain media content and service access without storing data unnecessarily at third party cloud services is an intelligible desire. The core webinos architecture is based on the state of the art widget and web runtimes, which consist of rendering components, policy and permission frameworks, packaging components and extended APIs. Realizing cross-device communication, webinos splits out the packaging, policy and API extensions from the renderer. By loosely coupled components unlike hitherto monolithic structures, it is easier to expose application centric components to the renderer, such as different TV hardware implementations by exposing it through a Web API. Applications call the interface that will invoke the service provided by the implementing component via remote procedure calls rather than having the service implementation integrated within the renderer. Similarly, the abstraction from specific file systems through a set of common operations provided by a discoverable file service, enables access to stored media content distributed across different devices.

3. Requirements for a cross-device media management application
One of the key requirements for this demo application is to use a phone to take control of the television. Scrolling through long menu structures and channel lists should be eased by not just replacing the remote control with its software equivalent. Using a mobile device, such as a phone or a tablet, searching for channels can be made faster, for example by applying list filter operations. Additional and relevant information can be retrieved on demand by utilizing each device's capabilities in the most meaningful way. Another requirement for our cross-device media management application we addressed is the unified access to media stored on all accessible personal devices composing a distributed file system and not necessarily relying on cloud storages [2].

3.1 Controlling TV from web applications
To achieve the cross-device, cross-domain experience we use a smartphone to navigate the menus on television innovatively and easily through gestures with the mobile device. To our knowledge Web technologies based gesture control in this domain is a novel approach [3]. A further advantage of using the smartphone are its mobile device input capabilities that are not provided by usual remote control units - accelerating the process of searching for specific items, like channel names or stored media content. Despite this added value on the one hand, touch input based smartphones lack physical feedback provided by hardware remote control buttons as well as infrared units. Therefore, the intuitive control of the television is accomplished in our demo application through the combined use of the device orientation API to detect roll and pitch movements of the smartphone and the TV module API to access the tuner data and functionality on the television side.

Meta data or media content can be pulled synchronized from the television onto the smartphone. For example, if watching

television with friends and popping out the room, instead of halting the show for everyone the media can be consumed seamlessly on the smartphone without interruption. Similarly, additional media or channel information presented on the personal handheld does not disturb the presentation on the TV and it's audience.

In this demo application we utilize the webinos infrastructure for discovery of the remote services accessed by the webinos APIs and an event based messaging for the cross-device communication. Both the mobile as well as the television parts of the application are written with Web technologies. The permission system and privacy built into webinos prevents the television being controlled by just any phone, due to the user being able to set explicit access permissions.

3.2 Seamless cross-device media management

The functionality to share media like images, audio or video across different personal devices and between friends is included in this demo application as well using webinos enabled open source technology. With security and privacy built in, users are able to select what they desire from any device and choose the location and device on which to enjoy their books, films and media.

Further, the media sharing aspect in this webinos demo application illustrates how with any webinos enabled devices, like for example television, smartphone and PC, remote control of media can be handled in a way that lets users choose which device they want to use to control or use for media output. This is done without the developer having to take care of the underlying communications protocols on the involved devices. Thus, a smartphone may be regarded as a window into the overlay network of all connected devices and printed objects, making seamless links between the media stored on all personal devices in a secure way with control resting firmly with the user. Having provided that, the actual rendering of media on a device may be decided if certain conditions met or performed on user's explicit command.

4. Demo overview

We have built our demo application for the reference implementation of the webinos open source platform. Figure 1 depicts a general architectural overview of webinos comprising the concept of the personal zone. The personal zone consists of a single personal zone hub (PZH) and multiple personal zone proxy (PZP) per user. Whereas a PZH is a logical entity that resides on a server the PZPs are present locally on every webinos device. Along with the access to the local device capabilities the PZP provides a service discovery and message exchange mechanism if connected to the PZH or directly to other PZPs over secure channels. A Web runtime communicating with the local PZP and the relaying of service calls to other PZPs facilitate the interoperable execution of webinos application on disparate devices from different domains. The remote service access is utilized in our demo application, for example, to access the television's channel list information while using smartphones. The application consists of two parts, each part to be distributed to a disparate device, for example by installing the appropriate widget from an application store. One device, e.g. a large screen connected to a PC or a webinos capable television set, acts as a media renderer and another device, for example a Android smartphone, is used as a control and input device. But rendering of content on the smartphone is as well possible to facilitate the second screen experience. Beside rendering of media of different types and providing trick functions, like to playing or skipping media items, browsing through media content is possible that is present in the unified webinos file system consisting of file services being discoverable and accessible on connected devices. Queries and filter for specific media items can be applied. Both distributed parts of the application communicate via the webinos Events API leveraging a set of defined application events to synchronize states and invoke trick functions and commands.

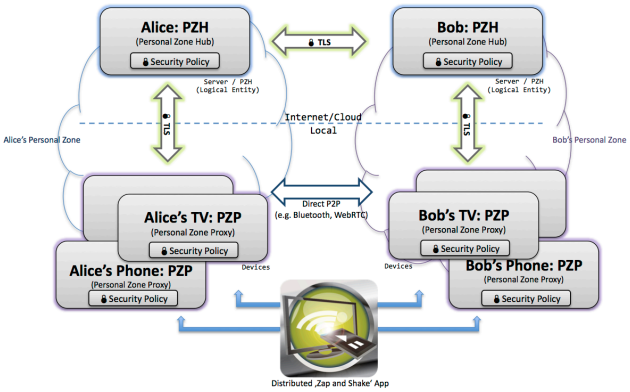


Figure 1: A distributed prototype application deployed for the webinos platform

5. Conclusion

Building applications on top of an overlay network for cross-device communication and interoperable service remoting shows that sophisticated scenarios for new television experiences are implementable with low effort. Applying standards technologies, defining JavaScript APIs and making them remotely available makes devices from different domains, such as television and smartphones, accessible for Web developers who otherwise would need to acquire expert knowledge and taking care for the secure cross-device communication on their own.

6. References

[1] Christian Fuhrhop, John Lyle, and Shamal Faily. 2012. The webinos project. In *Proceedings of the 21st international conference companion on World Wide Web* (WWW '12 Companion). ACM, New York, NY, USA, 259-262. DOI=10.1145/2187980.2188024 http://doi.acm.org/10.1145/2187980.2188024

[2] Serge Defrance, Rémy Gendrot, Jean Le Roux, Gilles Straub, and Thierry Tapie. 2011. Home networking as a distributed file system view. In *Proceedings of the 2nd ACM SIGCOMM workshop on Home networks* (HomeNets '11). ACM, New York, NY, USA, 67-72. DOI=10.1145/2018567.2018583 http://doi.acm.org/10.1145/2018567.2018583

[3] Mathias Baglioni, Eric Lecolinet, and Yves Guiard. 2011. JerkTilts: using accelerometers for eight-choice selection on mobile devices. In Proceedings of the 13th international conference on multimodal interfaces (ICMI '11). ACM, New York, NY, USA, 121-128. DOI=10.1145/2070481.2070503 http://doi.acm.org/10.1145/2070481.2070503

Enhancing Media Consumption in the Living Room: Combining Smartphone based applications with the 'Companion Box'

Regina Bernhaupt
IRIT
118, Route de Narbonne
31062 Toulouse, France
regina.bernhaupt@irit.fr

Thomas Mirlacher
ruwido
Köstendorferstr. 8
5202 Neumarkt, Austria
thomas.mirlacher@ruwido.com

Mael Boutonnet
IRIT
118, Route de Narbonne
31062 Toulouse, France
mael.moutonnet@irit. fr

ABSTRACT
When consuming media in the living room, a broad range of different devices and their individual (remote) controls are used. Switching between these devices and controls diminishes the overall positive experience of consuming media and is sometimes even a cumbersome task. The "Companion Box" solution is consisting of a hardware component and a smart phone application that aimes to solve this problem: The companion box is able to emit and receive infrared (IR), radio frequency (RF) and it offers a wirless local area network (WLAN) connection. The smart phone application can send commands via a WLAN connection to the Companion Box and the Companion Box in turn transmits an IR command to the televison (e.g. to change the volume), or an RF command to the set top box (to go to a menu). The demonstration of the companion box consists of the usage-centered smart phone application and the fully functional (hardware) companion box and demonstrates how to enhance media consumption in the living room.

Categories and Subject Descriptors
H. 5.2. **[User Interfaces]** I.5.2 **[Design Methodology]** Mobile Phone application design.

General Terms
Design, Human Factors.

Keywords
smart phone applications, remote control, control, entertainment environment, cultural probing.

1. INTRODUCTION
Media consumption in the living room can be a complicated task. To simply watch a movie, the user has to perform a sequence of tasks including several devices, remote controls and even smart phone applications or tablet applications. To depict how complicated a simple task like watching a movie can be, we use a short story: A French television (TV) channel broadcasted a movie saturday night. Frank recorded that movie on saturday on his external hard disk that is connected to the set top box (STB). Two days later he wants to watch the movie. To really enjoy movies, Frank owns a Digital Surround System. To watch the movie Frank has to: turn on the TV, the STB, the external hard disk and the Dolby Surround System. Since he wants to watch the movie from his STB, he uses the TV remote control to switch to the proper input where the STB is connected, so he can actually see the output on screen. Using the remote control of the STB, he has to go to the user interface/menu of STB, select the external hard disk and start the movie to play. Since he cannot hear the audio of the movie, he takes the remote control of the Dolby Surround System to switch to the input the STB is connected to and also sets the volume of the movie. Once the movie is playing he leans back and enjoys - only until the advertisement break. The sound is unusually loud and he tries to quickly control the volume. Based on his usual habit he grabs the remote control of the TV and changes the volume. But the volume is still too loud and he notices that he is using the Dolby Surround System - so he takes the remote control of the Dolby Surround System and changes the volume. He then changes the remote control and uses the remote control of the STB to skip the advertisement.

Goal of the companion box and the smart phone based companion application is to enhance the overall user experience of media consumptions in situations like these. A detailed description of the problem area, an ethnographic study on requirements and a set of design recommendations for such smart phone based applications can be found in [1].

2. State of the Art
There is a variety of applications available that were proposed to enhance the control of media consumption in the living room. In the scientific literature Lorenz et al [3] proposed a set of gestures on a smart phone to control the ambient media player. A complete survey of the smart phones as input devices for ubiquitous environments is available in [2]. From the industrial perspective, vendors and producers of iTV and IPTV solutions offer a variety of applications for smart phones and tablets. Solutions are also available from over-the-top (OTT) services for example Apples *Remote* developed for iPhone, iPod and iPad. To a limited extend, such applications are also available for smart phones with the Android operating system.

The limitation of all these applications is that they typically can only control the STB (typically via WLAN). Once the user wants to control any other device in the living room, it is necessary to use an additional remote control.

To overcome the limitations of WLAN based control, there is a set of devices allowing to control any type of IR-based device in the living room. Examples are L5 or myTVRemote for iPhones. These products are simple add-ons that are fixed on the phone via power consumption or USB, and that are able to emit infrared. A set of standalone solutions like Peel or Gear4 (Unity Remote) offer the user the possibility to put the device in their home. Other products additionally allow controlling the whole home infrastructure including heating or lights for example BeoLink from Bang & Olufsen.

Devices enabling the control of all different types using different transmission media (IR, RF) and protocols used for entertainment devices are (to the best of our knowledge) currently not available on the market, neither have they been investigated from a scientific media study perspective.

3. SMART PHONE APPLICATION AND SYSTEM ARCHITECTURE

The architecture of the "Companion Box"-Solution consists of two parts: the smart phone application and the companion box.

(1) Smart phone application: The user interacts with a smart phone application to control all devices in the living room. The user interface of the set top box can be directly controlled using swipe gestures allowing the user to go left, right, up, down or to simply confirm a selection with ok (tap in the middle) or go back (Figure 1). The control of devices is enhanced and optimized based on findings [4] that the majority of function/buttons of remote controls is only rarely used. The smart phone application is taking only the most often used functions/buttons on the remote controls into account, and offers simply a list of other functions for rarely used functionality.



Figure 1: Smart Phone application with direct navigation.

The application allows users to select their program in an electronic program guide (EPG), whenever the smart phone is having Internet connection. It enables the construction, management and usage of (pre-programmed) command sequences. Users can for example use a pre-programmed scenario to turn on all devices they need for watching a movie.

Figure 2 shows the architecture of the companion box solution: left the smart phone is depicted. The smart phone communicates with the companion box via WLAN. The companion box is able to emit and receive IR, RF and WLAN commands. It also provides access to a database with all types of IR commands for TV-related entertainment devices.

When interacting with the companion box solution, the user simply uses his smart phone application. For example user Frank selected the pre-programmed scenario to turn on all devices needed to watch a movie.

The smart phone application sends a command to turn on devices to the companion box. The companion box translates this command in a sequence of commands and sends an IR command to the TV indicating to turn on the device, it also sends and RF command to the STB to turn on the device and it additionally sends an IR command to the Dolby Surround System.

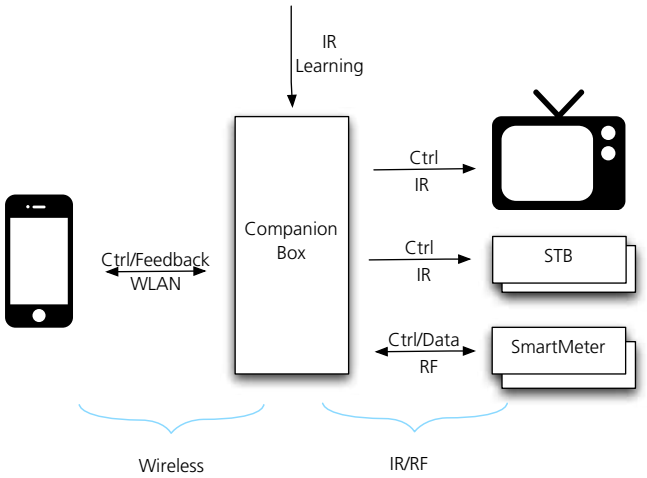


Figure 2: The companion box system architecture.

4. THE DEMONSTRATION

The companion box demonstration will allow conference participants to directly interact with the smart phone application of the companion box solution (direct control, EPG, video on demand, usage-oriented scenarios). It allows participants to see how the companion box translates commands to a set of IR/RF codes. Depending on available infrastructure, the IR commands will be sent to a real TV, STB and other IR/RF controlled devices, or results of the command will be shown in a simulator/demonstrator running on a PC.

5. ACKNOWLEDGMENTS

Our thanks go to ruwido that supported the development of the companion box and the related application. This work is partly funded by the project living room 2020 which is a joint-project between ruwido and IRIT.

6. REFERENCES

[1] Bernhaupt, R., Boutonnet, M., Gatellier, B., et al. 2012. A set of Recommendations for the Control of IPTV-Systems via Smart Phones based on the Understanding of Users Practices and Needs. In *Proceedings of EuroITV 2012*, to appear.

[2] Ballagas, R., Rohs, M., Sheridan, J. and Borchers, J. 2006. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 5(1): 70 - 77.

[3] Lorenz, A. and Jentsch, M. 2010. The Ambient Media Player - A Media Application Remotely Operated by the Use of Mobile Devices and Gestures, in *Conference Proceedings of MUM'10*, 372-380.Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[4] Mirlacher, T. and Bernhaupt, R. 2011. Breaking myths: inferring interaction from infrared signals. In *Proceedings of EuroITV 2011*. ACM, New York, NY, USA, 137-140. DOI=10.1145/2000119.2000146

A Continuous Interaction Principle for Interactive Television

Regina Bernhaupt
IRIT
118, Route de Narbonne
31062 Toulouse, France
regina.bernhaupt@irit.fr

Thomas Mirlacher
ruwido
Köstendorferstr. 8
5202 Neumarkt, Austria
thomas.mirlacher@ruwido.com

ABSTRACT
Users of interactive television can select from hundreds of channels, thousands of videos and an ever increasing number of services including catch-up TV offers, video on demand or various forms of electronic program guides. While the number of elements on the user interface was increasing the control in the user hands stayed the same. The continuous interaction principle for interactive system allows users to quickly and effectively search in large quantities of data. The principle was developed based on a set of identified contextual factors that influence interactive TV consumption today, including physical, temporal, social and economical context.

Categories and Subject Descriptors
H. 5.2. **[User Interfaces]** I.5.2 **[Design Methodology]** Mobile Phone application design.

General Terms
Design, Human Factors.

Keywords
remote control, control, entertainment environment, interactive TV, search, video on demand

1. MOTIVATION
The increasing number of services available on interactive TV and Internet Protocol based TV systems makes interaction with the systems more and more complicated. Interactive television systems today are often described as being unusable due to their limited interaction concept enabling either one-to-one functionality (a key on a standard remote control gives access to exactly one function) or a limited interaction through a menu (using navigation keys, number keys or color keys on a remote control and a user interface enabling the user to select within the user interface). New forms of interaction have been proposed in the scientific community to enhance the standard remote control interaction, e.g. by introducing Wii-like interactions [6], [5], gesture [4], speech based interaction [2] or employing second screen approaches such as smart phone applications or other ubiquitous computing solutions [3].

From a scientific usability oriented perspective, these new forms of interaction techniques promise to increase the bandwidth between user and system and they thrive to enhance the user experience overall when interacting with the system. From a more global software engineering and industrial perspective, usability and user experience are just two of many factors that have to be addressed when introducing a new form of interaction technique. While gestures might be "fun" when used for the first time, current systems are still in its infancy and do not offer the necessary reliability, security and performance to really allow a broad range of TV users to interact with the system [4]. The goal for a new form of interaction technique would be to enhance/extend the bandwidth between user and system, especially enhancing the selection and browsing so the user can navigate quickly in large quantities of data.

Other factors that seem to be barriers for a successful introduction of new interaction technologies in the living room are:

physical context: the living room is special in terms of physical characteristics: the TV screen is rather big, the user is typically meters away from the screen and the screen is typically between 5 to 10 years old. The question is how to take into account the distance between user and screen, especially how can we support the user to interact with the large screen without the need to continuously change his eye gaze between screen and remote control.

temporal context: usage of media in the living room is heavily time dependent, ... how can an interaction technique support the most frequently used temporal (media oriented) interaction aspects: time shift/pause, forward/back?

personal context: depending on personal characteristics, but in general watching TV is still predominately passive. How can the interaction technique support lean back behavior but give overall a user experience that the interaction technique would allow a certain degree of activity?

social context: media consumption on “the big screen” stays to be a social activity or an experience that people want to enjoy together. The “big screen” is a public device and engaging activities should be offered for all of them. Watching TV together is not like playing a Wii game together. What should be kept in mind, is that the social context leads to the requirement that the TV control can be used by and is accessible to anyone in front of the TV.

economical context: new interaction solutions for the TV are rather price intensive as they include motion capture, cameras, position sensors etc. which multiplies their cost by a factor of up to two hundred.

The goal thus was to develop a new form of interaction technique that allows to support the above described living room situation: enhancing the bandwidth especially for search in large quantities of data, enabling “blind” usage by the user, supporting most frequent activities, respecting passive usage and enabling every member of a household to access and use the interaction technique.

2. CONTINOUS INTERACTION

The continuous interaction principle for interactive TV shall allow to search in large quantities of data, enable the user to browse content in an enjoyable way but keep in mind that watching TV overall is still rather passive and that the interaction technique for the TV is used (and usable) by all members of a household.

The continuous interaction principle consists of two parts:

(1) an input device that enables continuous input and at the same time continuous feedback and (2) a user interface that corresponds to this mechanism.

(1) Input Device: In terms of interaction, we came up with a solution for the input device that uses a movable element, which can be pushed and pulled by the user. The movable element is based on a patented interaction solution [8] that enables continuous feedback to the user: the harder/faster the user pushes, the more haptic resistance the user perceives. The visual feedback reflected by the graphical user interface, is provided by the speed of the movement and presentation of the items on screen.

To support the overall user experience, the input device was designed as a monolithic structure [Figure 1 (left)]. The overall award winning design concept is called aura [1] and the goal was to communicate that this type of interaction is able to support users in their most preferred activities: searching through media content and simply watching media content.



Figure 1: The aura design concept consisting of a remote control (left) and the user interface metaphor (right).

(2) To support the continuous interaction moving forward and back, the graphical user interface takes up this metaphor and provides a list of elements that the user can browse through in only one direction [Figure 1 (right)].

To select an element within the presented menu, the user presses an ok/select button. The interaction and depth of menus is designed based on previous work on enhancing the usability of iTV systems [7]. Error recovery is supported as users can come back to the main menu by repeatedly pressing the back button.

The continuous interaction mechanism is currently focus of set of user experience and usability studies. Main goal is to make the interaction intuitive and emotionally engaging by respecting each individuals personal usage of pressure and speed. We started to investigate various usage forms to optimize what the user feels as feedback and what the user interface presents. For example: a user wants to move fast through the on-screen menu: the plate on top of the remote control is pushed fast and strong forward and then suddenly released. The user interface uses a different algorithm to visualize the end of the selection by slowing slightly down the speed for showing new items, giving the user a more comfortable feeling.

3. DEMONSTRATION

Participants of the conference can interact with the fully functional prototype installation of aura, consisting of the input device and the interaction mechanism. Depending on availability of technical equipment the user interface can be presented on a PC or a real TV.

4. ACKNOWLEDGMENTS

We would like to thank ruwido for the continuous efforts in developing the aura concept, especially Thomas Fischer and Christian Schrei.

5. REFERENCES

[1] Aura. Red Dot Award 2012.

[2] Balchandran, R., Epstein, E. P., Potamianos, G. and Seredi, L. 2008. A multi-modal spoken dialog system for interactive TV, *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*, 191-198.

[3] Ballagas, R., Rohs, M., Sheridan, J. and Borchers, J. 2006. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 5(1): 70 - 77.

[4] Bailly, G., Vo, D.-B., Lecolinet, E. and Guiard, Y. 2011. Gesture-Aware Remote Controls : Guidelines and Interaction *Techniques Proceedings of ICME 2001*, 263-270. DOI: 10.1145/2070481.2070530.

[5] Kela, J., Korpipo, P., Marvi, J., Kallio, S., Savino, G., Jozzo, L. and Di Marca, K. 2006. Accelerometer-based gesture control for a design environment. *Personal Ubiquitous Computing* 10, 5 (July 2006), 285-299. DOI=10.1007/s00779-005-0033-8

[6] Lin, J., Nishino, H., Kagawa, T. and Utsumiya, K.. 2010. Free hand interface for controlling applications based on Wii remote IR sensor. In *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry* (VRCAI '10). ACM, New York, NY, USA, 139-142. DOI=10.1145/1900179.1900207.

[7] Mirlacher, T., Pirker, M., Bernhaupt, R., Fischer, T., Schwaiger, D., Wilfinger, D. and Tscheligi, M. (2010). Interactive Simplicity for iTV: Minimizing Keys for Navigating Content, In *Proceedings of EuroITV 2010*, 137-140, ACM.

[8] Patent. Aura Interaction Solution 189/5609-Z.

Story-Map: iPad Companion for Long Form TV Narratives

Janet H. Murray, Sergio Goldenberg, Kartik Agarwal, Tarun Chakravorty, Jonathan Cutrell, Abraham Doris-Down, Harish Kothandaraman
Georgia Institute of Technology
Digital Media Program
Atlanta GA 30308 USA
+1 404 894 6202
{jmurray,sergio.goldenberg,kagarwal9,tchakravorty3,jcutrell3,abraham,harish7}@gatech.edu

ABSTRACT
Long form TV narratives present multiple continuing characters and story arcs that last over multiple episodes and even over multiple seasons. Since viewers may join the story at different points and different levels of commitment, they need support to orient them to the fictional world, to remind them of plot threads, and to allow them to review important story sequences across episodes. We present an iPad app that uses a secondary screen to create a character map synchronized with the TV content, and supports navigation of story threads across episodes.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - *Prototyping, Interaction styles, Graphical User Interfaces*. H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering*

General Terms
Design, Experimentation, Human Factors, Standardization

Keywords
Interactive Television, Dual Device User Experience, Second Screen Application

1. INTRODUCTION
The growth of digital formats has increased the consistency and continuity of television writing, by making the single season into a story-telling unit, and keeping events from past seasons current in viewer's minds, through distribution on DVDs and web-based fan activities. As a result, writers have incentives to create complex multi-character stories with plot-lines that arc across multiple episodes and even multiple seasons. But more complex plots and larger casts of recurring characters can leave viewers confused. Since the primary delivery is still one episode per week, new viewers need to be given contextual information about what has happened before, if a show is to gain viewership mid-season; and loyal viewers will also need reminders of characters or plot events that may be picked up from months before.

2. RELATED WORK
StoryLines [5] explores a timeline-based method of navigating news and episodic television in the context of rich archival resources. Users can identify individual story threads within a complex multi-threaded story environment by filtering the items on a composite timeline, which can be split into multiple timelines. Motorola Mobility [1] have prototyped a companion device experience that enhances TV viewing by providing synchronized semantically related auxiliary information and media on a second screen. The viewing aid released by the producers for watching Game of Thrones on HBO GO includes information synced with individual episodes [2] and provides additional information about some of the characters in an episode, but leaves a new viewer wondering about the many unexplained characters, relationships, and dialog references in this richly imagined fantasy world, without providing direct links to clarifying information.

In her 1997 book, *Hamlet on the Holodeck*, Janet Murray offers a prediction of a "hyperserial" that would grow out of the digital delivery of television content [4], predicting many of the applications such as virtual places and point of view information that have become staples of web sites associated with dramatic series and films. Henry Jenkins in *Convergence Culture* called attention to "transmedia storytelling" based around film and television properties, in which games and fan participation structures are not just marketing extentions but integrated parts of the canonical world [3].

The Story-Map application builds upon earlier experiments with tablets used as secondary synchronized screens, and upon the concepts of the hyperserial and transmedia storytelling, by creating an application that is intended to foster the viewer's understanding and immersion in a complex story-world by making the component parts of the story structure apparent, and fostering multiple paths of coherent exploration.

3. SYSTEM DESCRIPTION
3.1 Setup
The system consists of two main components. The first is the Internet capable television, which displays the television show as streaming digital video and the second is the companion tablet that provides viewers with auxiliary information about the show.

3.2 Features
The companion app provides three features that augment the viewing experience: Character map, Relationship Recaps, and Thematic recaps (in this case, Shootings by the protagonist).

The prototype was developed based on content from the first season of the FX television series *Justified*. The show chronicles the adventures of U.S. Marshal Raylan Givens as he shoots bad guys, befriends his ex-wife, and struggles to maintain his integrity as a lawman while returning to his native Harlan County which is filled with family members and old friends who have varying degrees of criminality.



Figure 1. Real time character map

3.2.1 Character Map

The character map (see Figure 1) is a real time, updating graph of characters and relationships. When a character appears within the television show, their thumbnail image appears in real time on the companion iPad app and a connecting line with an icon represents their relationship with other characters on the screen and in the world of the series. In addition, characters are arranged semantically across the area of the tablet screen. Icons on the lines connecting characters indicate the nature of the relationships between them. Touching the character thumbnail brings up a brief bio, which changes to reflect the unfolding dramatic revelations and events without exposing the viewer to "spoilers".

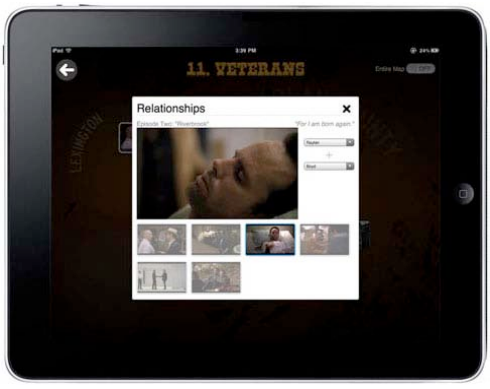


Figure 2. Relationship Recap

3.2.2 Relationship Recap

The Relationship Recap (see Figure 2) is available from the top level of the tablet app, or during a viewing of an individual episode, by tapping on the relationship icon between two characters. This brings up an overlay box that contains short video clips that summarize the plot thread concerning two key characters.

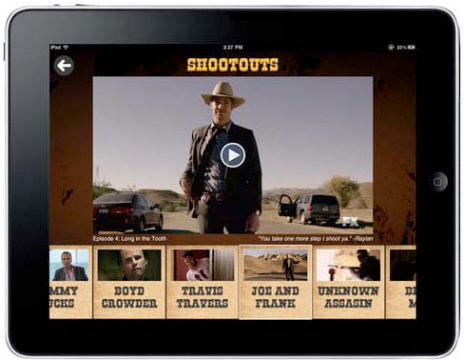


Figure 3. Video clips of shootings by the protagonist

3.2.3 Thematic Recap

The application also offers a thematic recap, in this case, all the iconic cowboy-style shoot out scenes featuring the quick draw hero (see Figure 3). The interface represents a list of characters who have been shot, and the shooting video plays on tapping.

4. CONCLUSION

This project offers an approach based on story structure to establish some conventions for making such aids helpful without creating unnecessary distractions from the viewing experience. User tests are being planned and will be reported on separately. Though developed for one series, *Justified*, the functionalities are presented abstractly so that they can be applied to the genre of television dramas.

5. REFERENCES

[1] Basapur, S., Harboe, G., Mandalia, H., Novak, A., Vuong, V. AND Metcalf, C. 2011. Field trial of a dual device user experience for iTV. In *Proceddings of the 9th international interactive conference on Interactive television* (EuroITV '11). ACM, New York, NY, USA, 127-136.

[2] HBO GO, 2011. Game of Thrones Interactive Experience. http://www.hbo.com/game-of-thrones/about/video/hbo-go-interactive-experience.html

[3] Jenkins, H. 2006. *Convergence Culture: Where Old and New Media Collide*. New York University Press, New York.

[4] Murray, J.H. 1997. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Free Press, New York.

[5] Murray, J.H., Goldenberg, S., Agarwal, K., Doris-Down, A., Pokuri, P., Ramanujam, N. AND Chakravorty, T. 2011. StoryLines: an approach to navigating multisequential news and entertainment in a multiscreen framework. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (ACE '11), Teresa Romão, Nuno Correia, Masahiko Inami, Hirokasu Kato, Rui Prada, Tsutomu Terada, Eduardo Dias, and Teresa Chambel, Eds. ACM, New York, NY, USA, Article 92, 2 pages.

Wordpress as a Generator of HbbTV Applications

Jean-Claude Dufourd
Telecom ParisTech
37-39 rue Dareau
75014 Paris, France
+33145817733
dufourd@telecom-paristech.fr

Stéphane Thomas
Telecom ParisTech
37-39 rue Dareau
75014 Paris, France
+33145818057
thomas@telecom-paristech.fr

ABSTRACT
This demo presents an innovative reuse of the very well known Wordpress blog management software, to help teams with scarce resources produce and maintain easily simple HbbTV applications. The demo shows two variants: a generator for a "modern" teletext application, and a generator reusing the concept of Wordpress widgets. The demo features an HbbTV TV set where the applications are displayed, and a laptop serving as both application server and management console. Viewers can change content in the management console for in either of the variants and can see the result immediately on the TV.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: Online Information Services – *Commercial services, Web-based services;* H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; H.5.4 [**Information Interfaces and Presentation**]: Hypertext / Hypermedia – *User issues*

General Terms
Algorithms, Experimentation, Standardization.

Keywords
HbbTV, interactive TV applications, Wordpress.

1. Introduction
This demo is for 2 templates.

2. Modern Teletext Template
This template consists in a Wordpress theme and a Wordpress plugin. It was designed specifically to target HbbTV terminals. It builds on a Wordpress plugin which helps personalize resources (header, footer, background, etc). This template adapts your blog by displaying its pages in a manner optimized for HbbTV. Posts are simply ignored. Page navigation uses JavaScript and Ajax in a manner compatible with HbbTV.

This template proposes two possible structures : one menu in the left third and one page in the right two thirds, or one menu in the left third, one sub-menu in the middle third, and one page in the right third.



Figure 1: 3-columns page



Figure 2: 2-columns page

The plugin adds a new settings menu in the Wordpress dashboard. For each page, one can choose the two columns or three columns layout and any resource. Left and middle menu elements are empty pages with just a title. Middle menu pages are hierarchically below left menu pages, and right pages are below a menu page according to the chosen layout.

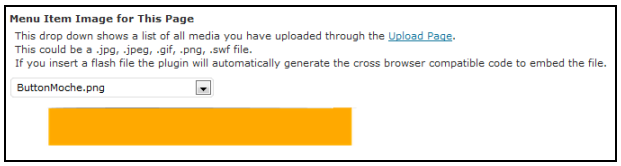


Figure 3 : changing the aspect of the menu item

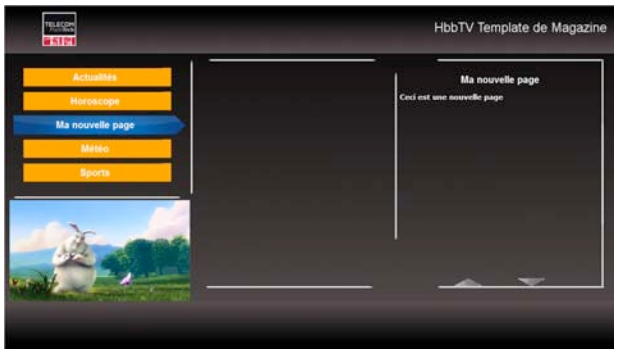


Figure 4: result of changing the menu item aspect

3. Widget Template

This template consists in a Wordpress theme and a Wordpress plugin. This template allows to use Wordpress sidebars while staying compatible with HbbTV. It enables the author to create an HbbTV application which consists solely of sidebars in which Wordpress widgets can be inserted. The theme only defines and manages the four sidebars, whereas the plugin defines a sample Weather widget, as well as prototypes of other widgets such as a widget to manage the video/broadcast object of HbbTV. Pages and posts are ignored in this particular template.

.

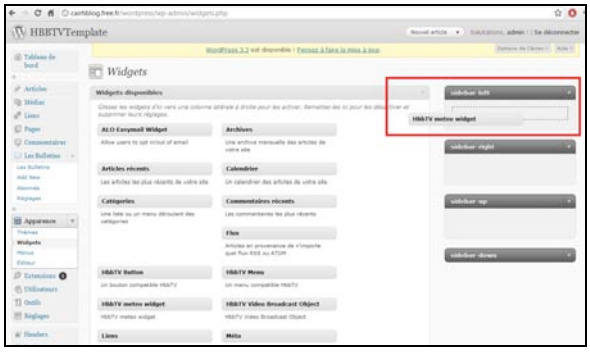


Figure 5: install widget by drag&drop



Figure 6: configure the widget



Figure 7: result



Figure 8: 3 widgets in the same sidebar

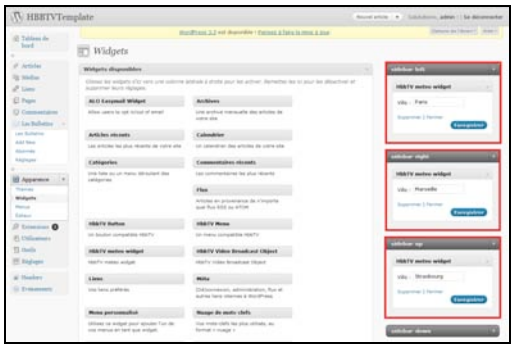


Figure 9: one widget per sidebar

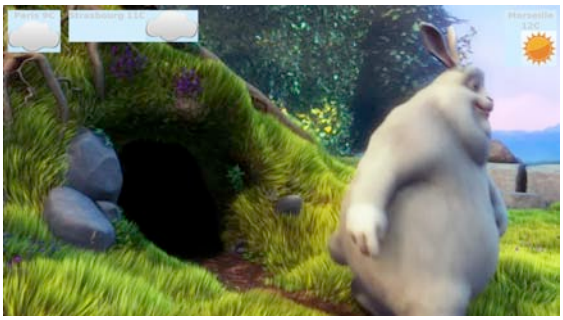


Figure 10: result

4. ACKNOWLEDGMENTS

This work was done within the French national project openHbb (www.openhbb.eu).

SentiTVChat: Real-time monitoring of Social-TV feedback

Flávio Martins, Filipa Peleja, João Magalhães

Dep.Informática, Fac. Ciências e Tecnologia,
Universidade Nova de Lisboa, Portugal

flaviomartins@acm.org, filipapeleja@gmail.com, jm.magalhaes@fct.unl.pt

ABSTRACT

In this paper, we present a Social-TV chat system integrating new methods of measuring TV viewers' feedback and new multi-screen interaction paradigms. The proposed system analyses chat-messages to detect the mood of viewers towards a given show (i.e., positive *vs* negative). This data is plotted on the screen to inform the viewer about the show popularity. Although the system provides a one-user / two-screens interaction approach, the chat privacy is assured by discriminating information sent to the shared screen or the personal screen.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]

General Terms

Design, Experimentation, Human Factors.

Keywords

SocialTV, chat, sentiment analysis, multi-screen interaction.

1. INTRODUCTION

According to Haythornthwaite [1] media popularity is linked to social interactions in the new media. They concluded that social ties and social media are important to users' media viewing habits. More recently, Harboe et al [2], conducted an experiment examining the influence of Social-TV watching, i.e., users watch television alone but get instant notifications of what their friends and family are watching. In a related user study Weisz et al. [3], examined the activity of chatting while watching video online. Their technological solution was designed to observe human factors and not to improve usability. Recently, Ariyasu et al. [4] proposed to detect the topics of twitter messages and to associate them to the correct show. In our system, the show is known and we proposed to infer the sentiment of the messages towards the show. Thus, we set out to develop a Social-TV system prototype that allows users to chat with each other, in an integrated way. The system's contribution is twofold:

- **Multi-screen interaction paradigm**: the system User Interface spans several screens and users interact with media through two screens.
- **Measuring TV chat feedback**: a novel viewer feedback metering method is proposed by integrating sentiment analysis technology into a Social-TV chat system.

With the proposed system, user messages are processed in real-time by a sentiment analysis algorithm, allowing the plotting of the emotions felt by viewers, [5].

2. A SYSTEM FOR SOCIAL-TV CHAT

The SentiTVChat system, allows "Social-TV-viewers" to communicate with each other during TV-viewing activities using a simple chat interface, shown in Figure 2 and Figure 3. The main goal was to develop a system for multi-screen TV-viewing activities and to enable the system to collect chat interactions with real-time analysis of the chat text to sense user moods.

2.1 System architecture

The system is divided in two parts: a Social-TV service and the SentiTVchat service, see [5]. The Social-TV service is responsible for the main media application and for binding user devices in the same session with a common communication channel. The chat system implements the chat service and messages analysis technology.

2.2 Social-TV service

Users can access the system by using any Web browser, capable of running JavaScript, which includes most modern browsers for laptops and the browsers included in iOS and Android (Figure 2). The users can use mobile devices, such as tablets, to remote control the TV and participate in chat rooms with other people.

TV-Screen: The Media Player UI

To build the TV-screen interface, in Figure 1, we used the resources provided in the Google TV documentation for HTML5. In addition, popular browsers started shipping preliminary support for the Full screen API allowing the player to turn most modern browsers into a full-blown media consumption screen.

Tablet-screen: The Personal Control UI

The remote-control functionality includes standard controls such as Play/Pause and other control tasks, it also offers directional keys for navigating between user-interface buttons on the TV-screen. After authentication, a bidirectional channel is opened between the Tablet-screen and the TV-screen. This allows the TV-screen application to send updated media metadata information such as program title and progress to the secondary devices. To authenticate the user, we use two factors: the user's Google account login information and a random alphanumeric 5-digit code that is available to the user in the TV-screen display corner. This provides better security and allows a single user account to be able to control a number of TV-screen devices.

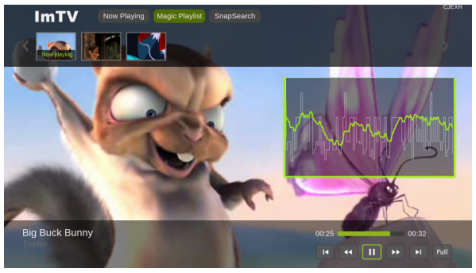


Figure 1. Media player with SentiTVchat graph.

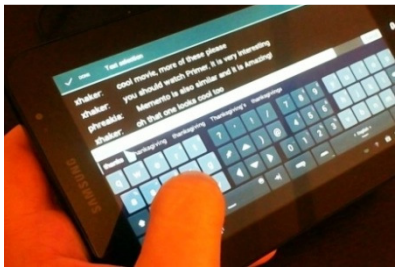


Figure 2. Tablet device in chat mode.

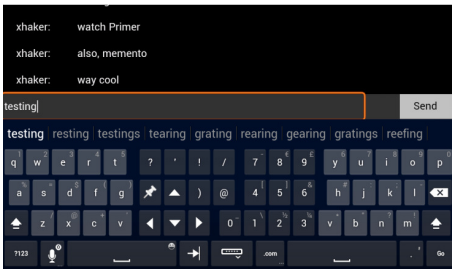


Figure 3. The personal control user interface

Binding devices: a session based communication channel

To bind the user's devices and to implement communication channels between them we looked into Web Sockets. Current browser support for Web Sockets is limited and implementations differ substantially between browsers providing this API. Thus, we decided to make use of Google App Engine APIs to build our system so we could utilize the Channel API available on this platform. The Channel API is similar to Web Sockets, it allows Web applications to establish bidirectional communication channels. However, browser support is improved, since the API is provided to the client-side in a JavaScript library. This means that any JavaScript capable mobile browser should be able to communicate with a multitude of HTML5 ready devices running the player, such as a Web TV, a PC or a tablet.

2.3 SentiTVchat service

The implemented chat system supports the chat communications among users seamlessly. The Tablet-side UI gets the current TV-channel from the updates received from the TV-screen application. Thus, when users access the chat section in the Tablet-screen it will automatically enter the corresponding chat room for the TV-channel being watched on the TV-screen. We considered that each TV-channel has a chat room with an identical *name*, so, if the user is watching the *tv1* channel, for example, the Tablet-screen retrieves the URL: *chat-host.tld/?room=tv1*. The page obtained displays the latest 1000 messages from the chat room. Moreover, a subscription is registered on the server, so that the client can receive further messages instantly. A subscription is registered with the chat room, the user, and the communication channel. When a message is received, it is relayed to all the subscribers of the corresponding chat room through each of their channels. However, to send messages, the client makes a POST request to the */newMessage* endpoint instead of using the channel.

2.4 Chat sentiment analysis

Consider a set of N user-comments, $D = \{(c_1, p_1),...,(c_N, p_N)\}$, where a user-comment c_i is labeled as positive $p_j = 1$ or negative $p_j = 0$. Taking a machine learning approach and considering the training set D, sentiment chat analysis aims at learning a classifier function, $\Phi : c_i \rightarrow p_i \in [0,1]$, such that for all new user-comment, $c_i \notin D$, the function Φ will infer a polarity value $p_i = 1$ for positive comments and $p_i = 0$ for negative comments. User-comments, are represented as a vector of *opinion words*, i.e., $c_i = (ow_{i,1},...,ow_{i,m})$ where each component $ow_{i,m}$ depicts the opinion word m of the user-comment i. An opinion word is a word that can express a preference, which can have different intensities (provided by SenitiWordNet) and different orientations (inferred by Pointwise Mutual Information technique), see [5] for details.

For classifying reviews, we used a linear classifier to assign a confidence value to that review. The classifier identifies the orientation and intensity of all opinion words of a comment $c_i = \left(ow_{i,1},...,ow_{i,m}\right)$ and computes its rating based on the sigmoid function,

$$\Phi\left(c_i\right) = \frac{1}{1 + \exp\left(-\sum_j ow_{i,j} \cdot w_j\right)} + b\,. \qquad (1)$$

The weights w_j are learned with a gradient descent algorithm to train the function to rate comments between positive and negative. The constant *b* is a bias factor for adjusting the sentiment curve.

3. DISCUSSION

This paper proposes a novel Social-TV chat system that measures in real time the viewers' mood / show popularity. By measuring and plotting viewers' mood, users that are zapping through channels will have instant access to the show popularity over the last minutes. Additionally, this measure will be closer to real viewers' preferences than by simply measuring audiences. As future work, we will improve the sentiment analysis for other languages and for chat slang.

Acknowledgements. This work has been funded by the Foundation for Science and Technology by projects UTA-Est/MAI/0010/2009 and PEst-OE/EEI/UI0527/2011, Centro de Informática e Tecnologias da Informação (CITI/ FCT/ UNL) - 2011-2012.

4. REFERENCES

[1] C. Haythornthwaite, "The Strength and the Impact of New Media," *International Conference on System Sciences (HICSS-34)*, 2001.

[2] G. Harboe, C. J. Metcalf, F. Bentley, J. Tullio, N. Massey, and G. Romano, "Ambient social tv: Drawing people into a shared experience," *ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, Florence, Italy, pp. 1-10, 2008.

[3] J. D. Weisz, S. Kiesler, et al., "Watching together: Integrating text chat with video," *ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose, California, USA, pp. 877-886, 2007.

[4] K. Ariyasu, H. Fujisawa, and Y. Kanatsugu, "Message analysis algorithms and their application to social tv," *EuroITV*, 2011.

[5] F. Martins, F. Peleja, and J. Magalhães, "SentiTVChat: Sensing the mood of Social-TV viewers," *EuroITV*, 2012.

Doctoral Consortium

Towards a Generic Method for Creating Interactivity in Virtual Studio Sets without Programming

Dionysios Marinos
University of Applied Sciences Düsseldorf
Josef-Gockeln-Str. 9
40474 Düsseldorf
+49-21-4351830

dionysios.marinos@fh-duesseldorf.de

ABSTRACT

In this paper, research aims and objectives towards a generic method for creating interactivity in live virtual studio productions without the need for traditional programming are described. The difficulties behind the use of traditional programming to create interactive virtual sets are presented and the need for a generic method that compensates for these difficulties is discussed. A brief description on current work as well as the necessary steps towards the theoretical and practical completion of the method is provided. The method will aim at minimizing the effort associated with creating virtual set interactions, by providing the necessary mental and software tools to help a production crew break down a desired interactive behavior into a combination of elementary modules, the configuration of which takes place based on examples, avoiding the need for traditional programming. Showing the superiority of the resulting method compared to current solutions will be a significant contribution to live interactive virtual studio research that could lead to a PhD.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *artificial, augmented, and virtual realities, evaluation/methodology.* D.1.7 [**Software**]: Programming Techniques – *visual programming.* H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, prototyping, theory and methods, input devices and strategies.*

General Terms

Performance, Design, Reliability, Experimentation, Human Factors, Theory.

Keywords

Interactive Virtual Studio, Programming by Demonstration.

1. INTRODUCTION

Virtual studios [3] nowadays provide a very attractive alternative to traditional television studios. This is due to the fact that virtual studios use computer graphics to create a virtual television set, in which one or more talents can act and moderate. The use of computer graphics for this purpose leads to a number of advantages regarding the costs of a production and its content. The same virtual studio can be used for a large number of different productions, each one having different requirements regarding style and content, making its use very cost effective. The modification of virtual sets is also a big advantage, because it allows for the iterative improvement of virtual sets during the life cycle of a television show without the need to interrupt the production in order to reconstruct or redesign the production set. Regarding the production's content, the use of virtual studios offers many new possibilities as well. Virtual objects with supernatural attributes (e.g. floating or flying objects) can be integrated into the production, creating interesting effects that would not be possible inside traditional television studios. The use of customized visual interfaces and presentation elements inside a virtual set opens up the way to creating shows that are highly informative and visually appealing to the audience. The virtual studio offers a generic medium for expressing creativity through the integration of real world objects and people inside a limitless computer generated world. Such a medium does not necessarily have to be constrained to news or weather programs. Theatrical plays or other kind of artistic performances could also take place inside a virtual studio [5].

In order to bring up the potential of the virtual studio, it is necessary to focus on and enhance the plausibility of the integration of real people (performers/moderators) inside a virtual world (virtual studio set) with respect to the audience. This could be done by increasing the resolution, offering more realistic lighting and shading or improving the keying process responsible for separating the talents from the background. The deployment of such methods in virtual studios is associated with big costs, originating from the need of replacing a lot of already established production and broadcasting hardware and software. However, there is another aspect that contributes to the realistic integration of real people inside a virtual world, which can be researched upon and improved without the need of costly replacements. This aspect is interactivity between the people and the virtual set.

Nowadays, creating highly interactive virtual sets that robustly react to the actions of the talents and provide interaction possibilities to navigate through, retrieve and present information during live show situations is so difficult, that broadcasting studios have to simulate such interactions by letting one or more human operators trigger the right events at the right time through the use of control interfaces, like buttons, sliders and knobs, from inside the control room. The difficulty in creating real interactive sets is due to the big number of different factors that have to be analyzed and evaluated at every moment of the program in order to support the ongoing interactions. This analysis/evaluation is specific to the requirements of the respective program and, at the moment, can be only addressed with a low level approach requiring programming (traditionally or visually). Such an approach involves interfacing many different components, like e.g. tracking systems, virtual set elements, talent feedback mechanisms, lightning and sound systems in regard to time related factors associated with the flow of the program. The programming

of such interfacing mechanisms is a laborious task and bears resemblance to programming virtual reality applications. Therefore, this method is not desired and cannot be easily adapted from a typical virtual studio crew, who is accustomed to work with high level authoring tools.

The objective of my research is to develop a generic method for creating highly customizable, robust and visually appealing interactions between talents and virtual sets without the need of traditional programming or scripting. The method should comply with the following requirements, which are already fulfilled by programming/scripting techniques:

- Compatibility with current methods and systems
- Possibility to iteratively refine or expand the created interactions
- Possibility to create novel, unexplored interaction styles

The potential superiority of the method should be examined based on the following criteria:

- How easy it is for a virtual studio crew to adapt to this method compared to traditional programming/scripting
- How much effort is required to turn an interaction concept into a functional prototype
- How much effort is required to turn a functional prototype into a working system ready for use in a live virtual studio environment
- To which extent the resulting interactions reflect the desired interactions, specified by the production crew during the conception, realization and test phases of the interaction development cycle.

In order to understand the principles behind such a method, an overview of related work is provided in the next section.

2. RELATED WORK

An approach that tries to solve part of the difficulties associated with traditional programming is *Programming by Demonstration* [2] or *Programming by Example*. This is a development technique that enables a developer to define a sequence of operations without programming, by providing examples of the desired behavior under certain conditions. Assuming the developer has provided enough examples for different conditions the system is then able to generalize and successfully interact even when the conditions change. In the field of creating interaction techniques by demonstration, Brad Myers presented in [6] a tool called *Peridot*. He concludes "*Peridot successfully demonstrates that is possible to program a large variety of mouse and other input device interactions by demonstration*". He also states "*Almost all of the Macintosh-style interactions can be programmed by demonstration...*". His tool was able to generate code for the created interactions, in order for it to be linked with an end-user application. A couple of years later, he created Marquise [7], an "*interactive tool that allows virtually all of the user interfaces of graphical editors to be created by demonstration without programming*". Another interesting example of a tool for creating interactive user interface prototypes by demonstration is *Monet* [4]. The designer is able to design the interface consisting of multiple widgets and provide examples that describe the relationship between mouse input and widget behavior. The application also allows the user to associate different widget configurations with states. These states can be used to create the interface logic by creating some kind of a finite state machine. The authors wrote "*Monet supports a wide variety of continuous behaviors, including both standard widgets such as scroll bars and unusual ones representing application semantics. These can be prototyped in a consistent and simple way...*". They also conclude with "*Monet greatly expands the visual language for informal prototyping and enables more complete tool support for early stage UI design.*".

A similar tool created by David Wolber almost 10 years before Monet was presented in [9]. The tool was named *Pavlov* and was an attempt to enable the development of highly interactive graphical interfaces without the need for programming. Wolber identifies the problem: "*Though interface builders like Visual Basic have significantly decreased the time and expertise necessary to build standard interfaces, the development of more graphical, animated interfaces is still mostly performed by skilled programmers.*". He presents a programming by stimulus-response demonstration model that also incorporates time, in order for time-based interactions to be possible. He writes: "*Stimulus-response provides a cohesive model for demonstrating interaction, transformation, and timing. The model seeks to minimize the cognitive dissonance between concept and design by allowing designers to demonstrate the behavior of an interface exactly as they think of it: "When I do A, B occurs", or "two seconds after the start of the program, this animation begins.""*". Wolber conducted a user study, in which inexperienced users had to create a couple of interfaces using Pavlov. The authors were "*...extremely encouraged by the results, as well as the enthusiasm the subjects expressed for exploring....*".

In a publication of mine and my colleges accepted for the EuroITV 2012 conference in Berlin [5], a new interfacing tool was introduced, named *OscCalibrator*. This tool enables the creation of multidimensional parameter mappings by utilizing examples provided by the user. The interfacing of this tool is based on the Open Sound Control (OSC) protocol [10], which makes the tool compatible with almost any modern virtual studio and media authoring environment. The user is able to pass tracking or other control values over to the tool and to define useful output values for the respective input values. This method, backed by a specific interpolation scheme, leads to the creation of continuous mappings between input and output values. These mappings are then used in real time to produce the necessary output values, which are able to drive the desired interactions. The peculiarity of OscCalibrator is that it promotes a novel method on how to deal with tracking data and interaction parameters in virtual studios in order to support the creation of interactivity. Instead of traditionally programming the necessary mappings, the user can exploratory define them with examples. The user changes the output parameters for the respective input data so, that the interaction looks the way it is desired. When a new case emerges, in which the interaction is not working properly, the user can add new mapping points, until he is satisfied with the outcome. This method provides a very attractive alternative to traditional programming, when it comes to location-based interactions in virtual studios, and it is associated with a significant effort reduction.

Belivacqua et al. developed in [1] an HMM-based system for gesture analyzing and tracking. The system is able to follow a gesture in real time and continuously output important parameters,

such as the current time progression and the likelihood of the demonstrated gesture. To make this possible, the user must first provide sample gestures through single examples. Specifying some additional application-specific parameters is also necessary. The system has been used successfully in a number of interactive systems that had mostly to do with music and video performances. This system opens up many new possibilities for interactive live performances, when it comes to the synchronization of movements with computer-generated sequences (animation, sound, and video). An examination of the system, as part of a method to create interactivity in virtual studios seem to be meaningful.

In [8] a generic method has been described on how fuzzy rules can be extracted from numerical examples. The generated rule set can be extended with additional rules originating from human experience. It has been proved that the fuzzy system generated from these rules "*is capable of approximating any nonlinear continuous function on a compact set to arbitrary accuracy*". This method could form the basis of an investigation, which is aimed at producing tools that create interaction logic for virtual sets based on examples.

3. TOWARDS A NEW METHOD

In the related work section, a series of research projects have been presented, in which examples, provided by the user, where used to define relationships that lead to the creation of interaction and control mechanisms. When we focus on a specific domain, like the interactivity in virtual studio sets, and we try to create a new generic method for supporting the creation of such interactivity, we will have to specify the kind of relationships that define every possible interactive behavior in this domain. After specifying these relationships, it is a matter of determining how these relationships can be expressed through the use of the available tools. The resulting method for creating the desired interactive behavior would then consist of the following steps:

1. Specifying the necessary components and their relationships
2. Using and configuring the available tools in order to express these relationships
3. Testing the resulting interactivity and identifying new relationships or problematic ones in order to iteratively expand and improve them (going back to 1. and 2.).

The steps above seem to be generic and to provide a guide for creating any kind of interactive system. The first challenge here is to concretize these steps so that the resulting method can be realistically used in a virtual studio environment. The second challenge is to demonstrate the superiority of this method compared to other methods, like traditional programming and scripting.

Focusing on the first challenge, the steps of the method have to be concretized. For that purpose, we will have to identify the kinds of relationships that may occur.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference'10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010…$10.00.

3.1 Identifying the Right Kind of Relationships

The relationships that may occur during the interaction of one or more talents with a virtual set can be safely classified in three main categories:

1. Spatial relationships
2. Temporal relationships
3. Logical relationships

These categories correspond to mental associations created when we interact with a system or watch someone interact. For example, if a moderator takes a certain position and the lighting of the virtual set changes, we are going to associate this change with this new position, leading to a spatial association on the side of both talent and spectator. Temporal associations may also be provoked. While the moderator walks, for example, from left to right and a virtual screen gradually appears, the spectator (by watching the program) and moderator (by acting and getting some kind of feedback) perceive two different actions that happen synchronously, suggesting a time relationship between the two. Additionally, if, for example, a new person appears in the show and a banner with his/her name shows up on the screen of the spectator, a logical relationship between the appearance of a new person and that of the corresponding banner is induced. Any interactive behavior in a virtual set could be described with a set of such relationships and even if not, it is a matter of identifying the right kind of relationships that make this description possible.

3.2 Creating Tools that Express these Relationships

Providing tools that can be combined and configured in such a way that expresses the relationships necessary to describe an interactive behavior is crucial. These tools would have to be able to interface with the virtual studio software and hardware systems as well as with themselves, in order to allow any combination of the desired relationships with arbitrary complexity to be expressed. In a best case scenario, every identified kind of relationship corresponds to a single tool, so that the interaction designer could unambiguously select the right tools to implement his/her concept. Assuming the kind of relationships suggested in the previous section, the corresponding tools could look like this:

1. A mapping tool, like e.g. *OscCalibrator*, could be used to define functional associations between position data and interaction parameters. Such a tool is able to express spatial relationships in an easy and consistent way.
2. A tool similar to the one developed by Belivacqua et al. developed in [1] could be used to express temporal relationships. With such a tool a sequence can be associated with another sequence, so that executing the first one leads to the synchronous execution of the second.
3. A tool based on the research for generating fuzzy rules from examples, like in [8], could provide a straight forward way for creating and expressing logical relationships.

As it is the case with OscCalibrator, the configuration of which takes place based on examples, the tools mentioned above also

work in the same fashion. This realization suggests that a method backed by such tools would provide a consistent way of creating interactive behavior. And apart from that, the fact that demonstrational examples are used to perform the necessary configuration indicates that during the design and development of an interactive set the actions of the participants would correlate better to the desired end-result.

For the creation and the refinement of the necessary tools, the work practices and experience of virtual studio interaction designers have to be considered. The first step in this direction would be to provide early versions of the tools to undergraduate students who are creating interactive virtual studio productions in our facilities as part of their studies. The collected feedback will be used to refine the tools, leading to more stable and easy-to-use versions that would be forwarded to our virtual studio partners and other interaction experts for further reviewing.

4. EXAMINING THE SUPERIOTY OF THE METHOD

Showing that the suggested method is superior to the available methods for creating interactivity in virtual sets would be the biggest contribution of my research, together with identifying its limitations. This could be shown with the help of a series of generic test cases that demonstrate how the suggested method can be applied to create arbitrarily complex interactivity in virtual studios. By comparing with the effort needed to create such interactive systems using traditional methods, like scripting or visual programming, the advantages of the method will become obvious. The cases, in which the traditional methods would perform better, would be used to identify the limitations of the suggested method.

Some indications that point out to the advantages associated with an example based method for live interactive virtual studio productions are the following:

1. Example based approaches tend to minimize the cognitive dissonance between concept and design, allowing for a more rapid development.
2. The participation of the talents during the configuration of the system, by letting them provide examples, makes them familiar with the interactions, helps with recognizing problems early and reduces the overall time needed for testing and rehearsal.
3. The use of examples, provided by the talents under the supervision of a coordinator, could lead to interactions that can be easily performed and look the way they are supposed to.

Normally, traditional programming has to be performed by an expert programmer, who tries to fulfill the requirements mediated by the interaction designer. If the requirements cannot be fulfilled as originally designed, the interaction has to be redesigned and eventually compromised. Providing a method that makes clear to the designer which tools he/she has to use and how to configure them in order to implement any desired interactive behavior in his/her domain is the key to decoupling the interaction design process from the programming aspects that compromise it.

5. FAZIT

In this paper, I presented the aims and objectives of my research towards a method that would help people create interactivity in virtual studio environments. The method should clarify what kind of interaction relationships it supports and should provide a toolset for expressing these relationships in a consistent manner. The configuration of these tools should be performed based on examples, avoiding the need for programming and reducing the effort associated with prototyping, testing, redefining and rehearsing interactions for live virtual studio productions. Such a method would open the way to the acceptance of virtual studios, not only as a rich broadcasting medium for news, weather or other information driven programs, but also as a generic interactive medium for expressing artistic and entertaining concepts.

6. ACKNOWLEDGMENTS

I would like to thank all scientists and researchers who provide *free* access to the source code and designs associated with their work and, in this way, support the verifiability of their contribution.

7. REFERENCES

[1] Bevilacqua, F. et al. 2010. Continuous Realtime Gesture Following and Recognition. *Gesture in Embodied Communication and Human-Computer Interaction*. S. Kopp and I. Wachsmuth, eds. Springer Berlin Heidelberg. 73-84.

[2] Cypher, A. and Halbert, D.C. 1993. *Watch what I do: programming by demonstration*. MIT Press.

[3] Gibbs, S. et al. 1998. Virtual Studios: An Overview. *IEEE Multimedia*.

[4] Li, Y. and Landay, J.A. 2005. Informal prototyping of continuous graphical interactions by demonstration. *Proceedings of the 18th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2005), 221–230.

[5] Marinos, D. et al. 2011. Design of a touchless multipoint musical interface in a virtual studio environment. (2011), 1.

[6] Marinos, D. et al. 2012. Large-Area Moderator Tracking and Demonstrational Configuration of Position Based Interactions for Virtual Studios. *EuroITV 2012* (Berlin, 2012).

[7] Myers, B.A. 1987. Creating dynamic interaction techniques by demonstration. *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface* (New York, NY, USA, 1987), 271–278.

[8] Myers, B.A. et al. 1993. Marquise: creating complete user interfaces by demonstration. *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems* (New York, NY, USA, 1993), 293–300.

[9] Wang, L.-X. and Mendel, J.M. 1992. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*. 22, 6 (Dec. 1992), 1414-1427.

[10] Wolber, D. 1996. Pavlov: programming by stimulus-response demonstration. *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground* (New York, NY, USA, 1996), 252–259.

[11] Open Sound Control.

Providing Optimal and Involving Video Game Control

Thomas P. B. Smith
School of Computing Science
Newcastle University, UK
thomas.smith@ncl.ac.uk

ABSTRACT
In this paper I discuss my PhD research in advancing video game controller design, how I wish to expand this into interactive television and how my work can help others in the area of interactive television.

Categories and Subject Descriptors
K.8 [**Personal Computing**]: Games – *miscellaneous*

General Terms
Design, Human Factors, Standardization, Theory.

Keywords
Video games, video game controllers, video game interaction, interactive television, trans-media.

1. INTRODUCTION
Video games create rich and engaging experiences for players and are enjoyed across the world. Gamepad and motion controllers for consoles and the PC's keyboard and mouse have become the primary means of video game input even though their general-purpose nature means that they are not perfect for most games but the most viable option to settle for.

Interactivity is what differentiates games from any other entertainment medium; therefore it is paramount that the control method employed to provide interactivity is as optimal for the game as possible. A poor control device, or poor appropriation of a control device, will very simply lead to a poor game experience for the player and increases the developer's workload as they wrestle with non-optimal controllers.

In my research I propose two approaches for producing optimal video game controllers, which will solve the problems of using general-purpose controllers by suggesting the use of player-facing and developer-facing dedicated devices.

2. RESEARCH SITUATION
I am currently conducting research for a full-time PhD as part of the Digital Interaction group in the School of Computing Science at Newcastle University. I have been on the program for five months and so anticipate attaining my doctorate in three and a half years.

The structure of my PhD research will follow a portfolio approach, where my research into video game controllers will culminate in the evaluation of produced test cases at the end of my first year. This will allow me to use my findings in the different but related application areas of interactive television and augmented tabletop gaming and incorporate other customization and personalization methods into my research such as user-generated content and social media.

3. RELATED WORK
3.1 Alternative Game Controllers
For many years, dedicated controllers have been created when general-purpose controllers haven't been sufficient. [9] shows an interesting and comprehensive taxonomy of console controllers including dedicated controllers such as light guns and music/rhythm-based controllers. [1] describes the creation of a music/rhythm-based controller for use with a custom, banjo music game. The research describes the process of creating a custom controller from start to finish including hardware and software design. [8] discusses that we are currently seeing an emergence of emotional controllers based on interaction through gestures and tactile inputs. The research itself used the facial emotion of players as input to the game, requiring players to look happy, sad etc. to manipulate the game world.

3.2 Customizable Game Controllers
The VoodooIO Gaming Kit [11] is a framework to create real-time, customizable game controllers. The interface allows players to have complete control over their control schemes. Users can choose different inputs and arrange them where they want them on the controller. Commercially, Madcatz released a customizable game controller [5] although the customization is very limited; giving the player three areas on the controller to choose between four different forms of directional input.

3.3 Summary
The fact that all of the work cited above is very recent shows that control is becoming important to players, developers and researchers alike. The commercial viability of customizable controllers is shown by the production of [5]. The feedback from [11] shows that players are comfortable with using a customizable controller framework and that it benefits their play. Also, the results of [8] show us that emotional input in games is fun for players. Using this information, I have decided to develop a customizable controller framework for the production of dedicated controllers by players (optimization-led) and developers (narrative-led).

4. RESEARCH CONTEXT

4.1 General-Purpose Controllers

Video game controllers fall into two major conceptual categories if we look at control methods for all modern game platforms, general-purpose and dedicated devices. General-purpose devices are what we see in terms of the major consoles and the PC.

The problem with general-purpose controllers is that players receive less than optimal controls for the game/genre they are playing and developers are restricted to producing less than optimal player experiences due to this. For example, a very popular genre called the First-Person Shooter is widely played on both PCs and consoles. Aiming using a joystick on the controller is not as optimal as aiming with a mouse as shown by [7] but the continuous states provided by the joysticks allow for much more control over the speed and acceleration of directional movement than the binary states provided by the directional keyboard keys. Therefore, a more optimal controller for the First Person Shooter genre would be a mash-up of the keyboard/mouse and console controller input devices.

4.2 Dedicated Controllers

The Fightstick [5] is a modern reinterpretation of the controls provided on the arcade cabinets where the "Fighting Game" genre was born. Modern fighting games are released on home consoles but the Fighting Game enthusiast community quickly found console controllers were not viable for providing the accuracy or durability they required and so Fightstick controllers were developed based on the original arcade controllers.

The Driving genre also enjoys a large community of enthusiasts who require a much more optimal experience than a keyboard or console controller can provide. Steering wheels and pedals exist to provide much more accurate input for driving games but also output is very important to these controllers. Force-feedback technology [3] creates a more realistic experience; by driving motors inside the steering column, the steering wheel fights against players when they attempt sharp turns at high speeds.

Both Fightsticks and force-feedback steering wheels are more optimal controllers for their genres than traditional general-purpose controllers but for different reasons. Fightsticks deliver better control by providing the exact number of buttons desired and the required durability due to the forcefulness of game play, in essence providing a better experience through optimizing redundancy and improving build quality. Whereas the steering wheel provides a better experience through replicating the real-world experience of driving a car; force-feedback technology deliberately delivers a more difficult experience than a steering wheel without force-feedback but also a more accurate experience.

Therefore, we can see a difference in the design approach between Fightsticks and Force-feedback steering wheels; the Fightstick is designed as an optimization-led controller whereas the steering wheel is designed as a narrative-led controller.

4.2.1 Optimization-Led Dedicated Controllers

I define optimization-led dedicated controllers as those in which the player decides her exact control over the game. This stems from the notion that the controller should be invisible to the player as it is a cumbersome necessity in the experience. [4] and [5] are examples of this.

I have devised the following requirements for an optimization-led dedicated controller:

- Provides inputs only for the controls required.
- Provides the best inputs for the controls required.
- Provides for the play style expected from the control scheme.

As can be seen, the ability to provide the "best" inputs for the controls required is subjective. This acknowledges that not all players will think the same controller is optimal even if there is no redundancy and the inputs map the game's control scheme well. This is recognized by games developers as games usually allow the player to alter which input maps to which in-game control. The problem is that not enough customization is given to the player and there is a potential for input redundancy. Therefore, I propose that hardware customization as well as control scheme customization would provide a much more functional and useful controller for players. This solution is a player-facing solution as the player has complete control over the customization of their inputs.

4.2.2 Narrative-Led Dedicated Controllers

I define narrative-led dedicated controllers as those in which the player's control directly replicates the game as much as the narrative affords. This stems from the notion that the controller should be an integral part of the game as the player's control is an important message to convey for narrative purposes. [8] and [3] are examples of this.

I have devised the following requirements for a narrative-led dedicated controller:

- Provides inputs only for the controls required.
- Provides controls through integration with the narrative elements.
- Provides a specific style of play expected from the narrative.

The second requirement states that control should be provided through integration with the narrative which puts control into the hands of the developer rather than that of the player. This is supported by the third requirement which states the style of play is expected from the narrative, something only the developer knows.

4.2.3 Problems in Developing Dedicated Controllers

The lack of attention on dedicated controllers can be attributed to the practical and financial risks associated with in-house hardware development at predominantly software-based games developers. It is much safer to invest all of the project's time and money into the software side of player experience which can be widely distributed to many gaming platforms.

5. AIMS AND OBJECTIVES

5.1 Aim of Research

The primary aim of this research is to enhance players' experiences with video games through producing practically viable, novel controllers and design methods for game designers to produce optimization-led and narrative-led controllers.

5.2 Objectives

The objectives which must be met to attain the aim are as follows:

- An examination of controller research literature and commercial ventures.
- Development of hardware and software to aid in the creation of optimization-led and narrative led dedicated controllers.
- Work with professional games developers to help in the creation of dedicated controllers.
- Creation of design methods to aid game designers in incorporating narrative-led controllers into games.

6. RESEARCH METHODOLOGY

The research will follow a very practical path, supporting hypotheses through the development of test cases which will be analyzed through real-world use. I will develop the hardware and software systems for these test cases myself.

6.1 Towards an Open Modular Framework

I propose building a modular, customizable hardware and software framework as the basis for dedicated controllers. Two major problems with dedicated controllers are solved using this method. Obtaining an in-house hardware division or outsourcing the creation of controllers to other companies is solved by this approach. Also, the problem of requiring players to purchase every bespoke controller is solved as the customization of modular components allows for each possible controller to be created.

The responsibility of acquiring the correct or desired inputs and outputs is given to the player. For developers, the framework will allow them to write software to send and receive data to and from the different inputs and outputs without the need to produce or distribute these hardware components.

6.1.1 Benefits for Optimization-led Controllers

For optimization-led controllers the modular framework means the player has complete control over the controller they want to play with. This fulfills all of the requirements for an optimization-led controller.

The developer has the responsibility of making sure the framework is compatible with their game which would only entail software development. They could also provide recommendations for hardware configurations to further engage with their community.

6.1.2 Benefits for Narrative-led Controllers

Using the framework for producing narrative-led controllers is more complicated as they are developer-facing rather than player-facing. Simple narrative-led controllers could use the same pattern as optimization-led controllers but require certain inputs and outputs rather than allowing the player to have a choice. Yet this does not leverage the full range of opportunities provided by narrative-led controllers. More interesting narrative-led controllers would require additional hardware such as custom controller cases and additional but inexpensive hardware add-ons. This reduces the work of developers from creating bespoke controllers and would be an enjoyable experience for the player, akin to following instructions to assemble a Lego creation.

6.1.3 Framework Technology

Microsoft Gadgeteer [6] is a rapid prototyping platform for small electronic devices. Assembly of components is modular and solder-less, allowing for the quick and simple construction of devices. Gadgeteer is open-source, allowing modules and mainboards to be created and sold by multiple electronics manufacturers. There are currently many different input and output modules from simple buttons to the more complex gyroscopes, cameras and networking modules. Due to the simple assembly and vast array of interesting modules, I believe Gadgeteer is the best technology for the framework.

6.2 Evaluation Methods

A challenging but very important aspect of the research is how the notion of player immersion will be evaluated. There has been much research into player immersion and an often used method is the IEQ [2]. The IEQ consists of 31 questions relating to the player's feeling of immersion in six different sections. Comparing these results using the dedicated controllers to an experience of the same games using general-purpose controllers will give a good indication of how beneficial the dedicated controllers are. Further to this, quantitative data will be collected based on the player's in-game interactions to further evaluate how they play the game.

It would be beneficial for these tests to be carried out over multiple sessions to attain more realistic results of a player's experience with a game, which isn't usually limited to one session. This would also be supported by playing the games in the players' regular gaming environments such as their living room/bedroom.

6.3 Towards Design Methods for Designers

Although the framework makes it easier for designers to produce dedicated controller-based games, it does not guarantee a successful implementation, as it is possible that some implementations could disrupt immersion in a game more than if a general-purpose controller was used. Through my experiences of using the framework and testing controllers and games I produce, I will be able to provide insights into the difficulties faced when producing optimization-led controllers and narrative-led controllers as fluid extensions of the software part of the game.

7. PROGRESS SO FAR

7.1 Open Modular Framework

The framework is currently in development and allows basic optimization-led controllers to be used with existing games. By creating an additional piece of software which captures the controller's inputs and spoofs mouse and keyboard events, any existing game can be played. This will allow tests to be performed with current games rather than spending a lot of time developing my own games which isn't necessary for testing the framework.

7.2 Narrative-led Controller Test Case

I am also developing a small game in collaboration with a PhD student in the Digital Interaction group. The game is called *Bib* and uses a narrative-led controller as its input. *Bib* is the protagonist within the game but also the controller. We are exploring ideas of co-dependence between the character and the player as well as agency within the controller to show how unique and involving a narrative-led controller-based game can be. This will demonstrate that this experience cannot be replicated without a game design which includes a narrative-led controller.

8. TIME PLAN AND FUTURE WORK

8.1 Time Plan

8.1.1 Development

The open framework is currently being developed alongside the *Bib* game. I have decided to build them alongside each other as insights I gain from working with the framework for *Bib* allow me to modify and improve the framework as it is being developed.

8.1.2 Testing and Analysis

On completion of the first version of *Bib* and the player-facing optimization-led controller software, I will test the controllers with players to gain feedback on their experiences. The data gathered from these trials will help inform how the framework could be improved.

I will use the IEQ in the evaluation context discussed in the previous section. This complete process should take between 7 and 9 months, bringing me to the end of my first year. I then plan to use the framework and knowledge gained in the distinct but related application areas of interactive television and augmented tabletop gaming while continuing to work with video games.

8.2 Interactive Television

8.2.1 Customization of Television Control

The research into optimization-led controllers will be transferred from the domain of video games to interactive television as they share many parallels when discussing control. The optimization-led controller framework will be expanded to allow for television controllers to be constructed.

Television controllers provide many input buttons, many of which will not be used by the majority of users. The user-facing customization provided by the optimization-led controller framework will allow users to customize the controller to be as optimal for themselves as possible.

8.2.2 Trans-media Interactive Television

The research into narrative-led controllers will be used to research trans-media products between television and video games. The ability to provide a playful experience to viewers of a television program is not a new idea; the *Captain Power and the Soldiers of the Future* television program [10] allowed interaction during its episodes but this was very limited. I intend to use the framework and the interesting inputs and outputs available to Microsoft Gadgeteer to create more compelling experiences.

For example, a children's television program based around *Bib* would show how *Bib* is feeling throughout the program and at specific points allow players to play a small game. Further to this, the player's success in these games would influence what content appeared in the *Bib* game when the player returned to it. This opens up interesting ideas in terms of episodic play and trans-media interaction, all of which I would like to explore.

9. CONTRIBUTIONS TO iTV

Although the project is primarily based around video games for the first year, the knowledge gained will be useful in adapting and expanding the hardware and software framework for use with television controllers. The possibility of research into a trans-media product as described above would contribute knowledge and experience to the interactive television community.

Through my practice-based research into customizable, user-facing and narrative, developer-facing controllers, I hope to advance video game experiences and allow for the easy creation of such experiences. Expanding this to trans-media experiences between video games and television and video games and tabletop games.

10. ACKNOWLEDGMENTS

The research reported in this paper is funded as part of the UK AHRC Project: The Creative Exchange and is carried out under the supervision of Professors Peter Wright and Patrick Olivier.

11. REFERENCES

[1] Ey, M., Pietruch, J., Schwartz, D. I., 2010. "Oh-No! Banjo": a case study in alternative game controllers. *Proceedings of the International Academic Conference on the Future of Gam Design an Technology*. Futureplay '10. ACM, New York, NY. DOI= http://doi.acm.org/10.1145/1920778.1920810.

[2] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), 641-661.

[3] Logitech. Logitech Driving Force GT. 2012. Web. 25/02/2012 DOI= http://www.logitech.com/en-gb/gaming/wheels/devices/4172.

[4] Madcatz. Major League Gaming Pro Circuit Controller for Playstation 3. 2011. Web. 24/02/2012 DOI= http://www.madcatz.com/mlg/PS3_controller.htm.

[5] Madcatz. Madcatz Fightstick. 2008. Web. 24/02/2012 DOI= http://www.madcatz.com/productinfo.asp?page=247&GSProd=4696&GSCat=97&CategoryImg=Xbox_360_Controllers.

[6] Microsoft Research. Microsoft Gadgeteer. 2011. Web. 21/02/2012. DOI= http://research.microsoft.com/en-us/projects/gadgeteer/.

[7] Natapov, D., Castellucci, S. J., MacKenzie, I. S. 2009. ISO 9241-9 evaluation of video game controllers, *Proceedings of Graphics Interface 2009*. Toronto: Canadian Information Processing Society, 2009, 223-230.

[8] Obrist, M., Igelsböck, J., Beck, E., Moser, C., Riegler, S., Tscheligi, M. 2009. "Now You Need to Laugh!" - Investigating Fun in Games with Children. *Proceedings of the International Conference on Advances in Computer Entertainment Technology*. ACE '09. ACM, New York, NY. DOI= http://doi.acm.org/10.1145/1690388.1690403.

[9] Pop Chart Lab. The Evolution of Video Game Controllers. 2011. Web. 20/02/2012 DOI= http://popchartlab.com/collections/prints/products/the-evolution-of-video-game-controllers.

[10] Salmans, S., TELEVISION; The Interactive World of Toys and Television. 04/10/1987. 27/02/2012 DOI= http://www.nytimes.com/1987/10/04/arts/television-the-interactive-world-of-toys-and-television.html

[11] Villar, N., Gilleade, K. M., Ramduny-Ellis, D., Gellersen, H. 2007. The voodooIO gaming kit: A real-time adaptable gaming controller. *ACM Comput. Entertain., 5, 3, Article 7 (November 2007)*, 16 Pages. DOI= http://doi.acm.org/10.1145/1316511.1316518.

The Remediation of Television Consumption: Public Audiences and Private Places

Evelien D'heer
IBBT- MICT- Ghent University
Korte Meer 7-9-11
BE-9000 Gent
+32 9 264 84 77
Evelien.Dheer@ugent.be

Pieter Verdegem, Ph.D.
Dept. of Communication Sciences, Ghent University
Korte Meer 11
BE-9000 Gent
+32 9 264 67 15
Pieter.Verdegem@ugent.be

ABSTRACT
This Ph.D. project aims to provide insight into everyday life in a media pervasive society. By simultaneously investigating mass communication (television broadcasting) and computer-mediated communication (i.e. social media conversation about TV programs), which we refer to as *remediation*, we shed light on interconnected modes of media consumption, emblematic for contemporary life. The new media ecology renders some well-established divisions in the field of media (sociology) anachronistic (e.g. public vs. private). The resulting theoretical as well as methodological issues will be addressed using a multi-method approach. First, both quantitative and qualitative content analysis will be conducted on social media texts (e.g. Twitter messages) on television consumption, with a particular focus on news and current affairs. This will advance our understanding of the consumption of television broadcasting in the new media era. More fundamentally, it will also allow us to rethink traditional conceptualizations of the public space. In a second stage, we will contextualize our findings in the offline domestic context using in-depth interviews. The home is indeed the private space where television consumption (and its remediation) takes place and is negotiated with household members. By considering the offline context, we advance our understanding on blurring lines between the public and the private in contemporary society. Although the remediation of television text is not equally present by all audiences, scholars identify the need to rethink media and audiences.

Categories and Subject Descriptors
J.4 [**Social and behavioral sciences**] - *Sociology*

General Terms
Human Factors, Theory

Keywords
Audience, Convergence, Remediation

1. INTRODUCTION
Life in post-traditional, or often called 'post-modern' society is characterized by a rapidly changing order [1]. Traditional separations become fluid and blurred and social institutions and structures are no longer adequate to maximize self-development and life stories. Reflexive modernity emphasizes individual agency above structure, with meaning and identity being grounded in the self [2]. Despite the opportunities of self-autonomy, creativity and participation, we must be sensitive to the challenges and paradoxes of reflexive modernity. We acknowledge and integrate this broader societal framework in our theoretical elaboration and empirical examination in order to provide insight in the articulation of the networked self in this new media ecology and the many 'hybrids' emblematic for this high modernity.

Mediation is a crucial constituent of everyday life [3]. In this media-drenched, data-rich, channel-surfing age, we no longer live with, but *in* media (cf. 'Media Life', [4]). Although the intrusiveness and ubiquity of communication and information technologies is widely acknowledged, very little empirical evidence exists on the way 'new' and 'old' media are interwoven in the fabric of our everyday lives. The ubiquity of Internet technologies takes the notion of audiences into the public space, whereby they become fragmented and fluid constructions. This raises the question to what extent conventional theories and methods for researching audiences can be extended to the new media ecology and how far some significant rethinking is required.

Abercrombie and Longhurst [5] make a plausible case for the diffused and eternal nature of the audience. People are more connected than ever, participating in and engaging with media through various platforms and channels, including social media platforms [6]. In this respect, we acknowledge the importance of the paradigm 'mass self-communication'[7]. This new type of communication is simultaneously mass communication but also multi-modal and self-generated, self-directed and self-selected. Convergence is not only about merging technological systems, but also about converging places and practices, hence, it is an aspect of our everyday life ('Convergence Culture', [8]). It blurs conventional distinctions in complicated ways (e.g. public and private), requiring not only the need to rethink, but to re-assess audiences. Yet, it seems in the analysis from the old to the new, the symbolic and social, audiences and users of 'new' media technologies remain 'underrepresented' [6]. In this respect, this Ph.D. project aims to contribute to the study of (television) 'audiences' in this fluid modernity, through conceptual elaboration and empirical support in the field of media sociology.

2. RESEARCH OBJECTIVES
2.1 The convergence of technologies and practices
This research project aims to unravel how the mundane (though not trivial) media life (or new media ecology) is interwoven in the fabric of our everyday lives. We conceptualize media life in terms of 'rituals and reflexivity', instead of (fears about) the effects of mass media [9]. Both mass media content as well as talks and activities that emerge around them (e.g. conversations on TV

programs) are incorporated into our everyday lives. The latter is what Fiske [10] refers to as 'secondary discourses' on media content. When conceptualizing and reflecting on audiences, we cannot ignore the diffusion of the Internet and social media (e.g. YouTube, Twitter, Facebook), constituting the era of 'mass self-communication'. As in contemporary media life, old and new communication practices collide, we will focus on the relation between mass communication (broadcasting television) and 'mass self-communication' [7] or what is called remediation. The notion of *remediation* [11] here, reflects the reception of and reflexivity on mass media consumption via social media platforms.

We will conduct audience research from a *constructionist* view ('third generation' audience research, [12]). It entails a broadened frame in which we place media (use). The main focus is not restricted to reception of media messages, but rather to understand our contemporary 'media culture' by studying media texts as an element of the everyday life. In addition, it adds reflexivity on top of the reception of media messages when studying media audiences. Hence, (1) the meaning of particular media texts (e.g. TV programs), (2) the media (and their representations of reality) as well as (3) the activity of being an audience need investigation.

Even in our contemporary life, television remains a focal point for the consumption of popular culture as well as our political life [6]. Concerning the former, a lot of in-depth audience research has scrutinized popular culture [13, 14], whereas news and current affairs have been largely unexplored. We identify the need to incorporate audience research into the context of news and information consumption.

In addition to the television set, we notice an increased density of home media devices. This new media ecology reconfigures existing media (use) and alters their meaning and experience. Therefore, we suggest the consumption of television (mass communication) to be investigated in relation to new and social media (mass self-communication). The ubiquity of Internet technologies, takes the notion of 'audiences' into a new dimension. Audiences are no longer singular objects but become fragmented and hybrid constructions, or as Warner [15, p. 65] refers to: 'multiple concrete publics', constituted via communication, based on shared interests and activated on topical concerns. This needs to be examined in an adequate way. The first research question of the Ph.D. project is:

RQ1: How is television consumption, with a particular focus on news and current affairs, being remediated through social media platforms?

2.2 The convergence of places

In order to fully grasp the remediated consumption of news and current affairs in our everyday lives, we want to advance our understanding of the dynamics that exist between the offline and the online world. Television consumption is predominantly private and thus connected to the domestic life, where consumption meanings are ideally constructed with family members [16]. The ubiquity of the Internet allows and stimulates viewer conversation to take place outside the physical boundaries of the home and renders the family activity of television viewing more complex and dynamic. Whereas much work has been concerned with online communities and identity in cyberspace, "*there is still very little that confronts the relationship between the offline and online world*" [9, p. 201].

Domestic television viewing is frequently combined with other activities, including media consumption (e.g. read e-mails when watching television). These overlapping uses are typically integrated in our everyday communicative routines [17, 18]. In this respect, we feel it is important to know how synchronous remediation practices (e.g. tweeting about television content via the tablet whilst watching) are incorporated the everyday routines and what their significance is in the private domestic space. The domestic environment is referred to as the moral economy and serves as a transactional system through which meanings are exchanged with the public world [19]. In addition, Lull's [16] relational uses of television allow us to re-evaluate interpersonal behavior concerning television consumption, as the Internet renders our understanding of the public and the domestic anachronistic. In this regard, it is valuable to co-assess the public and the private to advance our understanding of its complex constellation. The convergence of places [20] will be understood from a user perspective. Hence, the second research question of this project is:

RQ2: How does the remediation of television content, with a particular focus on news and current affairs, has an impact on the convergence of spaces?

3. METHODOLOGY

Through a mixed-method design, combining qualitative and quantitative research methods, we will achieve our theoretical commitment discussed earlier. The planned research conceptualizes audiences in the private context, with a focus on the domestic context. Simultaneously, we will also focus on audiences in the (online) public space, with a focus on social media texts.

3.1 A quantitative Survey

The quantitative Digimeter survey, monitors trends in new media use, with a sample of approximately 1500 respondents. We supplement this existing survey by including specific questions about social media use in relation to television content, with a specific focus on news and current affairs (e.g. use of Twitter while watching the news).

The purpose of our adjustments to the survey is two-fold:

(1) The survey allows collecting data from a representative sample of the Flemish population. This way it will provide us with information on the phenomenon of remediation of television consumption through social media. In addition, it also allows demarcating the socio-spatial contexts of use, the devices and social media platforms through which remediation practices take place.

(2) The second objective is to develop a typology that can be used to profile distinctive user types (e.g. based on participation level). Working with classifications allows us to gain profound insight in the phenomenon of remediation. Its value lies in the description and comparison between different users to shed light on the divers nature of a, to date, rather new phenomenon.

3.2 Content analysis

In this project, we want to grasp how television consumption is remediated through social media platforms. Both quantitative as well as 'interpretative' qualitative text analysis of social media data will be carried out. The data are available and can be analyzed using methods such as mining, aggregating and

analyzing online textual data in real-time [21]. Content analysis will allow exploring the concept 'mediated publics' (or social media platforms as meso-level spaces) [22]. Although these media are often used for individual purposes, they also play a role in reflection and communication, especially around time-sensitive events (e.g. elections). In this regard, we expect the elections of 2014 to be an ideal case as it triggers both intense television consumption and social media activity. Our corpus thus will focus on the remediation of television content about the 2014 elections on Flemish public broadcasting channels (Eén, Canvas) via social media platforms.

We opt for text-based social media platforms such as Twitter (www.twitter.com). It is a medium that enables (longitudinal) text analysis. Despite the specific type of messages ('tweets') (they are limited to 140 characters), they contain a wealth of data (e.g. thoughts, feelings, activities, opinions). Moreover, the Twitter API (Application programming interface) makes it possible to systematically capture tweets containing a given keyword (hashtag).

(1) Via a quantitative analysis of social media texts on news and current affairs programs in the context of the 2014 elections. Key topics and themes can be captured, as well as changes over time, as a sequential number of television programs will be selected. Next, various user roles can be demarcated.

(2) In a second instance, a more in-depth reading of a selected amount of social media texts will be performed. Of each television program that is analyzed, a subsample of data will be selected to conduct a qualitative content analysis. In correspondence with our theoretical outline concerning audience research from a constructionist view [12], we will study the reception of and reflexivity on television consumption, with a focus on news and current affairs, through social media platforms. We acknowledge social media texts as 'proxies' of the mental constructions people make about television content. Hence, retrospective interviews (phase III) allow triangulating the inferences we made based on our content analysis. The data will allow us to better understand the way television content becomes remediated through social media platforms (RQ1).

3.3 Qualitative research: In-depth interviews

As there is nothing inherent in a text, in a sense that it is always brought by someone, we will be conducting retrospective interviews with social media users to contextualize the prior findings. In addition, in-depth interviews are a suitable method to deepen our understanding of the converging online and offline spaces, advanced by the remediation of television content through social media (RQ2). By hand of the content analysis (phase II), whereby we also collect user identification, we will invite respondents for participation to this research. In total, approximately 50 interviews will be conducted. Respondents will be selected in accordance to the typology we have developed in phase I. Correspondingly, circa 15 interviewees per cluster will be selected for a qualitative interview. This way the qualitative research will enrich the findings of phase I and II.

The concept of the 'moral economy' of the home [19], allows to conceptualize the negotiation of cognitions and values concerning media messages within the household and how the communication with the public world is structured. In addition, Lull's [16] typology on social uses of television will be assessed in the new media ecology as the emergent virtual context reconfigures the interpersonal construction of television time.

4. PRIMARY FINDINGS

In what follows, we shed a first light on the dynamics of television consumption in the new media environment. First, we briefly elaborate on the quantitative survey and secondly we discuss the qualitative follow-up study. This research precedes the methodology outlined above.

(1) Based on the quantitative Digimeter survey, executed in 2010 (N = 1399, 52 % male, 48 % female, M age = 48.24, SD age = 17.62), we got insight in the way television confounds with other activities, more specifically other screen media (i.e. laptop, smartphone, tablet). The survey, representative for the Flemish population, shows that 21% (N = 292) of the respondents fit 'the multi-screen profile'. Here, a multi-screen household denotes the ownership of at least one television, one portable computer/tablet and one smartphone. When comparing the multi-screen respondents with the other respondents, it seems the availability of multiple screens does foster the simultaneous consumption of television and computer/internet technologies [23]. Nonetheless, their presence has no implications for the amount of time television is consumed nor the social context of television viewing, as most respondents report TV to be consumed socially [23].

(2) Consequently, by means of in-depth domestic interviews, with 26 owners of multiple screen technologies (15 males and 11 females, ranging from age 20 to 57), we interpret the integration of multiple secondary screens in the everyday television viewing behavior to shed light on the social dynamics of multi-screen consumption in the living room context. We see that the second screens (e.g. laptops, tablets) are used or are within reach when watching television on a daily base. In most cases, their use is related to content *other* family members are watching on the television screen. Hence, we notice a privatization of media consumption in the living room. As family members remain physically close, but mentally separated, the living room becomes a place where (screen) media technologies are consumed 'alone together'.

Nonetheless, we also found evidence of changing dynamics concerning public and private spaces, as people extend television texts on their secondary screens into online social spaces or more generally, the Internet. In most of the cases, respondents report making comments on television content (news, current affairs & reality TV) via Twitter or Facebook, *without* sharing these reflections with family members. Paradoxically, this multi-screen environment simultaneously induces both privatization (individual consumption) as well as socialization processes (online social interaction).

The information gateway provides us avenues to reflect upon television content as well as the media as an institute and the reality claims they make. It allows us to control and manage our consumption as well as our identity. The technologies at hand *collectively* provide us with possibilities concerning individual agency (empowerment), nonetheless we all need to perform these acts *individually*. In this respect, the importance of literacy to be able to benefit from this participatory digital culture, is salient.

5. DISCUSSION

In an increasingly complex media environment, capturing 'the viewer' becomes more and more difficult, both conceptually and methodologically. In our theoretical elaboration, we need to stop thinking about contemporary (media) society in terms of traditional, modern concepts [2]. The major purpose of the project

is to overcome some of these conventional theoretical and empirical approaches in the field of media sociology and afford new contributions.

We elaborate on some of the many 'hybrids' emblematic for this new media ecology. More specifically, we acknowledge the blurring boundaries between the public and the private, the offline and the online and mass versus interpersonal communication. In doing so, analysis of meso/macro socio-cultural activities (i.e. remediation of television content via online social media) is supplemented with a micro-analysis of domestic (offline) practices. As a result, we theoretically and empirically overcome some of our binary conceptions and recognize the many challenges and paradoxes, characteristic for contemporary (media) life.

We contextualize television in the complexity of the everyday, into which media are enmeshed. In doing so, we aim to complement the empirical richness, with a reflexive and critical approach, both on the part of the researcher, as well as on the part of the audience. The execution of project itself is a dynamic process in a rapidly changing environment, requiring us to be pragmatic and eclectic in our theoretical and empirical choices.

6. ACKNOWLEDGMENTS

The author would like to thank IBBT - i.Lab.o for sharing the data from the Digimeter survey, wave 3, 2010.

7. REFERENCES

[1] Bauman, Z., Liquid modernity2000, Cambridge: Polity Press.

[2] Beck, U., World Risk Society1999, Cambridge: Polity Press.

[3] Silverstone, R., Media and mortality. On the rise of the mediapolis2007, Cambridge: Polity Press.

[4] Deuze, M., P. Blank, and L. Speers, A life lived in media. digital humanities quarterly, 2012. 6(1).

[5] Abercrombie, N. and B. Longhurst, Audiences1998, London: Sage Publications.

[6] Livingstone, S., The challenge of changing audiences: Or, what is the audience researcher to do in the age of the internet? European Journal of Communication, 2004. 19(75): p. 75-86.

[7] Castells, M., Communication Power2009, Oxford: Oxford University Press

[8] Jenkins, H., Convergence culture: Where old and new media collide2006, New York: New York University Press.

[9] Silverstone, R., The sociology of mediation and communication, in The sage handbook of sociology, C. Calhoun, C. Rojek, and B. Turner, Editors. 2005, Sage Publications: London. p. 188-207.

[10] Fiske, J., Television culture1987, London: Metheun.

[11] Bolter, J.D. and R. Grusin, Remediation: Understanding new media2000, Cambridge: MIT Press.

[12] Alasuutari, P., Rethinking the media audience: the new agenda1999, London: Sage Publications.

[13] Ang, I., Watching Dallas: Soap opera and the melodramatic imagination1985, New York: Methuen

[14] Katz, E. and T. Liebes, Mutual aid in the decoding of Dallas: Preliminary notes form a cross-cultural study, in Television in transition, P. Drummond and R. Patterson, Editors. 1985, British Film Institute: London. p. p. 187-198.

[15] Warner, M., Publics and counterpublics2005, New York: Zone Books.

[16] Lull, J., Inside family viewing: Ethnographic research on television's audiences1990, London: Routledge.

[17] Bakardijeva, M. and R. Smith, The internet in everyday life. Computer networking from the standpoint of the domestic user. New Media and Society, 2001. 3(1): p. 67-83.

[18] Humphreys, L., Social topography in a wireless era: The negotiation of public and private space. Journal of Technical Writing and Communication, 2005. 35: p. 13-17.

[19] Silverstone, R. and E. Hirsch, Consuming technologies. Media and information in domestic spaces1992, London: Routledge.

[20] Papacharissi, Z., A private sphere. Democracy in a digital age. Digital media and society series2010, Campridge: Polity Press.

[21] Chew, C. and G. Eysenbach, Pandemics in the age of Twitter: Content analysis of Tweets during the 2009 H1N1 Outbreak. PLoS ONE, 2010. 5(11).

[22] Bruns, A., et al., Mapping the Australian Networked Public Sphere. Social Science Computer Review, 2011. 29(3): p. 277-287.

[23] D'heer, E., C. Courtois, and S. Paulussen. The dynamics of multi-screen media consumption in the living room context. In: Digital proceedings of Etmaal van de communicatiewetenschap. 2012. Belgium, Leuven.

The Role of Smartphones in Mass Participation TV

Mark Lochrie
School of Computing and Communications
Infolab21, Lancaster University
Lancaster, LA1 4WA, UK
+44 (0) 1524 510537
m.lochrie@lancaster.ac.uk

Paul Coulton
School of Computing and Communications
Infolab21, Lancaster University
Lancaster, LA1 4WA, UK
+44 (0) 1524 510393
p.coulton@lancaster.ac.uk

ABSTRACT
From the early days of television (TV) when viewers sat around one TV usually in their living room, it has always been considered a shared experience. Fast forward to the present day and this same shared experience is still key but viewers no longer have to be in the same room, and we are seeing the dawn of mass participation TV. Whilst many people predicted the demise of live TV viewing with the adoption of Personal Video Recorders (PVRs) it has not materialised. Shows that are watched live are often ones that have a greater social buzz. These shows regularly have viewers discussing what they are watching and what's happening in real-time. This paper focuses on the influence smartphones have on TV viewing, how people are interacting with TV, and considering approaches for extracting sentiment from this discussion to determine if people are enjoying what they are watching.

Categories and Subject Descriptors
H.5.1 Multimedia Information Systems

General Terms
Human Factors

Keywords
Mobile, second screen, interactive, television, Twitter, shared experience, mass participation, smartphone

1. INTRODUCTION
Dating back to the early 1950s, watching TV has always been a social experience, although limited to those in the same room (usually the living room). In 2012 the social experience of watching TV has potentially expanded to include anyone with a data connection, taking it beyond your living room into many other viewing areas.

In the early 1950s there weren't many channels to choose from as a viewer, and in some cases only one. TV shows that were being commissioned only had to be better than their counterparts being broadcast at the same time when competing for viewing figures. It wasn't until the 1960s that TV really took off worldwide, with the introduction of more channels, shows and greater emphasis was on ways to introduce TV to the mainstream. With the explosion of mainstream TV shows, TV guide producers were extending the number of pages to include extra information and advertise TV in a way that hadn't been done before. During this time people primarily discovered programmes through word of mouth (water cooler moments) and through the traditional TV guides. Now we have 100s of channels, 1000s of TV programmes. Broadcasters have to invest in new methods of engagement to attract viewers. Many are now engaging with the public through social media. We have seen the dramatic rise of social media services such as Facebook, Twitter and other services that link into existing social networks, being utilised to create forums for debate around a range of topics including TV shows. While Facebook is being used for its functionality of branding and approval systems - Twitter, with its ability to share topics through '*hashtags*' and '*re-tweets*' with anyone, TV audiences are now using such methods to communicate in almost real-time.

Furthermore PVR ownership has had a profound impact on the when/why/how we consume our entertainment. We discover programmes very differently these days, whether from social recommendations (water cool moments, an online advertisement or social media), browsing through the interactive TV guides (from set top boxes to mobile applications), to seeing a clip on TV of an up and coming programme, that invites us to schedule the recording/notification of the programme or the entire series (usually from viewer interaction by pressing the red button, which sets up the PVR to record or notify when the programme is broadcasted). The majority of people that use a PVR to time shift their entertainment, usually catch up later the same day (so they still have the ability to join in the water cooler moment at work the next day), some prefer to watch 10 minutes behind time to skip forward past the commercials and usually those programmes that aren't watched the same day are normally ones that don't have the requirement to be consumed right away, whether its from lack of interest to little social buzz, these are usually consumed later in the week [2, 5, 6].

Since the introduction of such time-shifting devices and their ever growing popularity within our households, the current trends seen in what shows are more likely to be time shifted are usually scripted for genres such as sci-fi, sitcoms and dramas, whereas we still see the need to watch live programmes such as sporting events and others alike, generally because the live audience wants to be part of something bigger, similar to a crowds participation in the stadium of a football match.

2. BACKGROUND TO STUDY
This constant emerging desire to converse, share and interact around TV shows isn't going away, we are seeing more shows attempt to integrate social media into the shows plot, narrative and format. Many social network platforms for instance Twitter allows social participation and discovery with the popularity of trending topics. Watching live broadcasted TV will always create a greater buzz than pre-recorded shows. It is these types of shows that the real-time advantages of social media (in particular Twitter) works well with, as these events are still generally viewed in real-time rather than on time-shifted devices [4, 5, 6].

More recently Twitter introduced new ways to discover and engage with current trending topics, in particular on mobile devices, through its "Twitter Discovery" service. Twitter is seeing how people interact with information differently from desktop to mobile devices. One way in which Twitter is improving its mobile experience is through ways in which people connect (follow) and interact. This is achieved by displaying prepopulate tweet

windows for *hashtags* and *retweets.* Twitter has also focussed on the discovery of information, by displaying real-time trends with hash information and an external article (that the trend refers to). Studies from Twitter [7] suggest that when broadcasters combine the real-time elements of Twitter, there is a direct and immediate increase of viewer engagement, anywhere from two to ten times more the amount of *mentions, follows* and *hashtags* used whilst the show airs. This is highlighted when you consider the 2010 Grammy Awards, which saw a 35% increase on viewing figures from 2009, one of the reasons for this increase, is suggested to be the integration of social media in the 2010 event.

It is becoming apparent that social media is having a significant impact on what and how we watch TV. Studies in social TV trends for the UK show that 17% of viewers will watch a TV programme based on influences from social media, this number rising to 39% when considering the main demographic (18-24 year olds) that are likely to adopt such technology[1]. The insights already seen in social TV has provided Channel 4 (UK free to view channel) the opportunity to launch a new social media based catch up TV channel. The channel aims to rebroadcast programmes over the last seven days based around their social buzz [2]. In the past, TV shows rating and viewing figure were obtained from television measurement organisations such as Neilson (US) and Barb (UK) using electronic metering technologies and census data. Recent studies by Neilson show a direct connection between traditional TV rating and social buzz.

Like television measurement organisations, TV shows can take advantage of social media buzz to predict and analyse viewer engagement through gaining sentiment from users interactions. Although the ability to derive sentiment analysis from Facebook statuses and Tweets is possible, its accuracy is open for debate, for example; studies into average lengths of tweets by Isaac Hepworth, indicated that users who tweet from the 'desktop' web client are more likely to write more content and use the full 140 characters, whereas those who tweet on their mobile client average around 30 characters per tweets. Therefore obtaining sentiment analysis in particular tweets from mobile devices is inaccurate [1, 3]. Twitter's API does provide a simple approach to this by using Emoticons (happy/sad faces depicted by punctuations i.e. ☺ ☹), however this method for detecting emotion in *tweets* is limited, as the majority of *tweets* composed do not contain such, this is confirmed in the *tweets* captured in this study.

In view of these limitations and to gain better understanding of how people use social media whilst watching TV a system was developed to record and analyse in real-time all the tweets associated with popular TV shows. The TV shows that we analysed included a range of show genres including panel shows (X-Factor, Strictly Come Dancing, etc.), reality shows (The Apprentice), award ceremonies (The Oscars and The Grammys), sporting events (Super Bowl 2010 & 2011, World Cup 2010, Wimbledon 2011, etc.), live news events (The Royal Wedding), scripted dramas (Homeland) and also the introduction of a new channel (Sky Sports F1).

For the purpose of this paper we will focus on the 2012 Super Bowl. The Super Bowl more than doubled its previous years social buzz, reaching a record breaking 12,200,000 *tweets* during broadcast. With this need to discuss, search for contextual information, engage and gain a better experience, we are seeing an enormous increase in traffic toward "social TV". When we consider a live sporting event such as the Super Bowl and compare it to something more globally watched like the Grammys, there is only a 6% difference in the number of *tweets* recorded at the time of broadcast. Whereas dedicated second screen apps for both events saw the Super Bowl doubling the amount of unique users the Grammy app had, this is more likely due to the fact that sports fans want more real-time statistics and in play tactics, formations and player ratings.

3. THE STUDY

To enable us to perform this study on each show we needed to capture and analyse people's public *tweet* data. The process involved capturing tweets from twitter.com, using its Streaming API. The system can stream all tweets that contain a certain *hashtag* in its raw form. The *tweet* data is then parsed, split and sorted into different tables (tweet data, mentions, tags, urls and users). This allows a subsequent deeper analysis of the *tweets* content, including its source (to determine if the tweet was sent from a mobile device), the tags used and who tweeted it.

For the purpose of this paper we will highlight the findings of a live sporting event (Super Bowl 2012), for which the system collected the *tweet* data relating to the hashtag #SuperBowl and #SuperBowl46 in real-time. The tweet data presented in this paper was captured on Sunday, February 5, 2012 at 6:30 pm EST till 11:00pm.

The *tweet* source data was analysed and grouped by platform in order to determine if the *tweet* originated from a mobile device. During the collection period 1,802 different clients were used to compose a *tweet*. In order to establish the platform used to *tweet,* the data had to be reclassified. This was achieved by analysing each client and arranging into either mobile, non mobile or mixed. Due to limitations of Twitters metadata, which details the agent used rather than the exact client, a mixed category was adopted for clients that are on multiple platforms. For example, 'TweetDeck', as the agent has many versions available the *tweet* could have originated from either desktop, mobile, browser app or web, therefore it is classified as *MIXED.*

Figure 1 shows the most popular clients used to *tweet* during the 4.5-hour period. The majority of *tweets* were sent from Twitter's dedicated services such as their website, iPhone, Android and BlackBerry branded clients. These findings coincide with similar client usage studies. Sysomos [//sysomos.com], found that 58% of *tweets* originated from official Twitter clients, web being the most popular with 35.4%, iPhone, BlackBerry, m.twitter and Android following behind. In order to fully understand how people interact whilst watching TV, we first needed to analyse the average amount of characters being composed over all platforms.

[1] Dissfusion.com
http://www.diffusionpr.com

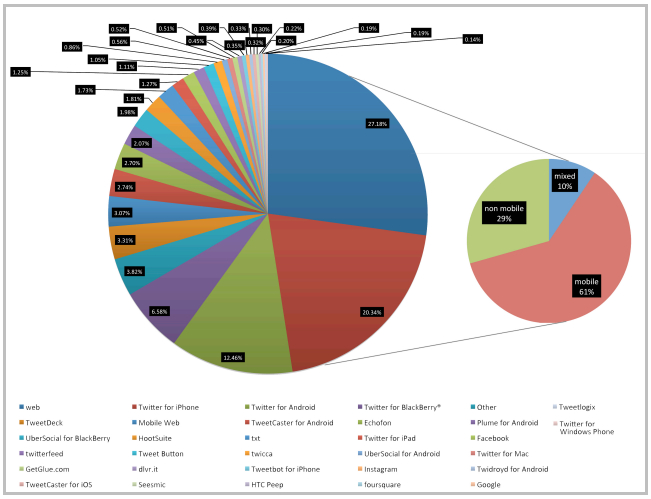


Figure 1. Tweet overview sources breakdown and platform classification, Super Bowl 05.02.2012 18:30 - 23:00

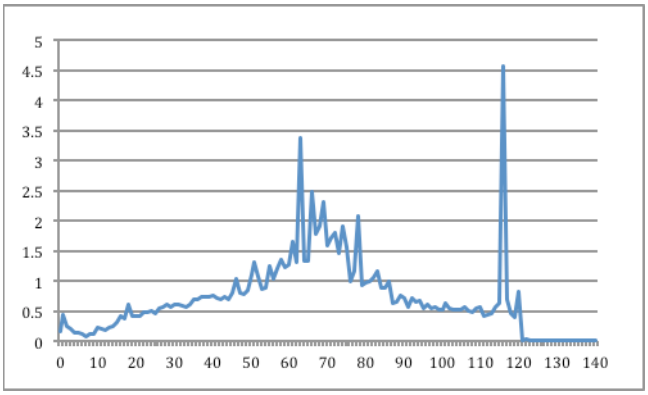


Figure 2. Percentage of characters per tweet for all sources, with no RT and URL removed, Super Bowl 05.02.2012 18:30 - 23:00

In the first instance it became apparent the majority of *tweets* 6.6% used the full limit of 140 characters. However when analysing *tweets* for sentiment its becomes obvious that the majority of *tweets* with 140 characters are filled up with URL's, *hashtags* and *RT*. When attempting to determine sentiment from *tweets*, we needed to establish how many characters are used in the majority of *tweets*. Figure 2 demonstrates the percentage of all tweets that haven't been *retweeted* and any URL removed. This time the trends have changed, it is apparent there is a shift in the amount of characters being used, the majority of these tweet contain anywhere between 60 – 120 characters. Also it is noticeable the number of 140 character *tweets* has decreased.

The results shown in Figure 2 do not consider *tweets* solely from a mobile device. Figure 3 demonstrates *tweet* with the URL removed and that weren't *retweeted* from the most popular mobile devices. It is clear from looking at Figure 2 and 3, that the percentage of characters differs dramatically depending on platform. The most popular mobile clients used during the Super Bowl 2012 were Twitter for; iPhone, iPad, Android, BlackBerry and Mobile Web, SMS, TweetCaster, Echofon, Plume for Android and UberSocial for; Android and BlackBerry. As you can see from Figure 3 the bulk of *tweets* composed were at the lower end of the spectrum, around 10 – 80 characters. *Tweets* that were less than 10 characters were statuses that referred to the URL, which was removed. The typing constraints on a mobile device and the nature of the show in question (people want to share their opinion, as its happening) have an impact on differences between mobile and desktop platforms. This is one issue when attempting to extract sentiment from tweets where the majority are sent from mobile devices (61% of Super Bowl related tweets originating from a mobile device, as shown in Figure 1), and this isn't decreasing any time soon, especially around TV shows with second screen experiences.

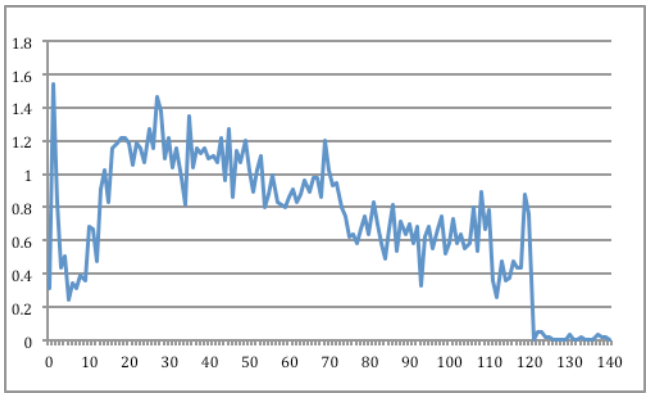


Figure 3. Percentage of characters per tweet for the most popular mobile clients that haven't been retweeted and had the URL removed, Super Bowl 05.02.2012 18:30 - 23:00

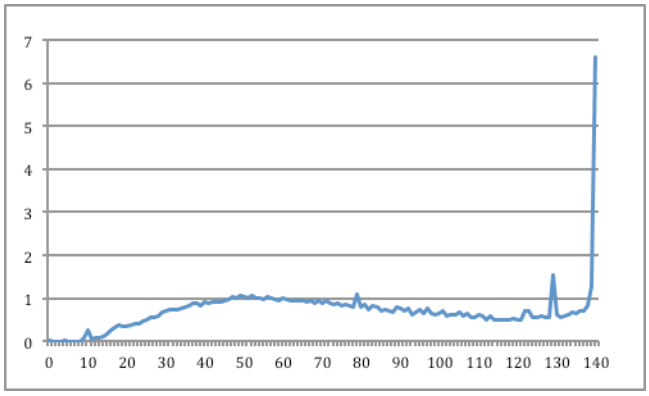


Figure 4. Percentage of characters per tweet for the most popular mobile clients, Super Bowl 05.02.2012 18:30 - 23:00

As the majority of *tweets* that contain the full 140 characters mainly contain URL's, *RT's* and *hashtags* it makes extracting sentiment from small pieces of information inaccurate (Figures 3 and 4). Nevertheless we are seeing TV shows like 'Homeland' involve viewers using *Twitter's hashtag* mechanism to gauge viewer consensus, by engaging with its audience to see if they are following the plot (this enables the show to fully understand if they deceiving viewers with their story). The show is achieving this by prompting the question "friend" or "foe" (#friendorfoe) at the end of each episode. This gives a clear and concise (Boolean like) understanding whether the viewers are following the story, therefore viewers that don't particularly watch or enjoy the show are unlikely to tweet the *hashtag*. Similarly on Facebook: 'Likes' used to gain sentiment with its Boolean value, users can 'Like' brands, shows, posts etc.

Otherwise in order to fully analyse the emotion over small bursts of information, we first need to study the users behaviour over a period of time, to gain a better understand of the language they use this would provide an improved baseline into determining sentiment from short pieces of information. The problems with

this approach would the time it takes to perform such operation and the amount of data it would require.

4. THE FUTURE OF SECOND SCREENS BEYOND SIMPLE ANALYTICS

Already some shows are using second screens to get real-time data to integrate into the show. This is seen in Dancing on Ice 2012, where viewers can score the skaters on their performances, share their scores and opinions with their social network friends, rate the judges and catch-up on achieved video highlights. The applications data is integrated with the live show, as each judge scores the performances the presenters compare those scores from the judges with an average from the public consensus. Although the viewers' scores have no real impact on the actual scoreboard (who essentially is in the bottom two), this could soon be an integral part to shows alike where viewers are the nth judge. Similarly seen on Homeland, Britain's Got Talent 2012 audition phase, are flashing *hashtags* for each act when the performer takes to the stage. This is a good way of engaging the audience with each act on a show like this, similarly to the way in which, where a shows format involves phone votes and SMS to determine a leaderboard.

There are many different types of second screen applications. Some are built for a specific show; others are for a more general watching of TV. Each year we are seeing a rise of specific second screen applications, typically for shows mentioned in this paper (panel and reality shows and sporting events). Zeebox [//zeebox.com] has taken a slightly different approach to the second screen market. Zeebox offers a white label second screen application that allows different TV shows to build upon the platform with specific show related functionalities.

5. CONCLUSIONS

The traditional viewing environment, where friends and family sit around, the one TV in the living room to consume their entertainment, has considerably changed over time. No more so than today, where we are witnessing broadcasters invent new ways to watch TV, from smartphones to tablets, laptops to TVs. As the traditional TV medium is essentially a shared experience, simply overlaying personal *tweets* on screen isn't a shared experience. Essentially this would mean users would have to opt in to share their social streams with everyone else in the living room. Not to mention this would require the viewer splitting their attention away from the core element, whereas utilising the ubiquitous smartphone/tablet allows viewers to focus their attention at once place at a time.

With recent high profile launches and decreasing prices, we are seeing tablets replace laptops in the home because of their ease of use, fast boot up, size, convenience, lightweight and mobility. What is clear is the ways in which we interact with these devices differ from smartphones. The tablet is a device in which we are more likely to share with others (43% share with others), mainly within families, whereas a smartphone is considered a more personal interaction. The other main difference is how the two devices are interacted with. Tablet users often interact and hold the devices differently (A tablet is usually held horizontal to the ground thus sharing the experience, whereas a smartphone is a more closed off experience as these devices are held vertical). All of which constitutes different approaches when designing social TV experience applications.

The real opportunity for second screen applications are when they fully integrate the viewer into the plot or narrative of the show, therefore connecting the viewer into the format of the show. When incorporating a second screen application into a shows format the linearity of the programme needs to be core. It is the growth in social network consumption, broadband availability, and the on going sales of smartphones and tablets that is driving social TV.

Whilst this study included *tweets* from a variety of mediums the information obtained about the clients used to compose the *tweets* indicates over 60% are from mobile, which is consistent with the figures reported by Twitter. Furthermore as reported, widely in the media, smartphone manufacturers are shipping more devices year after year outselling PC units worldwide. Not to mentioned Apple's iPad outselling their Mac series for the first time (June 2011). This trend is likely to continue throughout the industry amongst other manufacturers, thus the amount of connected devices to Internet services through some form of mobile is likely to increase dramatically in the near future.

Overall this study highlights that mobile phones are already becoming the second screen for TV but not through broadcaster provision of personalised services, or service providers enabling them to act as a new form of remote, but rather by audiences themselves creating their own forums for inter-audience interaction. It is therefore important for broadcaster and producers to be able to better understand the nature of this interaction otherwise TV itself may become the second screen.

6. ACKNOWLEDGMENTS

Many thanks to Lancaster University Network Services (LUNS) for providing a high bandwidth connection used to stream tweets from Twitter, for data consumed in this study.

7. REFERENCES

[1] Lochrie, M. & Coulton, P., 'Tweeting with the telly on!: Mobile Phones as Second Screen for TV', Paper presented at Social Networks and Television: Toward Connected and Social Experiences at IEEE Consumer Communications & Networking Conference, Las Vegas, United States, 14/01/12 - 17/01/12.

[2] Reuters, New TV and social media trend among the youth, http://uk.reuters.com/article/2011/03/08/uk-digital-social-tv-idUKTRE7275RZ20110308, last accessed on 08/06/2011.

[3] Simm, W.; Ferrario, M.-A.; Piao, S.; Whittle, J.; Rayson, P.; , "Classification of Short Text Comments by Sentiment and Actionability for VoiceYourView," Social Computing (SocialCom), 2010 IEEE Second International Conference on , vol., no., pp.552-557, 20-22 Aug. 2010.

[4] Social media chatter ups live TV stats,. Media Post, http://www.mediapost.com/publications/article/170743/social-media-chatter-ups-live-tv-stats.html, last accessed on 30/03/2012.

[5] The lazy medium,. The economist, http://www.economist.com/node/15980817, last accessed on 30/03/2012.

[6] The schedule dominates, still,. Deloitte, https://www.deloitte.com/assets/Dcom-Global/Local%20Content/Articles/TMT/TMT%20Predictions%202012/16470A%20Schedule%20lb1.pdf, last accessed on 30/03/2012.

[7] Twitter, Watching together Twitter and TV, http://blog.twitter.com/2011/05/watching-together-twitter-and- tv.html, last accessed on 25/03/2012.

Transmedia Narrative in Brazilian Fiction Television

LETICIA CAPANEMA
PhD student at Pontifícia Universidade Católica de São Paulo
Rua Monte Alegre, 984, Perdizes
São Paulo – SP I CEP: 05014-901
55-11- 3670-8000
capanema.leticia@gmail.com

ABSTRACT
This project aims to investigate the phenomenon of transmedia storytelling on brazilian fiction television, through the analysis of narrative processes of brazilian soap opera. It starts with the observation of some changes occurring in the media environment, as the convergence of content, the divergence of platforms, strengthening the participatory nature of the public, the hybridization and hypermediation of traditional media. In this context, there is constant feedback between television and new technologies, resulting in the expansion of the television narrative for several other medias. Thus, assuming that the way in which we tell stories is related to the limits and possibilities of language, this research investigates the challenges changes that occur in the narrative format of television fiction.

Categories and Subject Descriptors
H.5.1 [Information Interfaces And Presentation]: Multimedia Information Systems – animations, artificial, augmented, and virtual realities, audio input/output, evaluation/methodology, hypertext navigation and maps, video.

General Terms
Experimentation, Human Factors, Languages, Theory.

Keywords
TV Fiction; Transmedia Storytelling; Expanded Narrative; Brazilian Soup Opera.

1. INTRODUCTION
After some years of utopian and dystopian speculations on the future of media in the digital environment, it is observed that the predicted digital convergence did not culminated in the merger of various communication devices on a single machine. After all, the number of devices currently used for communication has not decreased, otherwise, increased. There is a growing number of technological devices - television, mobile phones, video games, computer tablets - all still far from converging into a single machine.

This misconception of media convergence is called by Henry Jenkins (2009) the *Black Box Fallacy*. According to the author, the expression means that the content will converge into a single black box in our living room. For Jenkins, part of what makes the concept a fallacy is that it reduces the transformation of the media to a technological change, and leave aside the cultural levels. (Jenkins 2009, p. 42). Today, it is observed that the convergence of communication technologies in a single unit is not consummated, however we can see another kind of convergence: the cultural one.

The *convergence culture*, title of Henry Jenkins's book (2009), is characterized by participatory culture, collective intelligence and media convergence. Convergence and divergence are two sides of the same phenomenon. The hardware diverges while the content converges into any communication platform. (Jenkins 2009, p 43).

Television responds to the phenomenon of convergence culture through the expansion of its contents in various media. In this context, the television fiction territory is suitable for renewal. As stated by Newton Cannito:

"The Digital has brought something that nobody expected: television has become more narrative. The screenplay for television series has never been so narrative and interconnected. The presence of good writers has become critical. The power has shifted into the hands of storytellers". (Cannito, 2010, p.18)[1]

There is a movement in academic investigations of television in order to understand the relationships that television sets with other media, especially the internet. In fact, "the extension of television narratives to new technologies is considered a major driver of the renewal of television fiction[2]" (Lacalle 2010, p. 82). And this process introduces important cultural transformations, such as strengthening the participatory nature of the public, the phenomenon of ownership of media content, stress the playful aspect of the narratives, as well as its complexity and expansion.

From this synergy between television and other media platforms, new forms of interaction and narrative models emerge, resulting in a phenomenon called *transmedia* (Jenkins, 2009). According to Jenkins, *transmedia storytelling* unfolds across multiple media platforms, where each new content contributes differently and valuable to the whole (Jenkins, 2009, p. 138). As pointed out by Scolari, the same phenomenon is also described by other investigators using different concepts: *cross media* (Bechmann Petersen, 2006), *multiple platforms* (Jeffery-Poulter, 2003), *hybrid media* (Boumans, 2004), *intertextual commodity* (Marshall, 2004), *transmedial worlds* (Klastrup y Tosca, 2004), *among others* (SCOLARI, 2010, p.189).

In the television industry, we can observe some experiences with the expansion of narratives. The television series *Lost* can be

[1] This text is a free translation of the author.

[2] This text is a free translation of the author.

considered as an example of successful experience of television transmediation. The 121 episodes of *Lost* were displayed on 6 seasons during the years 2004 and 2010. In addition to television episodes shown in the television grid, the narrative of the series was explored through books, websites, blogs, documentaries, television advertising, etc. All of them created by the producers of *Lost,* and many also produced by fans. Thus, two aspects stand out in the *Lost* experience: the expansion of narrative to other communication platforms; the mix between the production of fans and official producers of the series.

In Brazil it is possible to observe, over the past five years, some experiments in order to expand the soap operas narratives beyond the television platform. In 2009, in an effort still timid, *Rede Globo,* the most important brazilian television network, created the blog of the character *Luciana* from the soap opera *Viver a Vida*. The narrative universe of the character began to be exploited also in the internet through the blog, which functioned as a diary of the girl. *Rede Globo* soap operas that followed this, began to increasingly use the internet as a platform for expansion of the plot, adding other exclusive content such as videos, games, audiocasts, among others.

The period identified by some authors as the *post-television* (MISSIKA, 2006) representes more a moment for experimentation than actually the end of television. Television matures her language after each transformation, as seen before with the emergence of the VCR, the remote control, the dialogue with the Internet and the videogames. "Technological change makes television more television" (Cannito, 2010, p.49).[3]

2. A NEW KIND OF NARRATIVE?

It is possible to recognize cultural and semiotic changes in the ways stories are told through the current media. These changes are related to media and cultural environment, which is characterized by some aspects: the convergence of content and divergence of media platforms; the emergence of a participatory culture of the new generation; and the absorption of new technologies by traditional media.

Given this scenario, some paths are open: the narratives become more complex, which create stories that spread through a net of possibilities; the participatory nature of the public become more evident in the process of unfolding narrative. Janet Murray characterizes the current media environment by the lost of boundaries that separate the human expression forms:

We are on the threshold of a historic convergence in which novelists, screenwriters and filmmakers move toward multiform stories and digital formats; computer scientists begin to create fictional worlds, and the audience goes towards the virtual stage. How to figure out what comes next? Judging by the current situation, we can expect a continuing weakening of the boundaries between games and stories, films and simulation tours, between broadcast media (such as radio and television) and archival media (such as books and videotapes), between narrative forms (such as books) and dramatic forms (such as theater or cinema), and even between the audience and the author". (Murray: 2003, 71-72).

[3] This text is a free translation of the author.

In fact, television adapts to new technologies. This adaptation process is propelled in fiction television, genre that easily assimilates the phenomenon of media feedback. Coined by Henry Jenkins (2009) as *transmedia storytelling*, the narrative expansion in various media platforms is a factor that characterizes the current phase of television fiction. The marriage between television and new technologies provides territory for the transmediated construction of narratives (Lacalle, 2010 p 82). Currently, we observed several experiments on television industry to exploit the expansion of multi-platform narratives.

By studying these experiences, this project aims to find answers to the following question: **given the current media environment, characterized by the culture of convergence, transmediation and public participation, what are the transformations that occur in the fiction television narratives?**

To answer this question, we intend to investigate fiction television products in order to observe how the contemporary narrative structures change. This research will investigate the phenomenon of transmedia storytelling on brazilian sopa operas, through the analysis of current processes of fiction television narrative and its expansion into other media platforms.

The research assumes that the way in which we tell stories is related to the limits and possibilities of language developed so far. As in film history, which begins exploring simple narrative through a cinematic language poorly developed (without the use of the edition, for example), we can infer that the television adaptation of the logic of hybridization, convergence and transmedia results in new forms of narrative television.

This research is justified by enrolling in the Program in Communication and Semiotics at PUCSP[4], especially in the research area "Analysis of the Media". It is also a relevant study to understand the direction of television language and contemporary audiovisual narrative.

3. AIMS AND OBJECTIVES

This research aims to investigate the phenomenon of transmedia storytelling on television through the analysis of current processes of narrative fiction television, particularly in brazilian soup operas, and its expansion into other media platforms.

Also this project has specific objectives:

Develop theoretical basis on the settings of contemporary communication;

Review the literature on theories of television fiction;

Study deeply the expanded narrative and transmedia storytelling on television;

Select for analysis brazilian soap operas that use the expansion of narrative;

Investigate the form of interaction and immersion of the viewer / interactor with the selected works;

[4] Catholic University of São Paulo. "Communication and Semiotics" post-graduation program.

Analyze the relationship between transmedia narratives and game logic;

4. IN SEARCH OF A METHODOLOGY

Roland Barthes, french critic who carried out important studies of the narrative, states that it "is present at all times, everywhere, in all societies, begins with the history of mankind. (...) Does it come from the genius of the narrator or does it have a narrative structure in common which is accessible to analysis?" (Barthes, 2009, p. 19-20). [5]Barthes begins his researches from the perspective of structuralist studies, which aims to identify common structural features to any narrative. In his final work, the author shifts his studies for a post-structuralist approach, which extends the theory of narrative using methods of structural linguistics and anthropology. Thus, to the poststructuralist authors, the sense of the narrative is built both by the author and the reader. The deconstruction of the narrative emphasizes the role of a subject (reader, listener, viewer) in the process of semiosis, in other words,the interpretation of meanings.

More than ever, the audience plays a fundamental role in the process of narrative signification. The way the stories are told and the way we have access to their narrative elements characterize a new phase of communication. This phase is marked by the phenomenon of integration in various communication platforms and the participatory nature of the viewer. Besides, this study starts from the post-structuralist Roland Barthes to understand the role of the audience in the process of (de) construction of transmedia narratives. Based on the concept of *interactor*, developed by Janet Murray (2003) and extended by Arlindo Machado (2007), the research will explore the role of the subject (before viewer, listener or reader) who starts to act on the communication process. This interactor is required to make decisions and invited to participate actively interfering in the process.

The concept of media reformulation - remediation - developed by Jay Bolter and Richard Grusin (2000), is an important baseline to contextualize the current moment of communication, evidenced by the phenomena of convergence, hybridization and hypermediation. The characteristics of new media defined by Lev Manovich (2001), as well as those identified by Janet Murray (2003) are also important.

Manovich, in his book *The Language of New Media* (2001), proposes the abandonment of old categories and establishes new, derived from computational logic, in order to visualize the topology of the new media:

“Because new media is created on computers, distributed via computers, and store and archived on computers, the logic of a computer can be expected to significantly influence the traditional cultural logic of media; that is, we may expect the computer layer That Will Affect the cultural layer”. (Manovich, Lev, 2001, p 46).

From this perspective, the author seeks to identify a language to the new media. Manovich elects five aspects considered relevant to the digital condition - numerical representation, modularity, automation, variability and transcoding.

[5] This text is a free translation of the author.

Janet Murray identifies the following properties of the new media: procedural, participatory, spatial and encyclopedic. For the author these four aspects are interrelated. In other words, digital systems are based on description process; react to information that are inserted and require decisions and actions of its users. In that way, they are procedural and participatory. Digital systems make cyberspace seem so vast and rich as the physical world, so they are spatial, and have an unrivaled ability to store information. By situating the narrative process in the era of new media, the researcher states that a maze of adventure is especially suited to the digital environment, because the story is tied to the navigation of space (Murray, 2003, 131). In this sense, transmedia narrative deals with the interactive audience. As the story progresses, it acquires a feeling of great power, influencing significantly the events, which is directly related to the pleasure we feel with the unfolding of the narrative (Murray, 2003, p. 131).

Entering in the study of television narrative, the analysis will be based on the relationship between television and the Internet, especially those established in television fiction explored by Charo Lacalle (2010). For the author, in fact, the extent of television narratives to new technologies is considered a major driver of the renewal of television fiction (Lacalle, 2010 p. 82). Stories, that were restricted to the limitations of televisual format, begin a process of peripheral leakage to other platforms, such as internet and mobile phones, producing so extensive narratives that can not be contained in a single medium (Jenkins 2006, p. 95)

Finally, the research also adopts the concept of *trasmedia storytelling*, developed by Henry Jenkins in his book *Convergence Culture* (2009). According to the author, the characteristics of transmedia narratives refers to a new aesthetic that emerged primarily in response to media convergence - an aesthetic that makes new demands on consumers and depends on the active participation of communities of knowledge. (Jenkins , 2009, p. 49).

5. INTERACTIVE OPERA

To perform the analyzes, this research focuses on brazilian television fiction format that has the highest penetration in the country: the soap operas.

Soap operas have great importance in brazilian television culture. Based on the serialized structure of linear and daily exibition, the soap opera is considered the television ficition format most successful in Brazil. Since the last five years, brazilian broadcasters have explored the transmedia narrative of soap operas into other platforms, such as the Internet and mobile. This process is characterized by the complexity of narratives and also by the participatory nature of the audience. The viewers / interactors are called to interact with other platforms to complement the narrative followed on television. The narrative of brazilian soap operas exceeded the limits of space and time stipulated by the television schedule and have spread to others plataforms that are increasingly accessed by curious and active audiences.

6. THE RESEARCH STATUS

This research began in January 2012, at which time the project *Transmedia Narrative in Brazilian Television Fiction* was approved by the Pos Graduate Program in Communication and Semiotics at PUC São Paulo. But the interest in study of narrative

fiction television began during the preparation of my master`s degree *Television in Cyberspace*. The survey, conducted in the Pos Graduate Program in Communication and Semiotics (PUCSP) and completed in March 2009 under the guidance of prof. Dr. Arlindo Machado, was based on research demonstrations in digital television, such as Digital TV and webtvs. From the analysis of several examples of the reconfiguration of television in digital environments, the study identified characteristics of television to the stage in question. Among them, there is the playful and participatory television aspect, as evidenced by the fiction television programs produced in the mold of ARGs - Alternate reality games - and the experiences of expansion of narrative content in other media platforms.

Therefore, this PhD project proposes the extension of the master's research, through analysis of narrative fiction television, to investigate the expansion of the narrative elements to other devices such as internet and mobile, as well as the participatory nature of audience interaction. We believe in the relevance of this search as a study to understanding the direction of television language and contemporary audiovisual narrative.

We are in the first phase of research that corresponds to the literature review of theories dealing with the current panorama of communication, television narratives and the phenomenon of transmedia storytelling. In addition, we are developing the state of the art of transmedia experiences performed on television, and later, we are going to select the brazilian soap operas that will be analyzed.

7. REFERENCES

[1] Barthes, Roland. 2009. *Análise estrutural da narrativa:pesquisas semiologicas*. Vozes, Petrópolis, RJ.

[2] Barthes, Roland. 1980. *S/Z*. Edições 70, Lisboa.

[3] Bolter, J. Davis. Grusin, Richard. 2000. *Remediation. Understanding hybrid media*. MIT Press, Cambridge, MA.

[4] Cannito, Newton. 2010. *A Televisão na Era Digital. Interatividade, convergiencia e novos modelos de negócios*. Summus, São Paulo, SP.

[5] Gilder, George. 1990. *Life after television*. Whittle Books, Knoxvill.

[6] Jenkins, Henry. 2009. *Cultura da Convergência*. Aleph, São Paulo, SP.

[7] Lacalle, Charo. 2010. *As novas narrativas de ficção televisiva e a internet.* Revista Matrizes, São Paulo, SP.

[8] Lopes, Maria I. V. 2011. *Ficção televisiva transmidiática no brasil: plataformas, convergencia, comunidades virtuais*. Sulina, São Paulo, SP.

[9] Machado, Arlindo. 2007. *O sujeito na tela. Modos de enunciação no cinema e no ciberespaço*. Paulus, São Paulo, SP.

[10] Manovich, Lev. 2001. *The language of new media*. MIT Press, Cambridge, MA.

[11] Missika, Jean-Louis. 2006. *La fin de la television*. Seuil, Paris.

[12] Murray, Janet. 2003. *Hamelet no Holodeck. O futuro da narrativa no ciberespaço.* Unesp, São Paulo, SP.

[13] Scolari, Carlos A. 2009. *Ecologia de la hipertelevisión*. Sulina, Porto Alegre, RS.

The 'State of Art' at Online Video Issue

[notes for the 'Audiovisual Rhetoric 2.0' and its application under development]

Milena Szafir

PPGMPA – ECA – USP
[CAPES financial support]
São Paulo-SP, BRAZIL

milena@manifesto21.tv

ABSTRACT

Nowadays it has become clear that processes of digitalization and convergence have transferred audiovisual production from the television screen to the Internet environment. How can one use a video as a reference to be remixed without having to resort to complicated (and heavy) non-linear video applications? Or, inside a Found Footage perspective, let us use the enormous available database through online video platforms! As a consequence of this current state of affairs, open issues concerning this topic will be here discussed and analyzed.

Consequently, my research is based on a alternative solution for the daily breaking of the rule by millions of users, i.e. the audiovisual copyright infringement. Audiovisual platforms are the online places where all of us look for videographic material; the theme of my PhD research is how ordinary people can construct what I have called *Audiovisual Rhetorics*.

> *"YouTube seems to provide an inexhaustible content generated by its users. But this very abundance (McCracken, 1998) might discourage us to question materials which are not found there. Historically the DIY movement aimed at enabling groups that had no commercial penetration to tell their own stories. If we admit that YouTube operates with no previous history, then we would be giving up what we have been fighting for and end up with much less than expected"*[1]

This article wishes to show the bases of the online video-library remix editor project, part of my PhD studies. Aiming at organizing the online audiovisual content like quotes, the user will be able to take excerpts from one video and link (mix) it to other ones in a cloud computing app.

Categories and Subject Descriptors

D.2.2 [**Software**]: Design Tools and Techniques – *Computer-aided software engineering (CASE), User Interfaces* | H.5.4 [**Information System**]: Hypertext/Hypermedia (I.7, J.7) – *Theory, User Issues* | J.5 [**Computer Applications**]: Arts and Humanities – *Arts, fine and performing*

General Terms

Documentation, Performance, Design, Experimentation, Theory.

Keywords

Found Footage, Database, YouTube, Vimeo, Video, Live, VJ, Archive, HTML5, interactive-video, Remix

1. INTRODUCTION

The motivation for the app development and its main usability is to discuss how we should design and evaluate interactive audiovisual content platforms, trying to highlight some relevant aspects of remix issues toward a new participatory use of online video platforms like YouTube among other ones.

Online video platform represents a new challenge on how to spread, link and consume audiovisual content that maximizes the internet user experience. During this first decade of XXI century, research often focused on several aspects of this new audiovisual online participatory scheme, but neglected aspects that might be of interest when trying to understand the wishes of a new generation inside a digital culture and its available technologies. A new concept for the content production has been created to be used by ex-spectators.

The academic research practice, in a brief conceptualization, can be thought of as an audiovisual remix practice: we make our research from authors, journals, images, offline (and online) database etc; then we analyze these data, remix and transform them into a 'new' value, a thesis. But the main difference between these two practices is the copyright laws.

> *The existence of footnotes in audiovisual remix practice is quite unusual; we could mention, for example, some films in the 70's and television programs in the 80's (such as Debord's Society of Spectacle and Godard's Histoire(s) du Cinéma, respectively) or even in the 90's Masagão's Nós que aqui estamos por vós esperamos (Brazil) and several digital audiovisual remix works at beginning of our XXI century as Rebirth of a Nation, The Society of Spectacle – digital remix (made by american djs Spooky and Rabbi, respectively) and other ones by groups like GNN (USA), Media Sana or by girls from mm não é confete (Brazil).*[2]

Therefore, my application will be developed bearing in mind a methodological organization applied to the creation of an audiovisual remix practice for academic and educational purposes.

2. RELATED WORKS

(a framework for this short article: an inside view from the PhD research)

This PhD project is divided into two areas, a practical and theoretical one. However, throughout the period of mapping the online video's 'state of art', both research methods are blended, i.e. one doesn't survive without the other.

Therefore, in order to develop the application's prototype (the PhD's practical part) a flowchart image, shown below, was designed to face the possible paths that could be followed; it couldn't exist without a brief contextualization of shared online audiovisual (the existing video platforms beyond YouTube) and the video editor platforms currently online.

At the same time, the "new" video possibilities (and its interactive perspective) have been changed with the HTML5 forms and its <video> element – which allows to include video directly in the webpages without the need to install a specific plugin to watch it –, i.e. the importance of the native video support in browsers inside an open source code should be utilized, with its associated APIs, to design different "ways in which video can be controlled via JavaScript"[1] in a direct competition to other technologies for Web applications as Flash (the most used one).

Finally, my PhD researches aims to understand the "found footage" issue within the audiovisual theory as part of the digital remix practices and both together as a path to propose a more participatory way – a non-linear video production – for ordinary people inside the interactive television and so called "online revolution" (the huge amount of data can be used as references to creative purposes).

3. Two Different kinds of Audiovisual Online Platforms: the On-Demand & the Video Editor ones

YouTube – created in February 2005 by two ex-employees of eBay – reaches a surprising 100 million video views in July 2006 (representing 42.2% of the videos views on Internet that time)[3]. In the same year, around 65 thousand new digital videos were posted daily by common users. After some months, "Time Magazine" dubbed YouTube 'Invention of the Year'[4]. In October 2006, Google acquired YouTube for 1.65 billion dollars[5].

> *"What is revolutionary about YouTube is that it represents, in Pierre Levy's words, a 'natural and sound appropriation of discourse'; a website where Mass Media is quoted and remixed; where homemade media achieves public access and several subcultures produce and share media…The rhetoric of the digital revolution predicted that the new media would replace the old one; however, YouTube exemplifies a cultural convergence…the business model of YouTube creates added value by means of circulation…Although much or the Remix Culture is based on parody, this genre intensifies the emotional experience of the original material, bringing us deeper into the main characters' thoughts and emotions…The current status of YouTube makes it an inevitable platform for broadcasting content generated by its users…YouTube or not to YouTube, that is the question."*[1]

Hence, this videographic platform of a 'real time' dynamic flux can be perceived as a public archive of audiovisual memory from a Jean-Paul Fargier – related to Godard – writing:

> *"Television, since in its origins has been a device for the endless reproduction of the present and has a memory capable of boundless storage… one you can refer to not just as a testimony of the past but also to replace an impossible live image… it is therefore, a databank of all images, including those of the cinema."* [6]

So, YouTube becomes a global reference for on-demand video, and it is based on this concept that I consider this online platform as a main example for quoting, creation and aesthetic possibility of a networked audiovisual. It is beyond the scope of this work then, to talk about other different aspects of YouTube related to a "digital culture" such as its political and economical implications.

Besides *YouTube,* there are some other ones found along the research during the "on demand online video platforms" mapping period: HTML5, 5min, Activistvideo, Break, Blinkx Blip, Clipshack, Currenttv, Dailymotion, Dalealplay, Exposureroom, Flurl, Getmiro, Graspr, Howcast, Liveleak, Mega Video, Metacafe, Mixplay, Mojoflix, Myvideo, Sapovideo, Screenjunkies, Tu.Tv, Tvig, Veoh, Viddler, Videolog, Vimeo, Vodpod, Vxv, Wildscreen etc (not all of them referenced here – mapped from 2008 to 2011 – are still online available).

Also During the year of 2011 (until October) I had already mapped some online video editors and, with the aid of my undergraduate students[2] that had tested and briefly reviewed them to me, we found some quite interesting at that time, to which concerns me in these studies, as the following ones: YouTube Editor, Pixorial, Kaltura, VideoToolBox, Animoto, OneTrueMedia, Magisto, MixMoov, Stupeflix, Stroome, WeVideo, Cellsea and JayCut (the best of them – in terms of uses, tasks and also in the application interface –, but since January 2012 it unfortunately no longer exists).

The main goal of this mapping on the "online video montage platform" was to identify a potential existing stage – i.e. already developed and/ or in operation – within the same sort of web-application that I had proposed since the beginning[3] of my researches on the audiovisual methodologies for ordinary users both on- and off-line.

Therefore, this knowledge on "the state of art" in which audiovisual lies on nowadays is extremely important, i.e. for the development of an app type as I've been proposed, all information about a specific subject – as much "on-demand online video" as what "video editor" platforms – means understanding (learning/ thinking) about what has been transformed in terms of audiovisual, more specifically to the online one and its new distribution and production sorts.

[1] http://dev.opera.com/articles/view/introduction-html5-video/

[2] along a subject area called "New Technologies" in an Industrial Design undergraduate course, related to Interface and Usability topics.

[3] regarding to a part of my final undergraduate work (2003, a Flash application delivered), bearing in mind the "Web-Vj'ing-Cam" (2005, an online application prototype submitted to Transmidiale Festival) and also concerning to my master degree research (2008-2010).

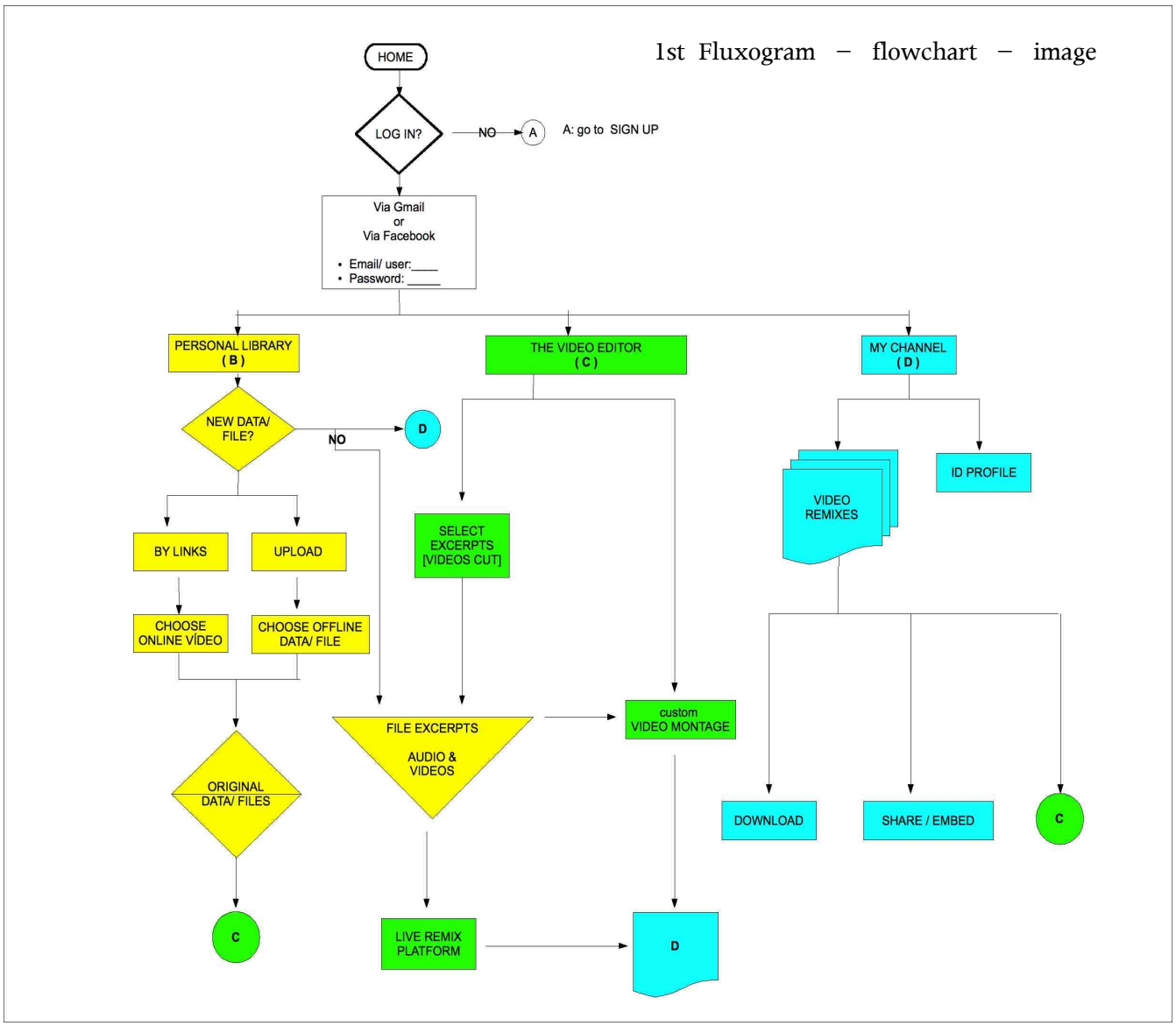
4. The HTML5 Possibilities ["the state of art" at online interactive videos issue]

> *HTML5 brings audiovisual pieces as part of this markup, i.e. in the near future browsers may not require many plugins (as Flash-Adobe) for playing videos and interactive features.* [7]

On March 18, 2009, Google launched a portfolio page about the new possibilities on the web through videos and animations – interactive videos – in HTML5 and JavaScript: "Chrome Experiments"[4]. My favorites are "Arcade Fire: The Wilderness Downtown" (by Chris Milk: *"An interactive HTML5 short created with data and images related to your own childhood. Set to Arcade Fire's song "We Used to Wait," the experience takes place through choreographed browser windows and utilizes many modern browser features. A collaborative effort between Google Creative Lab and Chris Milk."*) and "All is not lost", an OK Go videoclip.[5]

The significance of this focus – even if it does not seem to fit the type of research needed to develop the proposed application – is, for my PhD, a minimal understanding about current experiments carried out at video with online open source and its already developed "new technologies", especially some kind of projects under important sponsors like Google.

So, such creative experiences concern mostly to the online video field studies and its possibilities.



[4] http://www.chromeexperiments.com/

[5] Respective Images and more about it can be read in my last published text [see reference 7]

5. Few more lines about the proposed App

Everyday an enormous amount of information is posted on the web, but how can it be organized? Several online applications already take care of these ever increasing paths that sometimes stray or converge, but always link to one another. How can we make any audiovisual quote (link) easier inside a remix process? Maybe the expression should not be "make it easier", but rather "organize", assign a path for what has been researched for analysis and further creation. How can we stimulate people to create information from such varied paths in an environment where there is also the possibility of remixing this material and making it accessible to other users?

The application proposed here – the tech-methodological part of my PhD research –, is inserted in the so-called "digital culture" and represents a technological innovation, since it introduces an online audiovisual practice concept. As a study, it shows how videographic production and creation can be achieved using online videos. It may also be used as an important tool in learning environments, helping create rhetoric and expressiveness while working with both content and aesthetic aspects. The usage of this online application aims at promoting research and motivating networked participatory, shared experiences to develop cultural and digital skills.

Thus, I consider YouTube – and other current online audiovisual platforms – as a tool that opens a whole new spectrum of artistic and educational possibilities rather than just a video storage device. After all, the secret to make the student – or user/spectator – actively engaged in the process of creation is to transform him in a co-author .

> *"These paths [which facilitate movement of information between people] stimulate people to draw information from all kinds of sources into their own space, remix and make it available to others, as well as to collaborate or at least play on a common information platform. Barb Dybwad introduces a nice term 'collaborative remixability' to talk about this process: 'I think the most interesting aspects of Web 2.0 are new tools that explore the continuum between the personal and the social, and* ***tools that are endowed with a certain flexibility and modularity which enables collaborative remixability — a transformative process in which the information and media we've organized and shared can be recombined and built on to create new forms, concepts, ideas, mashups and services."*** [8](my highlights)

Let us refer to Manovich once again, "helping cultural bits move around more easily"; the proposed application, whose primary function was to facilitate the production and creation of the *Audiovisual Rhetoric* and its source noting, is inserted in a data terminology linked to "web 2.0":

> *"The Web of documents has morphed into a Web of data. We are no longer just looking to the same old sources for information. Now we're looking to* ***a new set of tools to aggregate and remix microcontent in new and useful ways."*** [8] (my highlights)

This online audiovisual remix application project aims at being a data organizer (a covered audiovisual path) for its own online switcher to create live ***Audiovisual Rhetoric***; i.e. the intention is to establish an online method of audiovisual writing with any digital videos (data) stored at video platforms a methodological elaboration for online rhetoric such as quick audiovisual quotes, like postings in a blog. The application may be a kind of audiovisual blog, where writing (text) is created through the remixing of videos linked in YouTube or other ones.

At the end of the prototype development the app will be evaluated: I intend to involve ordinary online users (some known stakeholders like public high school teachers and undergraduate students[6]) will be asked (invited) to test it. As this is an open source project, we hope – after the public launch of the prototype – tha

t the online community has an interest in developing new tools and functions according to some needs found and share it back (as has been occurred with the Wordpress Open Source Project).

Finally, this new app intends to be a simple online platform-system based on metadata technology in which some of the characteristics found in several current video-editor and/or vj'ing software are brought together. Its target, primarily, is to create an online *Audiovisual Rhetoric*.

6. References

[1] JENKINS, Henry. 2009. O que aconteceu antes do YouTube? in YouTube: digital media and society series. (Burgess & Green org.) [brazilian version] Aleph : São Paulo.

[2] *SZAFIR, Milena. 2010. YouToRemix: the online audiovisual quotting live remix application project. EuroITV2010 Proceedings, June 9th-11th 2010, Tampere, FINLAND. Copyright 2010 ACM 978-1-60558-831-5 /10/06*

[3] "Manifeste-se [todo mundo artista] – Mobile webTV Live Broadcast". 2006.

<http://www.manifesto21.com.br/manifestese2006/projeto.htm> accessed in February 2010.

[4] http://news.cnet.com/8301-10784_3-6133076-7.html

[5] http://techcrunch.com/2006/10/09/google-has-acquired-youtube/

[6] *Video Gratias.*2007. in Cadernos Videobrasil, número 03 <http://www2.sescsp.org.br/sesc/videobrasil/up/arquivos/200803/20080326_195612_Ensaio_JPFargier_CadVB3_P.pdf>

[7] SZAFIR, Milena. 2011. Breve 'Estado da Arte' do Vídeo Digital Online em 2011 [da Produção-Criação ao Armazenamento-Distribuição e Consumo] / ***A Brief 'State of Art' on Online Digital Video in 2011 [from Creation Production to Storage-Distribution and Consumption]*** /

[8] MANOVICH, Lev. 2005. Remixability <www.manovich.net/DOCS/Remix_modular.doc>

[9] SZAFIR, Milena. 2010. "Audiovisual Rhetoric..." Master Degree at Communication and Arts School in University of São Paulo.

<http://www.teses.usp.br/teses/disponiveis/27/27153/tde-17022011-150239/pt-br.php> acessed in March 2012.

[6] Both stakeholders types are part of my related jobs during these last years: I used to work as instructor in retraining course for public high school teachers and also as a teacher on several colleges of design and audiovisual.

Posters

Defining the modern viewing environment: Do you know your living room?

Sara Kepplinger
Ilmenau University of Technology
PO Box 10 05 65
98684 Ilmenau, Germany
Phone: +49 (0) 3677 / 69 2671
Sara.Kepplinger@tu-ilmenau.de

Frank Hofmeyer
Ilmenau University of Technology
PO Box 10 05 65
98684 Ilmenau, Germany
Phone: +49 (0) 3677 / 69 1267
Frank.Hofmeyer@tu-ilmenau.de

Denise Tobian
Ilmenau University of Technology
PO Box 10 05 65
98684 Ilmenau, Germany
Phone: +49 (0) 3677 / 69 2671
Denise.Tobian@tu-ilmenau.de

ABSTRACT
In this paper, we describe the activities towards the definition of average characteristics of the home environment and its respective viewing conditions concerning television consumption at home. This is based on a survey in which we are asking for general settings and focusing on the lightning conditions. This work points out the lack in definitions of viewing environments in the home context towards standardized quality evaluation tests representing the respective field of application. In order to overcome this shortcoming, we started to collect information about the viewing habits of the users and their different lightning situations. We present first results showing individual differences which we suggest to focus on in order to create different profiles about the home viewing environment.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – evaluation/methodology, user-centered design

General Terms
Measurement, Experimentation, Human Factors, Standardization

Keywords
Viewing Environment, Home Environment, Television

1. INTRODUCTION & RELATED WORK
By comparing the television environment and the viewing habits over the history since the invention of television (TV) one can recognize differences. Nowadays, several more aspects come along with the television device usage and trends forecast a multi-screen viewing habit for the future. However, there are also similarities concerning the viewing environment and watching TV is still seen as a relaxing and social activity. These are results based on studies defining the TV usage and the user experience (UX) in the home context like for example in [1]. Focusing on visual quality, comprising methods for the subjective assessment of visual representations, recommendations and standards propose certain viewing positions and lightning conditions. Thus, standardized conditions are needed to ensure comparable results of quality evaluations. These conditions and their correlations with other factors are relevant and examined topics. Other factors may be for example age [2], desirability and practicability [3], viewing distance in connection to screen size [4]-[5], dependency of scan line structure visibility [6], preferred viewing distance (PVD) versus design viewing distance (DVD) [7], and these in connection to different formats [8]. Several actual recommendations propose a different viewing environment for the home context than for the laboratory viewing environment in order to include the end user (e.g. [9]-[11]). However, recommendations define the home environment including the lightning condition and the viewing angle only sporadically. Figure 1 shows a picture about an exemplary home viewing environment. This does not contain a lot at the moment, as standardized data is missing. Simply creating average values about the characteristics building the home viewing environment is difficult, as it is a very private environment and may contain a lot of personal life-style. Therefore, this may vary a lot between different kinds of users.

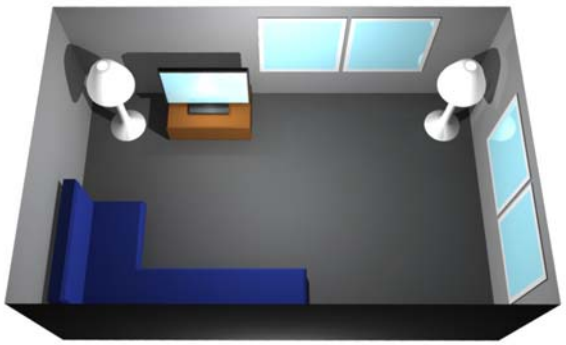


Figure 1. Scheme of a home viewing environment

Existing and often cited standards, for example the Recommendation ITU-R BT.500-13 about methodology for the subjective assessment of the quality of television pictures, define the home viewing environment in order to evaluate quality at the consumers' side of the TV chain [9]. Therefore, parameters have been selected which reproduce a near to home environment but define an environment slightly more critical than the typical home viewing situations using a cathode ray tube (CRT). The importance of taking in mind the home environment in conjunction with viewing distance, picture size and subtended viewing angle especially for three dimensional TV (3DTV) is emphasized by the actual report on features of three-dimensional television video systems for broadcasting, the ITU-R BT.2160 [10]. ITU-R BT.1438 [11] focuses on the subjective assessment of stereoscopic television pictures. Chen et al. [12] pointed out new requirements of subjective video quality assessment methodologies for 3DTV. Herein, they suggest a more precise definition concerning the background and room illumination, as based on different representation techniques the influences on the quality of experience (QoE) vary. The recommendation ITU-T

P.910 [13] on methodology for the subjective assessment of video quality in multimedia applications suggests a background room illumination of ≤ 20 lux under consideration of maximum detectability of distortions. This concrete suggestion varies from previous mentioned recommendations focusing on TV and stereo 3DTV. Herein, a ratio of luminance of inactive screen to peak luminance of ≤ 0.02 and a maximum observation angle relative to the normal of 30° are suggested for the general viewing conditions for subjective assessments in home environments as well as for the laboratory environment. Environmental illuminance on the screen in the home should be 200 lux. To whatever extends, besides others, illumination and viewing angle are important influencing factors (see also [14]-[18]), not only in the home context where TV consumption takes place. The actual viewing distance may not concur with the preferred viewing distance (PVD) recommended. In order to support a tighter definition of the actual viewing environment at home we conducted a qualitative study.

2. RESEARCH METHOD

In order to gain information about the reality at home concerning the illuminance, the viewing environment, the viewing angle, as well as the viewing distance to the used screen characteristics we conducted a survey in conjunction with light metering during the individual prime time. With individual prime time we define the time the participant usually uses the main TV device. With main TV device we define the device the participant uses usually and most regularly. The work is presenting our activities to collect information concerning the following research question: Is it possible to find average values in order to define a representative setting for a defined home environment for subjective evaluations and comparable results? Herein, we present the research principle with which we envisage a higher amount of data sets.

2.1 Test Environment

The test environment is the participants' respective room containing the main device during the individual prime time.

2.2 Data Collection and Analysis

We measured the illumination with an exposure meter. Additionally, we asked for the amount of used TV devices within the household, and how many people usually are using the main device together during the individual prime time. Besides a description of the respective room conditions (how many square meter and which properties or which kind of room, open or closed blinds or curtains) and the illumination, we asked for the viewing angle, the actual viewing distance and the TV screen size. We asked for the kind of device used (e.g. CRT, LCD, Plasma, 2D, 3D…). We measured the illuminance with the TV device showing a white screen, showing normal television program and being in off mode, respectively black. Therefore, we measured two times each, once in front of the users' eyes and once in front of the television. We also asked, if the participants used the possibilities of adjustments offered by their TV device. Finally, we asked for demographic data including age, sex, and profession.

Combining the data about the illumination, the PVD, and the TV screen size allows the definition of the peak luminance as well of a ratio of luminance of peak luminance (white screen) to inactive screen (black). This seems to be useful in a further step with more data sets. In general, an idea on the commonness of illumination in the home environment is acquired as well as accumulative values concerning the viewing angle and the viewing distance preferred and individual differences. Paying attention to differences and individuality can be used for the creation of several different viewing environments. Frequency of occurrence is used for now.

2.3 Participants

The selection of test participants at this stage was based on the readiness to open the living room to the researcher rather than on a representative sample selection which would be the theoretical quota. In sum, 19 participants answered our survey and we were able to measure the illumination from 10 home environments. The majority of the participants are middle-aged, but we also included seniors and best ager participants. Most of the participants (12) are students, two are retired and five are in full-time employment.

3. RESULTS

The results contain data sets from seven female and 12 male participants. The eldest is 73 years and the youngest is 22 years old. In average, the participants have an age of 31 years.

3.1 TV Usage

The majority of the participants (12) are using their TV device on their own, six of the participants are in front of the TV together with one other person, and one participant is watching TV in a threesome. Almost all participants (18) have and use one device which is also seen as the main TV device which is located in the living room. One participant has three devices, but defined the one in the living room as the main TV device. The other devices are in the kitchen and in the dressing room. Seven participants use no other device besides the TV device usage. Eight participants use one device, namely the laptop, during the usage of the TV. Three participants use additionally to the laptop their smart phone. One participant uses these two devices and the desktop PC additionally to the TV. The majority (13) uses the main TV device in the evening, one participant uses the TV during the night, and one uses it in the afternoon and the evening, and one in the morning and in the evening. These times are defined as the individual prime time by the participants, as they use the TV every day and regularly at these reported day times. Three people use the TV at any but not at a specific time.

3.2 Viewing environment

In average, the viewing room has 21 square meters. The biggest room reported has 32 and the smallest 9 square meter. The room conditions are most often (12) light dark with a dimmed indirect light source in the evening and bright during the day. Three participants have always bright light. Four participants always have it dark without any additional light besides the TV device. Asking for open or closed curtains or blinds, the majority (9) answered with closed, seven have them open, two are using open or closed conditions based on necessity, and one has no possibility to change. Adjustments offered by the TV device are only used sporadically. Seven participants have never adjusted any condition. Seven participants changed the brightness and contrast of their TV device. Two change the aspect ratio, use dynamic light adaption, and adjust the audio settings if necessary. One participant reported that the TV device was calibrated based on his requirements. One participant just adjusted the clock, and one only used the automatic program lookup when starting the TV usage.

The viewing angle reported varies a little bit. The majority (9) reports to sit in front of the TV with no horizontal or vertical shift. One reported a vertical shift of 15 degrees and one reported a horizontal shift of 15 degrees. Two reported a viewing angle from frontal to maximum of 30 degrees of horizontal shift. One reported a viewing angle of 20 degrees horizontal and 30 degrees vertical shift to the TV device. One reported a difference of 10 degrees horizontally. Two reported a viewing position of supine and a slight head enhancement towards a frontal position to the TV device. Two reported a constantly changing viewing position from lying to staying in front of the TV between 30 to 80 degrees horizontally. The display size is very different amongst all participants. The smallest device is 13 inches small, and the biggest device is 50 inches big. Taking in consideration all variations, the average size is 28.28 inches. The viewing distances vary a lot. However, the average distance is 240.79 centimeters. The minimum viewing distance is one meter and the maximum distance is 450 centimeters.

A collection of the lightning characteristics within ten data sets gives a small insight. Unfortunately, it was not possible to collect data with a white TV screen within four households, therefore they are marked with no answer (n. a.). Table 1 summarizes all data sets collected in front of the users' eyes during different TV device conditions, and shows the values collected in front of the TV display with the same conditions. Within this data set there is no 3D-TV available.

Table 1. Illuminance (lux) in front of the users' eyes and in front of the TV screen with different conditions

TV kind	TV off		TV white		TV program	
	TV	user	TV	user	TV	user
LCD	02.73	11.38	228.40	14.38	28.10	12.42
Plasma	19.27	63.00	147.10	70.90	25.50	66.90
CRT	25.00	00.94	163.00	07.89	30.00	02.20
CRT	01.01	03.15	n. a.	n. a.	104.30	06.39
CRT	10.00	27.90	n. a.	n. a.	231.00	33.00
CRT	03.17	18.62	n. a.	n. a.	237.00	21.12
Plasma	00.50	07.40	n. a.	n. a.	68.90	07.50
LCD	06.22	16.33	233.20	38.70	77.00	19.30
LCD	00.67	21.56	242.70	39.40	32.90	13.77
LCD	00.51	02.72	161.90	07.61	05.73	03.30

Taking in consideration the illuminance in front of the participants' eyes during the usual TV program, the minimum illuminance is 2.20 lux (CRT) and the maximum is 66.90 lux (Plasma). Figure 2 shows an arrangement of the reported data on viewing distance, display size and illuminance. Therefore, the size of the red dots visualizes the amount of participants with the same viewing distance conditions in meter. Based on this commonality, the respective average value of the display size (inches on the left side) is shown as well as the respective illuminance value (lux on the right side) if data on this was available. This figuration might not be fully acceptable in terms of reliability as too less data is computed and we already discussed that simply calculating an average value misses the representation of individuality and important differences (such as kind of display). However, already herein the diversity of the data combination from display size, illuminance and viewing distance is shown. As for example at the viewing distance of 3 meters, one would assume a rather bigger display than with less viewing distance and the illuminance (during the TV program) seems to vary independently from the distance.

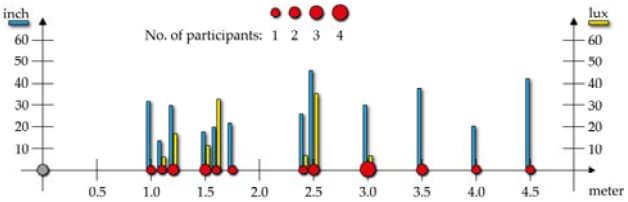


Figure 2. Average display size (inch) and illuminance (lux) based on the participants' viewing distance

To conclude this presentation of results, we would like to further discuss that the home environment representation in evaluation guidelines differs from the real home. Important factors, especially for new viewing possibilities like 3D, are different from user to user. These factors are for example viewing angle, viewing distance, lightning conditions, and display size. However, these differences may be concluded into several different groups in a next step and formulated into different viewing environment profiles representing special user groups. Based on a higher amount of data sets, such a solution is suggested in order to support comparable evaluation results from subjective quality assessment under consideration of the users' individuality.

4. DISCUSSION

This study is a first step towards the evaluation of general viewing conditions for subjective assessments in home environments. The purpose is to include the home environment for more standardized and comparable evaluation activities by creating typical viewing environment profiles. More data needs to be collected in order to give a scientific statement on possible commonalities and significant differences. Up to now, our results show, that there is the need of a tighter definition concerning the viewing conditions including differences in device usage and lightning conditions. Different displays create a different monitor contrast which can be influenced by the environmental illuminance. Additionally, the in theory changing TV consumption and device usage (e.g. second screens) has to be considered furthermore beyond just asking for the main TV device usage in a further step. However, dependent on the task which builds the basis for subjective assessment of quality in the home environment these differences may play an influence and have to be regarded.

5. NEXT STEPS

Based on this presented work we plan to collect a lot more data from the field representing the reality. We wish to create viewing environment profiles representing differences which would allow a more standardized evaluation environment for comparable results. Therefore, we plan to work together with an association for technical inspection in order to include information on contrast and colors together with different lightning conditions

and different kinds of TV devices. Further steps will be the combination of different user profiles with lightning conditions to see whether there are differences between TV, cinema and technology affine people, the average users, different generations, or age, gender and other aspects. After collecting enough information the resulting profiles based on average values may look like these examples:

The modest one:

- Device (monitor, processing, resolution, screen size, format ratio): LCD, digital, LV 120 cd/m², 28.28 inches, 4:3)
- Ratio of luminance of inactive screen to peak luminance: based on device used and resulting contrast > 0.02
- Display brightness and contrast: factory settings
- Maximum observation angle relative to normal: 45°
- Viewing distance: 3 meter independent of display size
- Peak luminance: > 200 cd/m²
- Environmental illuminance on the screen: > 200 lux

The discerning one:

- Device (monitor processing, resolution, screen size, format ratio): LCD, digital, LV 200 cd/m², 50 inches, 16:9)
- Ratio of luminance of inactive screen to peak luminance: ≤ 0.02
- Display brightness and contrast: set up via PLUGE [9]
- Maximum observation angle relative to normal: 10°
- Viewing distance: based on PVD rules
- Peak luminance: > 200 cd/m²
- Environmental illuminance on the screen: < 100 lux

These two, of course, are only an example which may be the result of such a data collection with its consequential average values and certainly has to be more concrete at that. After the definition of such samples, these findings may be applied for the subjective quality assessments including the home environment instead of only the laboratory as a context of use.

6. ACKNOWLEDGMENTS

Our thanks go to our local technician and external technical consultants as well as to all test participants. This work was conducted within the project no. BR1333/8-1 founded by the German Research Foundation (DFG).

7. REFERENCES

[1] Pirker, M., M., Bernhaupt, R. 2011. Measuring User Experience in the Living Room: Results from an Ethnographically Oriented Field Study Indicating Major Evaluation Factors. *In Proceedings of the Euro ITV Conference 2011 (EuroITV.2011)*, Lisbon, Portugal.

[2] Nathan, J. G., Anderson, D. R., Field, D. E., & Collins, P. 1985. Television viewing at home: Distances and visual angles of children and adults. Human Factors, 27, 467-476.

[3] Tanton, N. E. & Stone, M. A. 1989. HDTV Displays (Rep. No. BBC RD 1989/9 PH-295). BBC Research Department, Engineering Division.

[4] Diamant, L. 1989. The Broadcast Communications Dictionary. (3rd ed.) Greenwood Press.

[5] Ardito, M. 1994. Studies on the Influence of Display Size and Picture Brightness on the Preferred Viewing Distance for HDTV programs. SMPTE, 103, 517-522.

[6] Poynton, C. 2003. Digital Video and HDTV Algorithms and Interfaces. San Francisco, CA, USA: Morgan Kaufmann.

[7] Ardito, M., Gunetti, M., & Visca, M. 1996. Influence of display parameters on perceived HDTV quality. IEEE Transactions on Consumer Electronics, 42, 145-155.

[8] Lund, A. M. 1993. The Influence of Video Image Size and Resolution on Viewing-Distance Preference. SMPTE, 102, 406-415.

[9] RECOMMENDATION, ITU-R BT.500-13 Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union, Geneva, 2012.

[10] RECOMMENDATION, ITU-R BT.2160-2 Features of three-dimensional television video systems for broadcasting, International Telecommunication Union, Geneva, 2011.

[11] RECOMMENDATION, ITU-R BT.1438 Subjective assessment of stereoscopic television pictures, International Telecommunication Union, Geneva, 2000.

[12] Chen, W., Fournier, J., Barkowsky, M., Le Callet, P. 2010. New requirements of subjective video quality assessment methodologies for 3DTV. In *Proceedings of the Video Processing and Quality Metrics 2010 (VPQM)*, Scottsdale, USA.

[13] RECOMMENDATION, ITU-T P.910 Methodology for the subjective assessment of video quality in multimedia applications, International Telecommunication Union, Geneva, 2008.

[14] RECOMMENDATION, ITU-R BT.1128-2 Subjective assessment of conventional television systems, International Telecommunication Union, Geneva, 1997.

[15] RECOMMENDATION, ITU-R BT.1129-2 Subjective assessment of standard definition digital television (SDTV) systems, International Telecommunication Union, Geneva, 1998.

[16] RECOMMENDATION, ITU-R BT.710-4 Subjective assessment methods for image quality in high-definition television, International Telecommunication Union, Geneva, 1998.

[17] RECOMMENDATION, ITU-R BT.1210-3 Test materials to be used in subjective assessment, International Telecommunication Union, Geneva, 2004.

[18] RECOMMENDATION, ITU-R BT.1666 User requirements for large screen digital imagery applications intended for presentation in a theatrical environment, International Telecommunication Union, Geneva, 2003.

A Study of 3D Images and Human Factors
: Comparison of 2D and 3D Conceptions by Measuring Brainwaves

Sang Hee Kweon
Associate Professor, Journalism and Mass Communication, Sungkyunkwan University
82-2-760-0392
skwoen@skku.edu

Eunyoung Bang
Graduate School of Journalism and Mass Communication, Sungkyunkwan University
82-2-760-0392
sophiaey@naver.com

Bon Soo Kim
Bon Dental Medical Co., Ltd
MD
82-2-760-0392
Bon2828@hanmail.net

ABSTRACT
The study is based on the theses of McLuhan, 'Media is a message.' This study, which centers on 3D pictures and human facts, has been designed to observe the differences in receivers' cognition of three-dimensional (3D) stereoscopic and flat images through brainwaves tests. We attempted at understanding the characteristics of two-dimensional (2D) and 3D pictures and comparing the brainwave patterns of receivers when exposed to 2D and 3D TV contents. The main focus of this study is to grasp the differences between the human facts that affect cognition of stereoscopic and flat images in regards to their dimensional distinction by gathering concrete empirical data and comparing the alpha (α) and beta (β) wave patterns observed when watching 3D and 2D images. To this end, we adopted a 2X3 experimental research design and statistically processed the results gathered from 20 subjects. The 2D and 3D materials used in this study were divided into three categories—sports, animations, and promotional images—that have distinct structures. We exposed our subjects to these images and measured the differences in brainwaves according to the genre of the material. As a result, we found that β wave vibrations observed during subjects' exposure to 3D pictures were statistically significant compared to when subjects were watching 2D images. In addition, subjects showed stronger β waves in the frontal lobe when watching pictures with higher degree of stereoscopic effect and dynamic movements such as sports. In short, the greater the three-dimensional rate and the more frequent the transition of frames, the greater the brainwave activity, which also suggests a correlation with headaches and vertigo.

Categories and Subject Descriptors
D.3.3 Track 2: Media, Social and Economic Studies

General Terms
Human Factors, Experimentation, Measurement

Keywords
Experimental research, 3D TV, Human factors, Brainwaves, Cognition. α waves, β waves, Genre

1. INTRODUCTION
The We are living in an era represented by smart and three-dimensional (3D) technologies. In regards to 3D, relevant technology has developed significantly since its appearance in the 19th century but receiver preference remains low. Over the period covering the 1910s to the 1930s, there was a vibrant trend in the development of technologies for creating stereoscopic pictures. 3D was at full demand in the 1950s, resulting in the production of 69 motion pictures in 3D, but the trend quickly reversed when introduced to large screens. This required a new understanding about the process of reception, or to be precise, how receivers are accustomed to "low-involvement media," which, unlike telephones or other media for communication, engaged them in a passive process that allows them easy access with the least amount of cognitive and physical effort. In other words, 2D pictures provide receivers with what is necessary for processing information, thus decreasing brain activity. The spectacle element of the images also contributed to a dramatic development of the story. 3D pictures, on the other hand, are "high-involvement media" concerning not only brain activity, but also visual activity. Such a high level of sensory involvement accompanies fatigue, vertigo and headaches. Therefore, by comprehending the characteristics of 3D images by measuring brain activities during the reception process, it may be possible to grasp the course of how 3D images are received.

This study, which is based on an existing study concluding that 3D pictures and human cognition have a close relationship with brain activity, intends to step aside from the technological aspects of this subject and delve into the human facts of 3D image recognition. On the basis of the research question that explores the association of a receiver's cognition of images with the cognitive activity concerning images and brainwave movements, we assume that measuring receivers' brainwaves will provide insight into receivers' cognition of 3D pictures and media. Therefore, this study has been designed to determine how receivers' brainwaves differ according to the type of 3D stereoscopic image they were exposed to in an empirical manner. Through the course of this study, we will examine how 2D and 3D images induce different brainwave patterns. This will serve as an initial step to in-depth discussions on how changes in brainwaves movements result in different cognition of receivers. While the main objective of this experiment is providing empirical data through scientific methods, it is also aimed at suggesting a direction and questions for media contents in a cultural or social perspective.

Accordingly, this study attempts to approach 3D TV from a liberal arts point of view, whereas existing researches on this subject mainly focus on its technological aspect. In short, we intend to put focus on viewers of 3D TV and explore their cognition. To this end, we designed this study to measure and describe the brainwave movements of receivers during the reception process when watching 3D pictures or using 3D contents. The reception process of 3D images is complex, so a study that both describes and comprehends the characteristics of 3D images is likely to be of great academic and industrial value. This study compares the cognitive dimension of users when viewing 3D and 2D pictures and measures the level of brainwave movements according to genre.

2. BODY TEXT

2.1 Background

2.1.1 The McLuhan's Proposition about Media Effect

McLuhan explained media effect with the proposition "Media is a message". Even though there are same contents or message, users receive different effects depending on the media. Like McLuhan's saying, people can be affected by the dimension of images like 2D images and 3D images although the same contents. In other words, we can expect that active parts of user's brain will be different in 2D and 3D images. Also user's cognitive processing will be different depending on the images dimension like 2D and 3D Images.

McLuhan suggested the definition of cool media and hot media. The media is distinguished by definition and user's participation. Medium that has high definition and low user's participation like newspaper is a hot media. However, like television, medium that has low definition and high user's participation is a cool media. We can apply this proposition into image dimension. 3D images have high definition, and low participation for users to accept the images comparing to that 2D images has. So 3D images can be hot media. However, 2D images are with low definition and more participation is needed for users comparing to 3D images are. Hence, 2D images are hot media.

2.1.2 Study of the Operation of 3D Pictures and Human Cognitive Fact

There are various aspects when it comes to understanding 3D pictures; on one side there is the engineering element, and on the other side there is the experience by the receiver. There are also the aspects of social interaction and the social and cultural perspective. This involves deciphering the industrial and cultural meaning of 3D broadcasting and films and determining the structure of their consciousness. A creator of 3D graphics is required to literally have a convergence of abilities in regards to engineering, art and communication. To begin with, in order to acquire stereoscopic images with two cameras, the creator should be adept at camera engineering, and also well-versed in the principle of human image cognition. Furthermore, he or she should have ample knowledge in the aesthetic features of broadcasting and pictures, and the capacity to construct 3D images from a whole new scenario.

When comparing 3D and 2D pictures, 3D includes more than twice the number of images than 2D because a 3D TV is a device that sends left- and right-side images to the left and right eye respectively. Furthermore, 3D TVs have faster per-second screen refresh rate than existing 2D TVs. A majority of the 3D TVs launched recently have a refresh rate of 240 Hz (240 frame per second) while 2D TVs refresh at a rate of 120 Hz. At 120 Hz, viewers experience afterimages and lower luminance (brightness of the screen). Therefore, for a receiver, watching a 3D picture not meeting ergonomic standards or uncomfortable to the human eye will be an experience accompanied with eye-strain, headaches and vertigo.

2.1.3 Cognitive Analysis of 3D Stereoscopic Images

2.1.3.1 Necessity of new cognitive study

In regards to the study of media contents reception, traditional scholars who emphasize the synopsis of the contents, try to measure changes in a receiver's attitude or an individual's cognition in accordance with the linear way of expression adopted in the media contents' structure. For instance, these scholars still maintain that the "Arisotelian dramatic experience" can explain the "interactive digital content experience" from online games (Hye-Won Han, 2005). In this line of thought, they attempt at understanding 3D pictures as an extension of 2D and analyze them in this manner. They also apply a evolutionary perspective to the definition of 3D pictures. In their view, if films are a form of narratives that evolved from novels, 3D pictures are a new form of narratives that evolved from and are inter-usable with 2D films.

Meanwhile, those of the ludology school point out that analyzing reception of 3D contents with the yardstick applied to reception of 2D pictures, brings the focus on the "representation" of pictures. Inevitably, they are likely to overlook other aspects of 3D such as the sense of depth and vividness, in short the "simulation" element which is a crucial characteristic of multi-view interactional contents. Overlooking the "representation" of 3D and the relevant brain activity patterns of the receivers will naturally lead to a fragmentary conclusion. 3D and 2D pictures leave receivers with distinct visual and cognitive experiences. Therefore, it is necessary to analyze the receivers' cognitive response to 3D rather than analyzing the characteristics of the contents. In this sense, ludologists argue that the contents of 3D pictures are not the narrative, but the simulation of spectacle during the process.

2.1.3.2 Characteristics of reception and cognition of 3D images

Studies on the cause of vertigo or headaches that occur when receiving 3D pictures are still at their initial stage, but they are based on the notion that the human psychology and behavior prefers naturality and this is the cause of sensory discordance. Sheridan believed that there is a correlation between the amount of sensory information and the level of brain activity when perceiving the information, and that this is associated with the amount of sensory information processed (Sheridan, 1992). He adds that a receiver's sensory organ is related to the variables that elevate sensory fidelity, cognitive fidelity and sense of stability, which in turn is associated with the amount of information processed by the brain. Sensory fidelity can be defined as an energy pattern applied to sense. When we assume an increase of aligned energy by what has been displayed by the media, this is

likely to occur amid the sensory channels concerning sight rather than smell.

The receiver's cognition of 3D stereoscopic pictures is known to vary by various factors such as how the picture is edited, sound, sense of depth and spectacle. According to existing studies on the types of cognition, we can divide cognition of 3D pictures into the following categories: sensory reality, cognitive reality, arousal as a receiver's cognition, pleasure, and vision fatigue from stereoscopic images.

2.1.4 Brainwaves for Measuring Human Facts Cognition Factors

The core subject of human fact studies is the pursuit of brainwaves studies as a way to find solutions for noxious factors such as visual fatigue that may erect a barrier to the distribution of 3D contents. Methods for human fact studies range from surveys on receivers, experiments and measuring brainwaves when receivers perceive real 3D images. Also, it intends to understand the cognitive pattern in the process of receiving 3D pictures through empirical data gathered by measuring the difference in brainwaves according to the genre, three-dimensional rate and editing of the 3D materials used in the experiment.

In general, brainwaves are classified according to frequency as follows: Delta (δ) waves (0.2-3.99Hz), theta (θ) waves (4-7.99Hz), alpha (α) waves (8-12.99Hz), beta (β) waves (13-29.99Hz), and gamma (ζ) waves (30-50Hz).

Among these, alpha (α) waves are mostly observed in comfortable conditions such as relaxation. It increases in amplitude amid greater levels of stability and comfort. The waves, which show continued regular patterns, are most prominent in the parietal and occipital lobes and least noticeable in the frontal lobe.

Beta (β) waves are most often seen in the frontal lobe and usually appears when we are awake or performing all types of conscious activities such as talking.

Below our skull is the cerebral cortex, which is divided in to four lobes: frontal, parietal, temporal and occipital. Each lobe carries out a different role. The occipital lobe, located in the back of our heads, contains the primary visual cortex which processes primary visual information. The parietal lobe, which is near the crown of our heads, is where the somatosensory cortex is located and responsible for processing motor and sensory information.

When we show different types of materials to our experiment subjects, we are able to witness the structure and developmental function of the brain through the changes of wave lengths for each channel. α and β waves have an inverse relation. If α waves are prominent after viewing a certain material, it is highly probable that β waves are low. In such a case, we may assume that the subject has received much visual stimuli, leading to active emotional thinking. In contrast, when β waves are relatively higher than α waves, this suggests that the temporal lobe and the whole of the frontal lobe has been stimulated, which in turn means more auditory stimuli that lead to complex and planned thinking rather than emotional thinking.

Table 1. Measuring equinment's brainwaves and function depending on channel

Right	CH 1	CH 2	CH 3	CH 4	REF
	Red	Yellow	Green	Blue	Brown
	Prefrontal lobe (β)	**Frontal lobe (β)**	Temporal lobe	**Occipital lobe (α)**	Earlobe
	Comprehensive thoughts		Hearing	Sight + Emotional thoughts	Standard electrode
Left	CH 5	CH 6	CH 7	CH 8	GND
	Orange	Violet	Gray	White	Black
	Prefrontal lobe (β)	**Frontal lobe (β)**	Temporal lobe	**Occipital(α)**	Back of hand
	Comprehensive thoughts		Hearing	Sight + Emotional thoughts	Earth electrode*

2.2 Research Questions and Research Method

2.2.1 Research Questions

Please If this study helps us discover the difference in human cognition of 2D and 3D, it will pave the way for further in-depth studies of 3D stereoscopic pictures. In fact, there is an existing study that concludes that 3D pictures generate more negative effects such as fatigue compared to 2D images.

In order to find out the cognitive difference among different types of graphic materials, this experiment basically focuses on brainwave responses. To this end, we intend to build a basis for the experiment by first defining and understanding the fundamentals of the human brain structure and brainwaves, and subsequently measuring receiver brainwaves responses. After grasping the difference between the brainwave responses to 2D and 3D materials, we will be able to learn the difference in thinking processes by observing the brain structure and effect of brainwave activation.

The overall research question is whether receivers' cognition of 2D and 3D pictures will differ. To explore this, we have set forth the following sub-question.

Research question 1: When exposed to 2D (flat images) and 3D (stereoscopic images), would there be a difference in the α and β wave patterns of the receivers?

Another purpose of this study is to find out the changes in human cognition by observing the changes in wave width detected by each channel of the measurement equipment when receivers view 2D and 3D images. The equipment has a total of eight channels—four channels each for the left and right brain—of which each represents a certain region of the brain and its function. Therefore, by determining the wavelength differences, we may perceive the differences in the human thinking process according to whether the subject is watching 2D or 3D. To this end, we have set the following question.

Research question 2: Will the wavelengths detected by each channel differ according to 2D (flat images) and 3D (stereoscopic images)?

Scientific verification of the difference in cognition that is of a liberal-arts nature is a challenging, yet essential task. In this

regard, the key topic of this study would be to what extent we can clarify the difference in cognition by employing scientific methods. Accordingly, we need to grasp the structure and functions of the brain that brings about cognitive differences, and use this knowledge to understand the whole process of thinking. Therefore, the central question for this study is the understanding about cognitive differences according to different types of graphic material by measuring brainwaves in order to perceive changes in the brain.

This will provide us with the chance to apply the cognitive process of 3D stereoscopic images to various kinds of media contents such as animations, films and sports, and figure out which content is most or least adequate for 3D format. By doing so, we can clarify the effect 3D pictures have on humans and gain an efficient perspective in regards to the production of 3D, which is still in its fledgling stage, and its application to various forms of media and contents. To fulfill this purpose, we have set the research question as seen below.

Research question 3: Will there be a difference in cognition of 3D stereoscopic pictures according to the type of media contents?

2.2.2 Research Method

To ensure an elaborate experiment design and to minimize trial and error, we conducted a pilot test in February 2011. For the experiment, we installed an LG 40-inch CRT TV and used a Virtual FX3D converter to realize 3D stereoscopic images. We employed 10 students of S University, a school located in Seoul, as subjects, and then showed them 2D and 3D pictures in accordance with an assignment table we have prepared in advance and compared their brainwave patterns. We also took control variables into account.

We used a True3Di 24 inch monitor, a product of Redrover, which is optimized to display the 3D pictures used in this experiment. We selected this particular equipment because the button on the front enables us to transfer general pictures into 3D format and therefore make control easier. We installed it at a fixed viewing distance and adjusted the height so the center of the monitor meets eye level.

As stimulus, we used three types of graphic materials: 1. Sports, 2. Promotional image, and 3. Animation. In regards to the selection of genre, the nature of the experiment made it impossible to include all genres, so we focused on the characteristics of the materials in order to fulfill the purpose of this study. In detail, the sports material is made up of only actual images, and is about basketball game. The screen shift of this material is so fast, and includes big sound of audience. Basically this image is quite active. The animation consist computer graphics, and has a way that specific character lead the image. The character of this material make people pay attention to dance with fun music. As for the promotional material, it has, aside from the other two genres, much potential for utilizing stereoscopic images so we adopted it in this experiment. The PR material is made with the purpose of advertising Digital Media City. There is the narration to deliver information about DMC. So, subjects need continuous concentration and full understanding of the flow. To prevent bias due to different running time, we edited all materials to run for a duration of around five minutes.

Before we started, we explained about the experiment to the subjects, and then questioned them about any experience of viewing 3D stereoscopic pictures or 3D TV for at least an hour, aside from short encounters in exhibition galleries, to control for the influence of such experiences. This experiment is developed except students who feel uncomfortable or are vulnerable with 3D image. Before starting the experiment, subject's current condition is checked, and subject could not participate in the experiment until 10 minutes later after they arrived the lab.

When a subject watches all three materials, he or she may be affected by the order they were shown. To prevent this, we made a list of all possible orders of materials and matched it with the order of the subjects.

First, we selected 10 students as subjects and properly attached the electrodes to their heads. When doing so, we observed the principle of making a small mark on the precise spot where we attach the electrodes with a water-based pen. In accordance with a single guideline, we made sure that the same amount of time is spent for each subject when explaining and preparing for the experiment and commenced right after all preparations are made.

The proportion of 2D and 3D is even in the order for showing the materials. The viewing orders have been planned in advance to avoid any confusion, and to prevent any influence of the orders, we made sure that no student watch materials in the same order as the other. For instance, if subject 1 was shown materials in an order of A-B'-C-A'-B-C', the first material will be shifted to be shown at the end. Accordingly, subject 2 will view materials in the order of B'-C-A'-B-C'-A. For every material, we measured the subjects' brainwaves, and then compared them according to dimension or genre.

We experimented on subject at a time, and each of them was asked to close their eyes and relax for five minutes after a material was over. By doing so, we tried to exclude the overall influence of emotions or physical changes that occurred while watching the previous material.

2.3 Results

2.3.1 Data Analysis

We may begin with answering the first research question by comparing the level of α and β waves measured when the subjects were watching 2D and 3D images.

Table 2 Average comparison between 2D and 3D brainwaves

Categories		2D Image	3D Image	T value	Significance probability
Alpha (α) waves	Sports	16.41	11.60	3.059	.018
	Animation	17.27	11.00	2.526	.039
	PR	43.29	10.53	4.238	.004
Beta (β)	Sports	23.92	71.64	-8.393	.000
	Animation	26.40	64.054	-3.028	.019

waves	**PR**	35.04	56.84	-2.671	.032

The above table is a documentation of α and β waves measured when the subjects were watching 2D and 3D materials. In general, α waves are more prominent when viewing 2D pictures and β waves are observed more when watching 3D pictures. α waves have higher T-value than zero in Sports, Animation, and PR, but get lower p-value then .05. P-value which is smaller than .o5 means that the result of this analysis can be trust. So, α waves are more remarkable in 2D images than in 3D images.

B waves have negative T-value in all three materials, and get p-value that is lower than .05. In other words, B waves have higher value in 3D images than in 2D images.

In other words, when exposed to 2D images, the subjects depend on sight when receiving stimuli. During this process, the occipital and parietal lobes are stimulated. On the other hand, subjects showed higher levels of β waves when watching stereoscopic images, from which we can assume that the temporal, frontal and prefrontal lobes are stimulated. In conclusion, 3D pictures require more intricate and complicated thinking.

Table 3. The analysis of absolute value about alpha waves between 2D and 3D images

Categories			**2D Image**	**3D Image**	**T value**	**Significance probability**
Alpha (α) waves	**Sports**	**ch 1**	24.66	13.39	6.534	.000
		ch 2	26.12	13.63	6.127	.000
		ch 3	12.94	11.57	5.418	.000
		ch 4	13.19	11.26	4.583	.001
		ch 5	13.27	9.33	2.196	.056
		ch 6	12.89	10.82	4.719	.001
		ch 7	12.77	10.62	3.964	.003
		ch 8	15.40	12.15	3.482	.007
	Animation	**ch 1**	27.94	13.36	3.686	.005
		ch 2	28.74	13.58	3.906	.004
		ch 3	12.82	10.46	6.315	.000
		ch 4	13.39	10.37	5.953	.000
		ch 5	7.682	9.98	4.469	.002
		ch 6	11.91	10.57	3.827	.004
		ch 7	11.73	9.56	4.566	.001
		ch 8	23.87	10.13	2.094	.066
	PR	**ch 1**	45.65	12.69	1.879	.093
		ch 2	52.23	13.06	1.697	.124
		ch 3	19.37	10.20	2.509	.033
		ch 4	25.06	10.24	2.012	.075
		ch 5	53.67	8.65	1.121	.291
		ch 6	83.19	9.74	2.144	023
		ch 7	18.15	9.54	5.064	.061
		ch 8	48.90	10.12	3.144	.001

In order to measure α waves, we had the subjects watch three kinds of materials—sports, animation and promotional image—in one session. Afterwards, we analyzed the α waves for each material type and compared them in regards to 2D and 3D. The result shows that regardless of genre, the level of α waves were higher than that of β waves when the subjects were exposed to 2D images*. Given that α waves are mostly observed in a relaxed state, we can expect the occipital lobe, where α waves are usually detected, to be stimulated the most. Stereoscopic pictures arouse all senses other than not only our senses of vision and hearing, but other senses as well, so we can find that other regions of the brain that are responsible for hearing and rational thinking receive more stimuli. In comparison, 2D images concentrate on sending visual stimuli and relatively develop the occipital lobe which deals with sight. As a result, the wave lengths detected by channels 4 and 8, which both represent the occipital lobe, are significant. We can also find significant results in channels 3 and 7, which correspond to the temporal lobe that is in charge of hearing, as well.

Table 4. The analysis of absolute value about beta waves between 2D and 3D images

Categories			**2D Image**	**3D Image**	**T value**	**Significance probability**
Beta (β) waves	**Sports**	**ch 1**	35.17	81.78	-2.401	.040
		ch 2	33.16	85.05	-2.533	.032
		ch 3	17.62	59.65	-1.674	.128
		ch 4	14.18	49.75	-1.862	.096
		ch 5	39.37	121.73	-1.488	.171
		ch 6	25.66	79.13	-2.173	.058
		ch 7	11.46	41.40	-1.621	.140
		ch 8	14.74	54.57	-2.019	.074
	Animation	**ch 1**	42.57	77.02	-1.273	.235
		ch 2	39.78	83.35	-1.110	.296
		ch 3	20.20	40.47	-2.393	.040
		ch 4	16.10	30.82	-3.838	.004
		ch 5	28.95	134.80	-3.168	.011
		ch 6	22.10	92.75	-2.673	.026
		ch 7	15.65	25.32	-9.242	.000
		ch 8	25.77	27.87	-13.067	.000
	PR	**ch 1**	43.67	88.60	-2.921	.017
		ch 2	46.55	92.54	-2.098	.065
		ch 3	24.23	40.73	-4.228	.002
		ch 4	22.11	30.47	-5.408	.000
		ch 5	45.27	99.91	-2.934	.017

		ch 6	54.43	51.68	-5.671	.000
		ch 7	16.21	24.16	-6.311	.000
		ch 8	**27.83**	**26.62**	**-8.723**	**.000**

Likewise, we measured β waves for sports, animation and promotional images simultaneously and cross analyzed the results to compare figures for 2D images with that for 3D. Accordingly, we can see that β wave levels were higher for all three types of materials when they were shown in 3D format* so it produces significant results for all eight channels are significant. Since stereoscopic pictures are comprehensive in that it stimulate all other senses as well as those dealing with sight and hearing, the results for channels 1, 2, 5, and 6—which corresponds to the frontal and prefrontal lobes that are responsible for complex and systematic thinking—are significant. We also witness significant results for channels 3 and 7—representing the temporal lobe that is connected to the frontal lobe—and, due to the strong visual effects, channels 4 and 8—the occipital lobe. Compared with flat images, stereoscopic images stimulate and draw response from the whole brain. As β waves are more noticeable when shown 3D pictures, it inevitably activates not only the five senses, but also our regions for thinking. Consequently, we learn that the wave lengths for each channel are different according to the type of material. Furthermore, when the wave length per channel changes, we are able to determine which region and function of the brain has been activated and thus conclude that α and β waves alters according to the material format.

2.3.2 Brain Map

In this study, we mapped and analyzed the subjects' brains when exposed to 2D and 3D pictures.

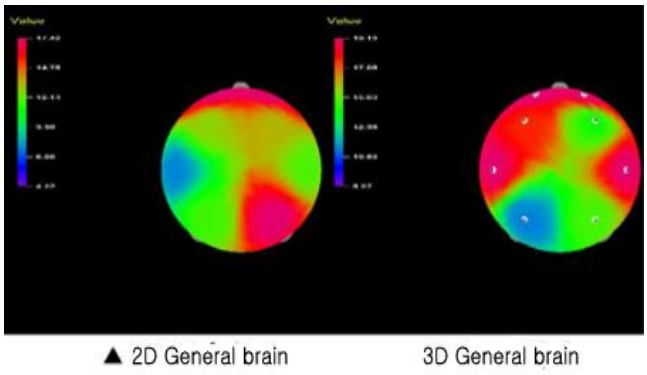


Figure 1. General brain image

The above figure is an analysis of average brainwaves when subjects are watching 2D and 3D images in general. When watching 2D, we observe significant development in the occipital lobe that represents sight. There is also activity in the temporal lobe that deals with hearing. On the other hand, the frontal and prefrontal lobes, which are involved in complex and planned thinking, sustains only a normal level of activity. In regards to watching 3D, the frontal lobe is extremely activated. The temporal and occipital lobes are also responding. When comparing the brain maps captured during 2D and 3D exposure, the frontal and prefrontal lobes, where β waves are usually observed, are much more activated when watching 3D pictures.

2.3.2.1 Sports

Sports footage features rapid transition of frame and requires concentration on the part of the viewer. In this sense, viewers show development in the frontal and prefrontal lobes to activate complex thinking when watching both 2D and 3D sports images. The left-hand brain map display activity in the temporal lobe, which is responsible for hearing, as well as the frontal and prefrontal lobes. Meanwhile, the map for 3D sports materials show a much significant level of activity in the frontal and prefrontal lobes when compared with not only 2D sports materials, but also other 3D materials. This means that out of all 2D and 3D materials, 3D sports footage raise β wave levels the most. For a more detailed result, we separated α waves from β waves and compared them when watching 2D and 3D sports materials.

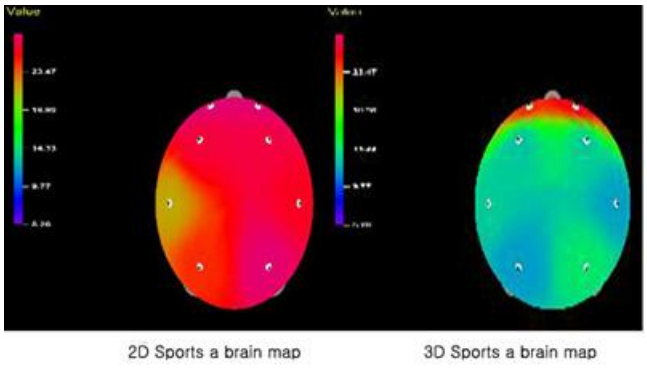


Figure 2. Sports alpha brainwaves

The above figure is an analysis of average α waves when the subjects were watching 2D and 3D sports materials. When the footage was in 2D format, α waves are active throughout the brain, particularly in the central region and temporal lobe. In regards to 3D footage, only the frontal lobe shows reaction, whereas the temporal or occipital lobes are slow to respond. α waves are frequently observed when we are comfortably or relaxed. Therefore, 2D sports materials induce higher levels of α waves than its 3D counterpart, suggesting that it makes viewers comfortable and relaxed.

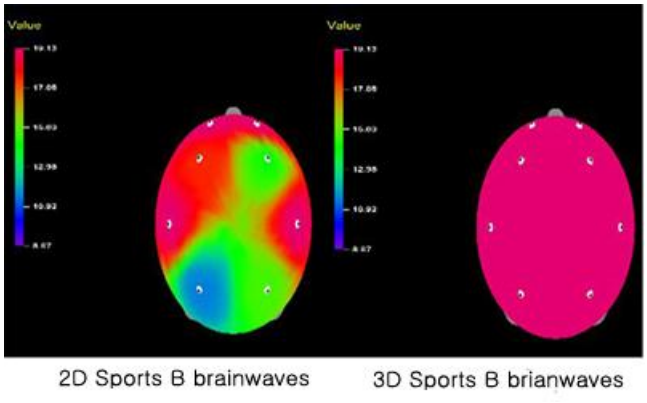


Figure 3. Sports beta brainwaves

β waves are observed when we are concentrating or tense. On average, 3D sports footage induce a higher level of β waves

than 2D footage. In short, watching 3D sports materials demand more concentration and tension. From the right-hand map, we can see that β waves are active regardless of region while the left-hand map shows higher β wave levels around the frontal lobe, which is the region that activates when concentrating or tense.

2.3.2.2 Animation

When exposed to 2D and 3D animations, the results were clearly distinct. For 2D animation, the brain activity is evenly spread throughout the prefrontal, frontal, temporal and occipital lobes. Given that animations are constructed of stories and characters that demand viewer's attention, β waves, which are basically needed for complex thinking, are inevitably active. 3D animation, however, is more elaborate and complicated than 2D, so the level of β waves is even more intense. The occipital lobe, which is where α waves are most prominent, shows stronger response when watching 2D animation. Given these results, how will they turn out when we separate α and β waves?

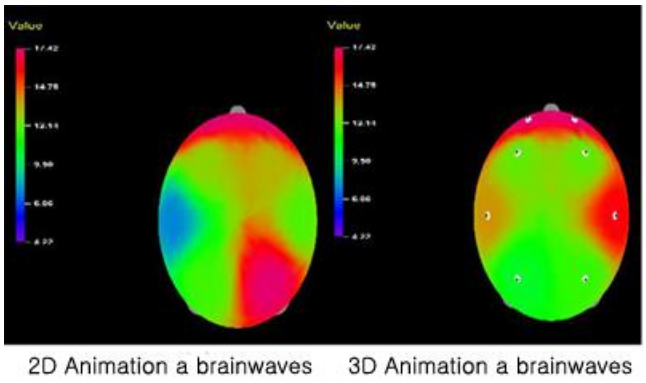


Figure 4. Animation alpha brainwaves

As mentioned before, α waves are most active at times of comfort and relaxation, so they appear weak for both 2D and 3D animations. Moreover, the brain maps for both formats are hardly distinct and nearly identical. In sum, both 2D and 3D animations are less involved with comforting and relaxing α waves.

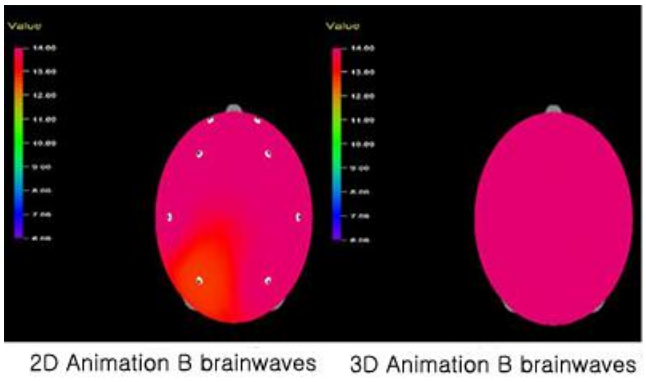


Figure 5. Animation beta brainwaves

β waves, which are active when performing conscious activities, are found to be prominent for both cases. Both 2D and 3D animations involve concentration and tension on the part of viewers. Yet, it is notable that 3D animations activate β waves more than 2D animations. The left-hand map shows that the occipital lobe that controls sight is relatively less active when watching 2D animations. 3D animations, in contrast, elevate β wave levels regardless of the lobes.

2.3.2.3 Promotional image

The objective of promotional images is delivering and introducing information, so viewers may find it difficult to pay attention and get easily bored unless they are interested. So when watching a promotional image in 2D format, viewers are likely to have more active α waves rather than β waves. In addition, the occipital lobe is more stimulated than the prefrontal and frontal lobes. When watching in a 3D format, we can observe increased activity of β waves, but this is still lower than when watching sports or animations. Also, since viewers' attention span tends to be shorter for 3D promotional images, we can see a more active occipital lobe and higher level of α waves compared to other contents. For more precise results, we compared 2D and 3D materials in regard to α and β waves.

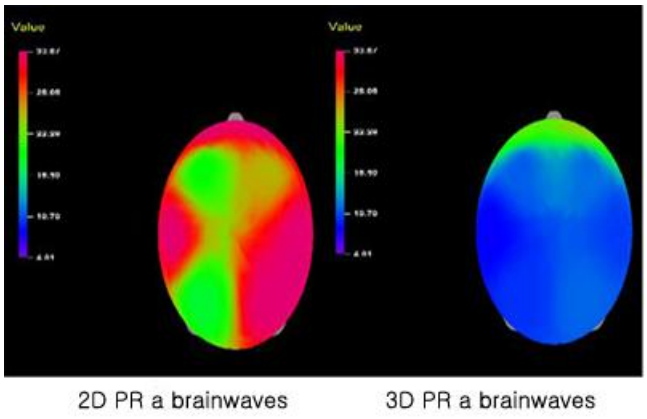


Figure 6. PR alpha brainwaves

We can see a stark difference in α waves between 2D and 3D. For the former, we find that α waves are partially active in the frontal, temporal and occipital lobes. In contrast, there are almost no α wave activity throughout the brain except for some limited activity in the frontal lobe when watching 3D promotional images. In conclusion, 2D promotional images are less complicated than 3D and therefore demands less concentration from the viewers.

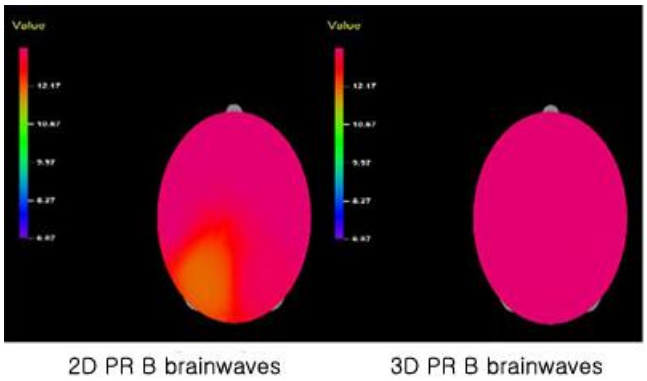


Figure 7. PR beta brainwaves

For promotional images, we can see active β waves regardless of their dimensional format. Whether it is 2D or 3D, viewing promotional images requires both concentration and continued tension. 3D promotional images activate β waves in

almost all parts of the brain. The same can be said about 2D images although there was a relatively lower level of β waves in the occipital lobe.

2.4 Result Summury and Discussions

2.4.1 Summary of result

In this study, we conducted an experiment in which we observed differences in brainwaves (α and β waves) when subjects were shown the same image in 2D and 3D format. The first type of material, sports footage, features rapid transition of frames and actual images of a real basketball match, so viewers are able to experience a great deal of liveliness and speediness even in 2D format. This experience is further enhanced when converted into 3D format, so viewers feel a sensation of being at the scene of the image and receiving all kinds of sensory stimuli at the same time, and will feel as if the footage is a part of reality. Accordingly, brainwave levels were higher when viewers watched the 3D footage.

The animation was of high quality in both formats, so viewers showed a great deal of engagement even when in 2D. The animation also had rapid frame transition in addition to the cheery and lively sound features. In 3D, the images are even livelier as the characters seem to pop out of the screen and come into our reach, thus viewers engagement is even stronger. Viewers may feel as if the images are part of the real world. Due to these characteristics, we could see different brainwaves for each channel.

Lastly, the promotional image had dull sounds and the least number of frame transition among the materials. Consequently, the 3D format can generate an advertising effect by facilitating the strength of stereoscopic images, but many of the scenes were still rendered identical to that of the 2D version. This led to the lowest level of engagement and sense of reality on the part of the viewers when compared to other 3D pictures. Brainwave levels were also low with not much distinction in α and β wave activities.

2.4.2 Study Implication

It is possible to project the result of our experiment on the basis of the features of our materials and existing studies. β waves are prominent when we perform complex and planned thinking and is mostly observed in the prefrontal, frontal and temporal lobes of the brain. In contrast, α waves are seen most when we are sleeping or relaxed and are relatively found most in occipital lobe. Therefore, β wave levels were relatively higher than that of α waves in channels 1, 2, 5, and 6 and also in channels 3 and 7, which are effected by 3D stereoscopic images.

However, α wave activity was somewhat high in channels 3, 4, 7 and 8. In channels 1, 2, 5 and 6, which corresponds to prefrontal and frontal lobes that are responsible for complex thinking, α waves were less prominent. In addition, we took into account space perception, sense of reality, presence and level of viewer engagement to assume what type of material demands the most complex and planned thinking, and measured brainwaves while the subjects viewed the material in 2D and 3D. In general, the level of α waves observed when watching flat images was higher than that found when viewing stereoscopic images and vice versa for β waves.

When analyzing the types of materials, promotional images had significantly lower presence, sense of reality and engagement than animations and sports materials. For promotional images, α waves were more active than β waves, and channels 3, 4, 7 and 8, which corresponds to α waves, showed relatively higher results. On the other hand, sports and animation both have extremely high level of engagement, sense of reality and presence. Therefore, they both provide a greater sense of liveliness and in 3D format and viewers are required to perform complex thinking in the process of cognition to perceive the depth. Accordingly, β waves were active and the results for channels 1, 2, 3, 5, 6 and 7 were high.

This study explored receivers' human cognition fact variables and image reception by measuring brainwaves, and its results may be applied to future experiments, physiological response studies and social psychological approaches. It may also be utilized to document the trend in the development of cognitive media with the advent and advance of diverse media. This involves how the sense of depth and reality, spectacle, sound, image, color and motion are converged and delivered to our sensory organ and how the receiver responds.

Furthermore, the human's cognitive understanding of 3D and the spectacular ways of expression by each genre and the establishment of a database of new media are areas with new dimensions for research. Such studies can provide valuable material for selecting characters and movements appropriate for stereoscopic effects, and determining genre or program production purpose best fit to materialize such contents.

3. ACKNOWLEDGMENTS

This study is supported by Samsung Academic Supports.

4. REFERENCES

[1] Barfield, W., & Weghorst, S.(1993). The sense of presence within virtual environment: A conceptual framework. Proceedings of the Fifth International Conference on Human-Computer Interaction. 5. 699-704.

[2] Emoto, M., Niida, T., & Okano, F. (2005). Repeated vergence adaptation causes the decline of visual functionsin watching stereoscopic television. Journal of Display Technology, 1. 328~340.

[3] Cho, Eun Joung, Kwan Min Lee & Yang Hyun Choi (2011). Effects of Stereoscopic Movies: The Position of Stereoscopic Objects (PSO) and the Viewing Condition. AEJMC 2011 Boston Confernece Paper.

[4] Cho. Eun Joung, Kweon, Sang Hee, Cho, Byungchul. (2010). Study of Presence Cognition Depending on Genre of 3D TV Images. Journal of Korean Broadcast. 24(4), 253-292.

[5] Choi, Tae Young. (2006). Effect for Brainwaves Changes by 2D Images and 3D Images. Korean Engineering Association Paper. 15(5), 607-616.

[6] Choi, Yang Hyun. (2009). Study of Work Flow of shooting technology for 3D images contents. Korean Broadcast Engineering Conference.

[7] Fukuta, T. (1990). New electronic media and the human interface. Ergonomics, 33(6), 687~706.

[8] Heeter ,C. (1992). Being There : The Subjective Experience of Presence. Presence : Teleoperators and Virtual Environments. MIT Press, fall, 1992.

[9] Heeter, C. (1995). Communication research on consumer VR. In Frank Biocca & Mark R. Levy (eds.), Communication in the age of virtual reality (pp. 191-218). Hillsdale, NJ: Lawrence Erlbaum Associates.

[10] Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks,

[11] M. S. (2008). Vergence-accommodation conflicts hindervisual performance and cause visual fatigue. Journal ofVision, 8(3):33, 1-30, http://journalofvision.org/8/3/33/, doi:10.1167/8.3.

[12] Ijsselsteijn, W.A., de Ridder, H., Hamberg, R., Bouwhuis, D., & Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. Displays, 18, 207-214.

[13] Ijsselsteijn, W.A. & de Ridder, H. (1998). Measuring Temporal Variations in Presence. Presented at the 'Presence in Shared Virtual Environments' Workshop. BT Laboratories, Ipswich, UK, 10-11 June.

[14] Ijsselsteijn, W. A., & Riva, G. (2003). Being there: Concepts, effects and measurement of user presence in synthetic environments. in G. Riva, F. Davide & W.A. IJsselsteijn(Eds.), Being there: the experience of presence in mediated environments, Amsterdam, The Netherlands; Ios Press.

[15] Kim, Tae Yong. (2003). Users' Traits Affecting Experience Possibility for Tele-presence. Journal of Korean Broadcast. 17(2)2, 111-141.

[16] Lang(1980). Behavioral Treatment and Bio-behavioral Assessment: Computer Applications. In J. B. Sidowski, J. H. Johnson, & T. A. Willians(Eds.), Technology in mental health cara delivery systems. 129-139. Norwood, NJ: Ablex.

[17] Lee, Hyung Chul. (2005). Step for Technological Development for 3DTV: 3D Human Factor. Journal of Korean Engineering Association. 13, 65-71.

[18] Lee, Hyung Chul. (2010). Study of Human Factor about measurement method of subjective fatigue in 3D Images. Journal of Korean Engineering Association. 15(5), 607-616.

[19] Lee, Ok Gi. (2009). The Study of Development of Measurement Scale for Presence. Journal Korean Journalism Informaion. 48.

[20] Livingstone, M., & Hubel, D.(1988). Segregation of form, color, movement, and depth : Anatomy, physiology, and movement, Science, 240. 740~749.

[21] Lombard, M., Reich, R. D., Grabe, M. E., Campanella, C. M., & Ditton, T. B. (1995, May).Big TVs, little TVs: The role of screen size in viewer responses to point-of-view movement. Paper presented to the Mass Communication division at the annual conference of the International Communication Association, Albuquerque, NM.

[22] Motoki Tosio, Yano Sumio. (201). 3D Images and Human Science : focusing on 3D Human Factor. Seoul, Jinsam Media.

[23] Neuman, W. R, Crigler, A. C., & Bove, V. M. (1991). Television sound and viewer perceptions. Proceedings of the Joint IEEE/Audio Engineering.

[24] Nielsen, J, (1993). Usablity Engineering, Academic Press, San Diego.

[25] Seuntiens, Vogels, Keersop(2007).Visual Experience of 3D-TV with pixelated Ambilight. PRESENCE

[26] Park, Il Woo. (2010). Discussions and Solutions for accepting 3D Images rightly. Journal of Korean Video Culture. 135-168.

[27] Sheridan, T. B. (1992). Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments, 1(1). 120-126.

[28] Brain Structure http://blog.naver.com/rscaresystem

[29] http://cafe.naver.com/1inchstyle.cafe

[30] http://blog.naver.com/adhdclinic

Subjective Assessment of a User-controlled Interface for a TV Recommender System

Rebecca Gregory-Clarke
BBC Research and Development
Centre House, 56 Wood Lane
London, W12 7SB
+443030409643
rebecca.gregory-clarke@bbc.co.uk

ABSTRACT

Many personalised recommendation systems rely on information explicitly provided by users, such as a set of item ratings, on which they can base their recommendations. This information can often be difficult to elicit, and the overall aim of this research is to present a recommender interface for an online catalogue of TV content, which would encourage users to provide information about their viewing preferences.

In this study, a field trial was conducted to gather initial feedback about the design, before a working version can be implemented and tested. The results showed that people were generally positive about the interface as a whole, citing its simplicity, and its emphasis on their personal preferences. However, a number of possible improvements have been identified, and will also be discussed here.

Categories and Subject Descriptors

H.5.2 [Information Interface and Presentation]: User Interfaces – *evaluation/methodology, graphical user interfaces, interaction styles*; H.3.3 [Information Storage and Retrieval] Information Search and Retrieval – *information filtering, relevance feedback, selection process.*

General Terms

Design, Experimentation, Human Factors.

Keywords

Personalised recommendations, user interface, user experience.

1. INTRODUCTION

This paper presents the results of a qualitative study of a novel user interface for recommender systems. The interface is designed to work with a recommender which uses a 'binary' rating system. It was thought that this might be simpler and more meaningful to the user, rather than a more subjective form of rating such as awarding a number of stars. While there is some research to suggest that higher scale granularity can produce better performance from some recommenders [1], there are also some problems associated with them. YouTube[1] previously employed a 5-star rating system, but this was eventually replaced with a thumbs up/thumbs down system, when it was noticed that the great majority of ratings awarded were the maximum 5 stars [2]. In this instance it was concluded that, whilst people tended to rate videos they really like with a high rating, they will not bother rating videos that they don't like at all. Our hope is that the negative impact of a reduced granularity rating system will be offset by the increased use of the system. However this is a long-term research question, which will not be evaluated here.

In the interface presented here, users are able to drag and drop programmes from the TV catalogue into a 'like' or 'hate' box, the contents of which will dynamically change the list of recommendations presented to them. The user is also able to interact with the recommender by adding the recommendations themselves to either the 'like' or 'hate' boxes, in order to refine and improve their recommendation set. Preliminary research had suggested that, for certain rating-based recommendation algorithms, this iterative method of refining the recommendations could produce good results, and that providing some negative feedback (i.e. programmes that the user doesn't like) was found to be beneficial. We therefore wanted to consider a user interface that would make this possible.

A printed mock-up of the interface was shown to a number of users in a field trial, the results of which will be discussed here, along with suggestions for improvements.

1.1 Research Questions

The main questions we wish to address in this study are given below:

1. What do users think of the interface as a whole (including the drag and drop function)? How easy is it to understand?
2. What do users think about adding programmes to a like/hate box?
3. What do users think about the general principle of personalised recommendations?

1.2 The Field Trial

A total of ten field trial participants took part, which was considered to be a suitable number for this form of qualitative usability testing, which is designed to provide preliminary feedback in order to make iterative improvements to the design. There was a 1:1 male to female ratio, and a good spread of ages from 18 to 65 amongst the participants. All considered themselves to be TV enthusiasts and watched an average of at least 2 hours of TV per day, and all were familiar with on-demand/catch-up services. The users were given a brief introduction to the trial, before being shown a mock-up of the interface and given some time to look it over. The users were invited to give their initial feedback as well as to ask any of their own questions about it.

[1] www.youtube.com

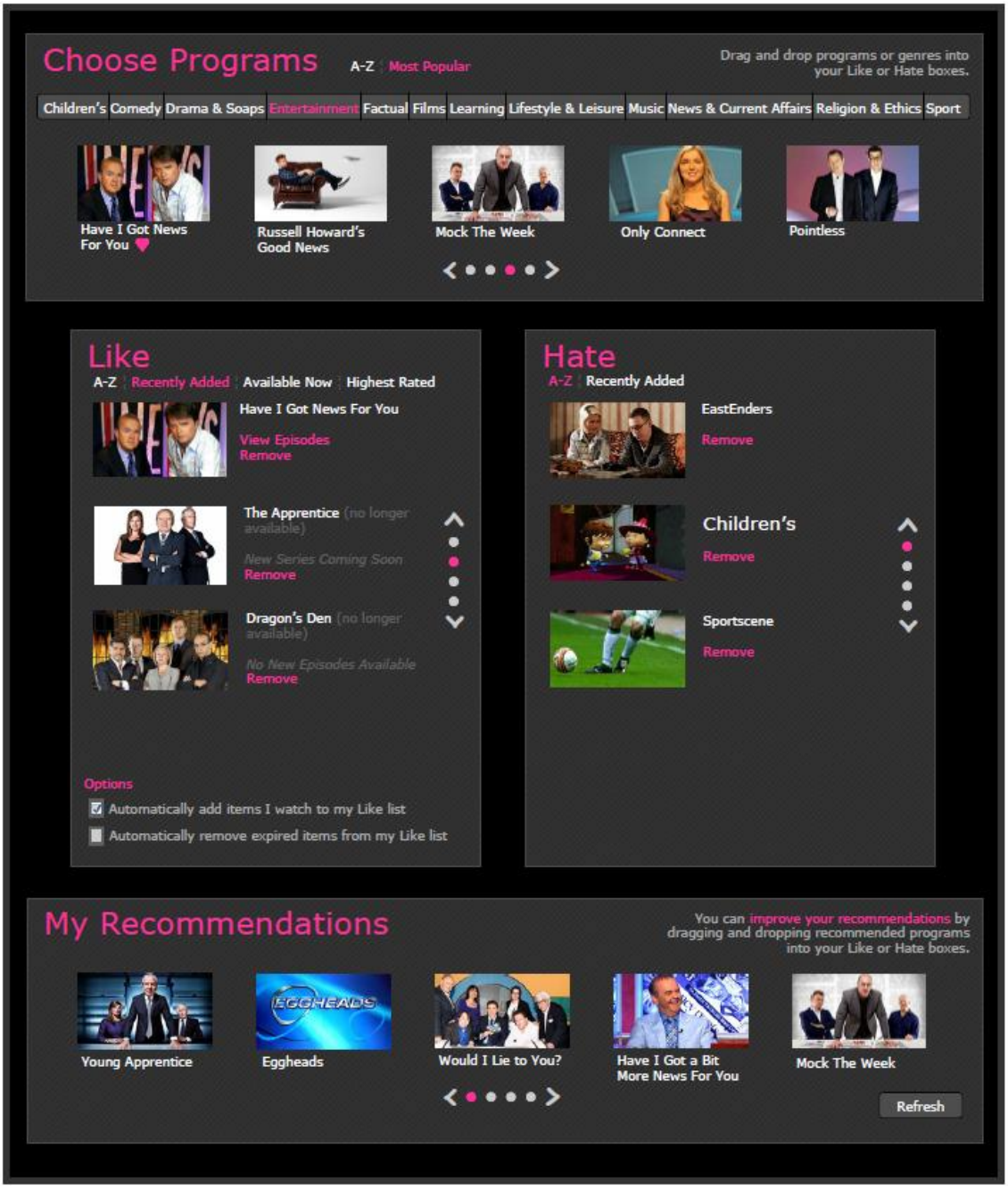


Figure 1. User Interface Shown to the Field Trial Participants

After this, the users were asked some more targeted questions about each of the features of the interface. All questions were based on the research questions described above, and left as open-ended as possible so as not to inadvertently influence the users' opinions.

2. OVERVIEW OF THE PROPOSED USER INTERFACE

A core idea of the design was to make the recommender interface an *interactive* and *transparent* experience for the user. We did not simply wish to present a 'black box' recommender in which the user has little control or understanding of the inputs and outputs of the system, as there is some evidence to suggest that transparency within recommender systems helps to improve user confidence in the system [3].

It is important to note that the information gathered through the interface is intended only to improve the user's private experience of the service, i.e. we were not considering the case where any of the information would be made public or used for social media applications. The key features of the interface are as follows:

- A section called *Choose Programmes* allows users to find and select programmes from the TV catalogue.

- Users can drag and drop programmes from the catalogue to either a *Like* or *Hate* box, to indicate whether they like or dislike programmes.
- *My Recommendations* is a set of personalised recommendations which automatically updates as the user adds programmes to the *Like* or *Hate* boxes.
- The user can say whether they like or dislike any of the recommendations by dragging them from the *My Recommendations* section, directly back into the *Like* or *Hate* boxes (this information can be used to refine and improve the set of recommendations).
- The *Like* box also has some extra features which may be of benefit to the user, which allows them to easily find and keep track of their 'liked' programmes (see Section 3.7).

A more detailed discussion of the key features along with the observations made during the field trial is given in the following section.

3. DETAILED DISCUSSION OF THE INTERFACE FEATURES AND FIELD TRIAL OBSERVATIONS

3.1 Like and Hate Boxes

The name 'like' is a familiar and broad term that was chosen to encourage people to include a large range of programmes that they enjoy watching. The name 'hate' was chosen to be deliberately strong, as it was expected that users would use this box much less than the Like box, reserving it mainly for a few programmes that they feel very strongly about. It was also thought that it might attract more interest than a milder word such as 'dislike'.

➔ There were some mixed opinions about the idea of 'disliking' programs. Some quite liked the idea of being able to say which programs they didn't like, referring to the 'thumbs up/thumbs down' rating system on some Personal Video Recorders (PVRs). On the other hand, some didn't see the point of the hate box as they assumed that their dislikes would be implicit from their likes.

➔ A few also mentioned that they liked the idea that these likes and dislikes might apply to the whole of the TV on-demand website. For example, if you have said you dislike something, then it will not appear prominently on the homepage, or might only appear at the end of the list when searching by category.

➔ Most agreed that they would not use the Hate box as much as the Like box, and several people suggested that it might be good to hide the Hate box away once they had finished using it, as they "don't need to see it" all the time.

➔ There was an almost unanimous feeling that 'hate' was "too strong", and a word such as 'dislike' would be better.

➔ Two trialists said they liked the idea of a 'middle ground' option in addition to Like and Hate, but most said they thought that would confuse things, and keeping it to two options would be simpler.

3.2 Series-based Interfaces

This interface would enable the user to add programmes to their Like or Hate boxes by series, however they may also select a whole genre if necessary, if they wish to filter it out entirely.

It should be noted that most online TV services present programmes by episode rather than by series. In the context of collecting information about a user's programme preferences we felt that a series based method would be more practical for the user, however it should still be easy to navigate to individual episodes using the interface.

3.3 Automatically Updating Recommendations

The *My Recommendations* section should update dynamically as items are added to the Like or Hate boxes. It was thought that this would help to emphasise the relationship between the user input and the recommendations, and thus make the system more transparent. The user should see the recommendations improve as they add more items, highlighting the benefit of the system.

➔ Generally the users thought that this made good sense, and was fairly intuitive.

3.4 Interactivity with Recommendations

As we have described, the user is able to fine-tune the recommendations by dragging them into the Like and Hate boxes.

Preliminary research had also indicated that people can have a very strong reaction against recommendations that they deem to be completely unsuitable. It was thought that giving the user some control over the recommendations may improve user satisfaction, as well as improving the recommendations.

Again, it is hoped that the transparency of the system may be improved by dynamically updating the recommendations as the user drags programmes from the set of recommendations into the Like and Hate boxes.

➔ The users thought this was fairly intuitive, with some making reference to other websites where you can indicate which recommendations are no good.

➔ A few trialists said they found the label "My Recommendations" a bit confusing, as they weren't sure if it meant programmes that they were recommending to others, or that the system was recommending to them.

3.5 Drag and Drop

The drag and drop method was thought to be a more enjoyable and interactive user experience than an alternative based simply on clicking buttons. It was also thought that dragging and dropping was particularly suitable for the feature which involves adding programmes to the boxes directly from the *My Recommendations* list, as it emphasises the link between the Like and Hate boxes and the set of recommendations.

➔ Most people liked the idea of a drag and drop user interface, however a few said they preferred to have the option of a button to click on instead, especially if they were not using a tablet or other touch screen device.

3.6 Incorporating Explicit and Implicit User Information

Whilst this interface is primarily designed to collect information *explicitly* provided by the user, there is the possibility that this system could incorporate *implicit* information as well. For example, by default the system could automatically add all items that the user has recently watched to the Like box. While the user can opt out of this if they wish, it has the advantage that the

recommender system will have something to work with, even without explicit feedback from the user. It is also a transparent way of showing users the items in their recent history, and may be another useful way for users to easily find programs that they watch regularly.

➔ The majority thought this would be very useful.

➔ Three of the users said they would like to be given a prompt at the end of the programme, giving them the option to add it or not.

➔ One user suggested that, whilst this is useful, the Like list might become very long, and so it would be good to keep this 'history' separately within the Like box, perhaps under a different tab.

3.7 Extra Features of the Like Box

As the Like box is essentially a record of programmes a user likes to watch, it follows that it could also be used as a way for the user to easily keep track of, and navigate to, their favourite programmes. In the interface presented here, this might include features such as:

- Being able to view the Like list in whichever way is most convenient (e.g. according to what's currently available, recently added, alphabetically etc.)
- Viewing all available episodes of a 'liked' series
- Being informed of any upcoming new episodes of a series (even if it is not currently available)
- Choosing whether to have expired items automatically removed from the list, or kept for future reference

These extra features were thought to be important when addressing the problem of cost versus benefit to the user. We want to maximise the advantages of the system to the user, and it was thought that combining the functionality of both a 'favourites' section and the recommender system could be a good way to address this.

➔ Most trialists thought these extra features were very useful. One user described it as like "a massive compendium of stuff you're going to watch, like a great diary."

4. CONCLUSIONS

The main conclusions with reference to the research questions are summarised below.

1. What do users think of the interface as a whole (including the drag and drop function)? How easy is it to understand?

Generally, the trialists were very positive about the interface, particularly with respect to its "simplicity" and the fact that it "saves time and effort". They liked the fact that it can be used as a method of "customising" your experience, and "reducing clutter". It was generally felt that it was fairly clear what the purpose of the interface was, and how to use it. The drag and drop idea was fairly intuitive to most, although a few also liked the idea of an option to click a Like or Hate button instead.

2. What do users think about adding programmes to a like/hate box?

People thought that the Like box was very useful, although it was thought that the Hate box would not be used as much, as people are unlikely to continue adding to it continuously - unless, for example, they spot something completely unsuitable in their recommendations.

There was a fairly unanimous feeling that the word 'hate' was too strong, and a word such as 'dislike' would be better.

3. What do users think about the principle of personalised recommendations?

The majority of trialists were very positive about the idea of personalised recommendations, as they saw it as a more customised approach to browsing online TV content. The general feeling was that users would like to see a more personalised approach to navigating online TV content than is currently available. Users were generally positive about any system which made their browsing experience simpler or more bespoke, and would therefore save them time and make it easier to find the programmes they were interested in.

4.1 Improvements to the interface

The following improvements are suggested for any future iterations of the interface:

- Preserve the parts of the interface that were most positively received, particularly its simplicity, and the prominence given to the user's own preferences.
- Include a separate 'viewing history' section of the like box to allow users to differentiate between items they have actively added to the box, and those that have been added because they have recently been watched.
- A milder, less polarising name such as 'dislike' would be better for the Hate box.
- The Hate box could feature less prominently, or have the option to hide it away once the user had finished with it.
- A different name for the *My Recommendations* section could be considered, as this name caused some confusion.

5. ACKNOWLEDGMENTS

Thanks to Libby Miller and Vicky Buser from BBC Research and Development for their help and support throughout the field trial. Thanks also to Maxine Glancy and Chris Newell from BBC Research and Development for all their helpful advice.

6. REFERENCES

[1] Cosley, D., et al., Is Seeing Believeing? How Recommender Systems Influence Users' Opinions. 2003. In Proceedings of *CHI '03 Human Factors in Computing systems*

[2] The Official YouTube Blog, 22 September 2009 http://youtube-global.blogspot.com/2009/09/five-stars-dominate-ratings.html

[3] Sinha, R. and Swearingen, K. The Role of Transparency in Recommender Systems. 2002. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* DOI= http://dx.doi.org/10.1145/506443.506619

MosaicUI: Interactive media navigation using grid-based video

Arjen Veenhuizen
TNO
Brassersplein 2
Delft, The Netherlands
+31 8886 61168
arjen.veenhuizen@tno.nl

Ray van Brandenburg
TNO
Brassersplein 2
Delft, The Netherlands
+31 8886 63609
ray.vanbrandenburg@tno.nl

Omar Niamut
TNO
Brassersplein 2
Delft, The Netherlands
+31 8886 67218
omar.niamut@tno.nl

ABSTRACT

Intuitively navigating through a large number of video sources can be a difficult task. The problem of locating a video asset of interest, whilst keeping an overview of all the available content arises in video search and video dashboard application domains, utilizing the new possibilities provided by recent advances in second screen devices and connected TVs. This paper presents initial results of an attempt to solve this problem by creating real-time mosaic-like video streams from a large number of independent video sources. The developed framework, called MosaicUI, is the result of cooperation between the European FP7 project FascinatE and the Dutch Service Innovation & ICT project Metadata Extraction Services. Using the combination of an intuitive user interface with a media processing and presentation platform for segmented video, a new method for user interaction with multiple video sources is achieved. In this paper, the MosaicUI framework is described and three possible use cases are discussed that demonstrate the real world applicability of the framework, both within the consumer market as well as for professional applications. A proof-of-concept of the framework is available for demonstration.

Categories and Subject Descriptors

H.5.1 [Information Interfaces And Presentation]: Multimedia Information Systems – *video navigation, video search, immersive media, interactive media, metadata extraction*

General Terms

Experimentation, Verification.

Keywords

Natural user interfaces, second screen, TV, mosaic, video search

1. INTRODUCTION

With an increasing number of video content available to viewers on multiple screens, it becomes more difficult to find videos of interest. Furthermore, to keep videos accessible to all viewers on all their devices, semantic cue or metadata-based access has become a necessity. Several content-based video retrieval systems or video search engines, that enables a user to explore large video archives quickly and with high precision, have already been developed. Janse et al [1], were one of the first to present a study on the relationship between visualization of content information, the structure of this information and the effective traversal and navigation of digital video material. Currently, the MediaMill Semantic Video Search Engine [2] is one of the highest-ranking video search systems, both for concept detection and interactive search. Most of these systems include a video browsing or navigation interface for user interaction. For example, both the Investigator's Dashboard and the Surveillance Dashboard prototypes, developed in the MultimediaN project [3], use video browsing as a method to assist users in their video-related tasks.

TV viewers face a similar task when trying to find content to their taste in an ever-expanding offer of TV channels and on-demand content. Barkhuus et al. [4] describes how techniques have changed the experience and planning of TV watching. The active choosing of what content is to be watched resembles other types of media consumption such as reading, listening to music, or going to the cinema. However, interactivity in TV deployments has not yet seen significant advances, as one of the key issues that need to be resolved is how to interact with and navigate through the content. Most of the solutions available so far have been based on advanced remote controls and hierarchical menus. Natural user interfaces and second-screen applications are seen as promising candidates for a next step in TV interaction, adding a new dimension and introducing new possibilities to navigating video content.

In this paper, we describe the concept of MosaicUI as a framework for interactive video browsing and navigation using, for example, a tablet or smartphone. A number of video sources are combined into one large grid which is displayed on a screen. The combination of high resolution, multi-layer, spatially and temporally segmented video with a state of the art metadata search engine and associated media content creates a multitude of new interactive video applications, which enable the user to interactively navigate videos, selecting the one of interest and exploiting the new possibilities in media navigation introduced by the latest advances in media interaction. Video selection can be performed in numerous natural ways, e.g. motion tracking or gestures.

MosaicUI finds its origin in results from two ongoing projects. The Metadata Extraction Service (MES) [5] project aims at near real-time extraction and indexing of multi-modal metadata from live and stored multimedia content (video, audio, text, images, etc.). Metadata extraction includes, but is not limited to, video concept detection, extraction of text embedded within an image or video (OCR) and speech recognition. Source multimedia content is stored using a long term storage array, and by combining this archive with a metadata index which can be queried, one is able to search for content in a multimodal manner.

Within the EU FP7 project FascinatE [6], a capture, production and delivery system capable of supporting pan/tilt/zoom (PTZ) interaction with immersive media is being developed. End-users can interactively view and navigate around an ultra-high resolution video panorama showing a live event, with the accompanying audio automatically changing to match the selected view. The output is adapted to their particular kind of device, covering anything from a mobile handset to an immersive

panoramic display. The FascinatE delivery network uses spatial segmentation and tiled streaming to enable interaction on mobile devices.

In the remainder of this paper, the MosaicUI framework is explained in further detail and its application domains are discussed.

2. MOSAICUI FRAMEWORK

The potential synergy between the MES and Fascinate Project has been explored by developing the MosaicUI framework. The capabilities of the two initially unrelated platforms are combined to form a new platform enabling a number of new use cases. For example, one could create a grid of live TV shows and display it on a HDTV, while the end user interactively browses and selects his program of interest on a second screen (e.g. a tablet). Alternatively, a number of CCTV feeds could be joined together into a video grid in real time. Another example would be to take the MES metadata search engine into the equation, enabling interactive high resolution multimedia content search and navigation, potentially using a second screen. Figure 1 shows the basic concept of a MosaicUI video grid.



Figure 1. A mosaic of multiple video sources.

In order to achieve the desired level of user interaction and allow for each of these use cases, a number of high-level requirements have been identified. Besides the ability to support a wide variety of media sources, such as live TV, internet videos, local videos, images and segmented content, a controller interface is required to navigate the content. Most importantly, a system is required which is able to combine the different sources together to a grid-like live video feed in real-time, which can then be streamed to a play-out device. Based on these use cases, two, largely equal, architectures have been developed, as shown in Figure 2 and 3. These architectures differ from one another in that the first architecture uses client side media processing, while the second architecture uses server side media processing. This allows the MosaicUI framework to be used both on devices that have a limited amount of processing power available as well as on more powerful devices, creating a versatile platform which can be utilized on numerous types of devices and which introduces new dimensions to the previously explored possibilities of mosaic based video content navigation.

2.1 The MosaicUI Functional Components

From a functional point of view, five key components and one optional component have been defined that together constitute the MosaicUI architecture. First, one or multiple media sources are required. In the MES system, this source can be any form of media, e.g. a live broadcast, a local audio file, a YouTube video or CCTV capture stream. Second, this media source must be indexed in some way so that the user is able to "find" and retrieve that specific media source. Indexing could be for example a form of metadata which describes the geographic location of a CCTV capture stream, a transcript of the closed caption of a broadcast item or the tags associated to a specific YouTube video. A metadata database is required to store and index this information. This enables one to actually find the media that is of interest to the user. It is envisioned that this metadata index could be a multi modal platform and could be based on existing (metadata) databases (e.g. YouTube). The current implementation is able to use both the MES database and a set of TV channel listings. Third, a user interface is required to control the play-out of the mosaic and to allow the user to adapt the mosaic to his liking. A controller facilitating this interaction (scrolling, zooming, fast forwarding/rewinding the content, but also changing the selection of the sources being shown in the mosaic) can be implemented by e.g. a touch-based, gesture recognition or motion tracking platform. The current implementation features a touch-based controller application on a tablet. Fourth, a combiner is required which places the requested media sources together in a grid to form a single mosaic media stream with which the user can interact. The role of this combiner is to stitch together multiple media sources into one high resolution stream in real-time. The most difficult aspect of the combiner to control is its ability to react as fast as possible to changing user requests, in order to get an intuitive user experience. The current implementation of the combiner is largely based on the FascinatE tiled streaming platform described in [7]. Fifth, a play-out function is required. The play-out interface actually displays the mosaic video stream with which the user can interact. This combined media content could for example be shown on an HDTV, beamer, smartphone or tablet. Sixth, an optional stream server can provide the means to stream the combined media content to a client. This is required in case the server-side combiner architecture is used. As stated in the previous paragraphs, an important distinction in the potential use cases is the fact that some of them require that the combination of media sources is performed locally (e.g. client side), while others require server side combination of media sources. From a functional point of view these two use cases seem to be identical, but from an architectural point of view, a clear distinction can be identified. Figure 2 shows the MosaicUI architecture with a client-side combiner, while Figure 3 shows a server-side combiner.

In the current framework, the media sources, the metadata index and play-out are provided by the MES project while parts of the combiner and controller are based on work performed in the EU FP7 project FascinatE [6].

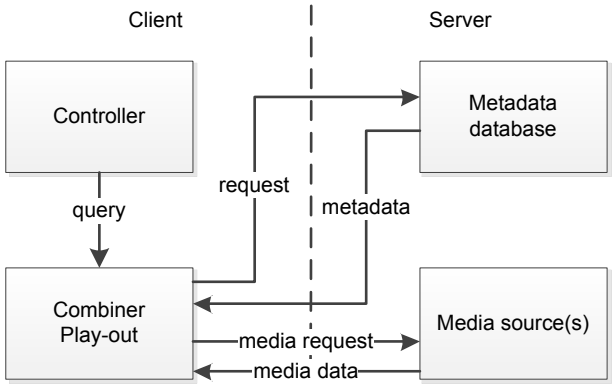


Figure 2. Client side media source combination.

2.2 Creating a mosaic video stream

The basic data flow in the MosaicUI framework can be described as follows. First the user requests a particular set of video sources using the controller. This request could for example be a fixed set of video streams or a search query to find a specific type of video. This query is sent from the controller to the combiner, which is either local or remote. Next, the query is relayed as a request from the combiner to the metadata database in order to look up the metadata describing the relevant media fragments and their location. This information is sent back to the combiner, which then requests the media streams from the identified sources.

Upon reception of the first frames of the media streams, the combiner starts the decoding process, running multiple parallel decoders. After a frame has been decoded, the combiner places it in the grid. In case a server-side combiner is used, the resulting grid frame is sent to an encoding process and streamed to the client. In case a client-side combiner is used, the resulting grid frame is sent to a video buffer for output on the display. It should be noted that in order for the recombination process to start, it is not necessary for the combiner to wait until the first frames of all video sources are available. The combiner can just add sources to the grid as they become available. Since the combiner needs simultaneous access to all video sources, the necessary bandwidth can become problematic. However, since the resolution of the resulting mosaic video will in most cases not be much larger than the full resolution of a single conventional video source, it is sufficient for the combiner to access a low-resolution version of each media source. Sources that are available in multiple resolutions, such as is often the case with adaptive bitrate content in the form of e.g. MPEG DASH or Apple HLS, are especially useful in this regard. A further method for limiting the bandwidth requirements at the combiner is by using FascinatE tiled streaming technology [7], which provides an efficient and scalable delivery mechanism for streaming parts of a high-resolution video – the video grid in this case.

3. APPLICATION DOMAINS

The MosaicUI architecture is suited for use in a wide variety of application domains. This section will examine the possibilities for MosaicUI in three of those domains: as a novel method for browsing TV channels; as an intuitive user interface for searching video and as a method for viewing multiple CCTV streams on a mobile device.

3.1 TV browsing with a second-screen

One of the more obvious applications of MosaicUI technology is as a method for interacting with, and navigating between, TV channels. In recent years several new types of TV user interfaces

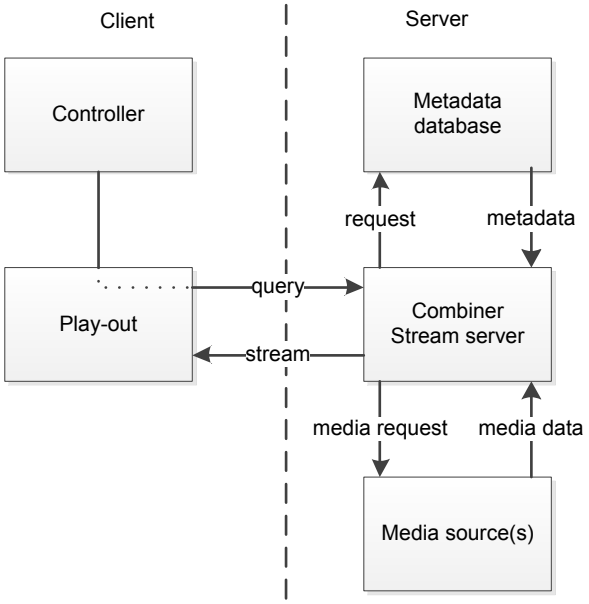


Figure 3. Server side media source combination.

have been proposed both in the academic community as well as in commercial applications: from touch-based interfaces to gesture recognition and voice recognition. The underlying principles of navigating between TV channels have not changed however. Instead of pressing a channel up or down button on a remote control, one might make an analogous gesture to a TV. And instead of pressing a specific channel number on a remote control, one might now shout the name of the channel to the TV. However, the essence of navigating between TV channels is still either linear (the channel up/down button) or directed (pressing '7' on your remote control). The only major innovation in this regard has been the introduction of the EPG, which allowed users to see an overview of the available programming before making a selection. MosaicUI can be seen as an evolution of this EPG. By giving users a visual overview of all available content, in this case TV channels, they can make a selection in a more intuitive way than by reading program titles from an EPG.

In figure 4, one can see an example of a second screen device being used to navigate through TV content. In this case the second screen device receives a mosaic video that incorporates the live video streams of all the available TV channels. By clicking on a particular channel, the second screen device instructs e.g. the TV to switch to the selected channel. Alternatively, the second screen device itself switches to a high resolution version of the selected channel, allowing the user to watch the particular channel on the second screen device instead of on the TV. Depending on the application, the selection of the TV channels that are included in a mosaic can be generic, such as a default mosaic including the 25 most popular channels, or personalized. In the personalized case, the selection of channels in the mosaic could for example be based on the user's viewing history or on a pre-selected list of channels. It is also possible to further configure the mosaic by for example including metadata information or channel logos either in the mosaic stream itself, or in the second screen application being used to display the mosaic. The second screen TV channel application might also be used to allow a personalized picture-in-picture stream. For example, a user might want to watch multiple channels simultaneously. In this case, he selects the desired channels from the mosaic (see the visual overlay in Figure 4), presses a button, and the second screen application sends the newly created video stream, consisting of the selected videos, to

the TV, for example through Apple Airplay. As discussed in the previous section, this newly configured mosaic stream can either be generated by a network-side process or as part of the second screen application.



Figure 4. Interactively selecting videos in a mosaic grid on a second screen device.

3.2 Interactive video search

Another possible application for MosaicUI is as a new method of displaying video search results. In most current video portals that allow for video search, e.g. YouTube and Vimeo, the results of a particular search query are displayed as text accompanied by either a static or animated thumbnail while navigating the results utilizes, again, a linear interface. This mostly text-based and static display of results can make it difficult for a user to assess which of the listed videos best matches with what he was looking for. The MosaicUI framework can be used to visually present the results of a particular search query, e.g. by including the 25 most relevant video results in the mosaic video. Upon seeing the results in their actual video form, a user would then be able to more quickly assess the results and either change his search query or select a particular video from the mosaic grid. This system could be extended with a function that works in a way similar to Google Instant, which updates your results continuously while you type. In a MosaicUI system, this would mean that the videos included in the mosaic would continuously be updated while the user refines his search query. Certain videos in the mosaic might be swapped out for others, while the position of those that are in the mosaic might change depending on their relevance. The most relevant videos might for example be placed in the middle of the mosaic. Also, the size of each tile in the grid could vary, depending on for example the relevance of that search result, content type or user specific criteria. Such an interactive video search system could support browsing video along multiple threads, as described in [8].

3.3 Video surveillance

One of the advantages of the MosaicUI framework is that it allows for simultaneous playback of multiple videos on devices that normally are not capable of doing so. Examples of such devices are smartphones and tablets that do not have the processing power for media decoding and are limited by a single hardware decoder. One application where this kind of functionality is useful is in the area of video surveillance. In this case, MosaicUI can be used to present security officials, working in the field, an immediate overview of a particular area or location by placing all relevant camera feeds into a grid and sending the resulting mosaic to e.g. a smartphone. By clicking on a particular video, the security official can quickly get a higher resolution version of a particular camera feed.

4. CONCLUSION AND FUTURE WORK

In this paper, a novel platform for navigating through multiple video sources has been presented. Three potential use cases, navigating between TV channels, intuitive video search and interactive video surveillance have been implemented, discussed and demonstrated. There are, however, a multitude of other applications and media sources which could be implemented by or connected to this platform. As part of the future work, we will look at the aforementioned combination of, and integration with, different (types of) media sources and experiment with different forms of interaction. Currently, a tablet has been functioning as user interface. In the future we will investigate the use of motion tracking and gesture recognition platforms (e.g. the Kinect) for interacting with the mosaic video. Connecting the platform to a platform like YouTube or Vimeo will be investigated as well. Preliminary results show that although technically possible, the respective APIs do not allow for a large number of concurrent connections. It is concluded that new intuitive user interfaces like multi-touch tablets, combined with the ability to create grid-like compositions of multiple video sources (optionally augmented with indexed metadata) creates a number of exciting new possibilities. In contrast to earlier research in the field of grid based media navigation, the utilization of new user interfaces greatly extends the flexibility and possibilities in media navigation. Application of these new possibilities proves not to be limited to over the top concepts, but to real world user centered cases as well with immediate usage in end-user environments and security environments, to name a few.

5. ACKNOWLEDGMENTS

Part of the research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 248138.

6. REFERENCES

[1] Janse, M.D., Das, D.A.D., Tang, H.K. & Paassen, R.L.F. van (1997). Visual tables of contents: structure and navigation of digital video material. IPO Annual Progress Report, 32, 41-50. http://www.tue.nl/publicatie/ep/p/d/144348/?no_cache=1

[2] Snoek, C. G. M.; van de Sande, Koen E. A.; de Rooij, O.; Huurnink B.; Gavves, E.; Odijk, D.; de Rijke, M.; Gevers, T.; Worring, M.; Koelma, D.C.; Smeulders, A. W. M. "The MediaMill TRECVID 2010 semantic video search engine". In Proceedings of the 8th TRECVID Workshop. Gaithersburg, USA, November 2010.

[3] MultimediaN Golden Demos http://www.multimedian.nl/en/demo.php. Visited March 2, 2012.

[4] Barkhuus, L.; Browns, B. "Unpacking the Television: User Practices around a Changing Technology". In ACM Transactions of Computer-Human Interaction 16, 3, September 2009.

[5] SII Metadata Extraction Services. http://www.si-i.nl/project.php?id=82. Visited March 2, 2012.

[6] FascinatE. http://www.fascinate-project.eu/. Visited March 2, 2012.

[7] van Brandenburg, R.; Niamut, O.; Prins, M.; Stokking, H.; , "Spatial segmentation for immersive media delivery," in Proceedings of 15th International Conference on Intelligence in Next Generation Networks (ICIN), 4-7 October, 2011.

[8] de Rooij, O.; Worring, M. "Browsing Video Along Multiple Threads," In IEEE Transactions on Multimedia, Vol. 12, No. 2, February 2010.

My P2P Live TV Station

Chengyuan Peng
VTT Technical Research Centre of Finland
Vuorimiehentie 3, Espoo
P.O. Box 1000, FI-02044, Finland
+358 20 722 6839
Chengyuan.Peng@vtt.fi

Raimo Launonen
VTT Technical Research Centre of Finland
Vuorimiehentie 3, Espoo
P.O. Box 1000, FI-02044, Finland
+358 20 722 7053
Raimo.Launonen@vtt.fi

ABSTRACT
Peer-to-Peer (P2P) live streaming has the potential to stream live video and audio to millions of computers with no central infrastructure. It enables people to broadcast their own live media content and to share the streams online without the costs of a powerful server and vast bandwidth. In this paper, we present an open source based live event broadcast injecting system using P2P live streaming technology. This PC-based system can distribute HD quality video and audio to user's desktop and laptop computers without using a central server. The low cost system could be applied to everything from family weddings and community elections to company seminars to local festivals. From system testing we conclude that P2P live streaming is able to overcome the limitations of traditional, centralized approaches and to increase playback quality.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information System**]: Video

C.2.2 [**Network Protocols**]: Applications (P2P)

C.2.4 [**Distributed Systems**]: Distributed Applications

General Terms
Experimentation, Performance, Verification

Keywords
Live broadcast, P2P live streaming, encoding, player plugin.

1. INTRODUCTION
Basically there are two alternative technologies for delivering live video to large scale end-users: Content Delivery Networks (CDN) and P2P systems or Hybrid CDN-P2P approach that incorporates the advantages and removes deficiencies of both technologies [3].

CDNs deploy servers in multiple geographically diverse locations and distribute content over multiple ISPs. User requests are redirected to the best available server based on proximity and server load etc. [3] [8]. In CDNs content is served by the hosting site to all connecting users, and it enables content providers to handle much larger request volumes. Thus, end-users are offered more reliable service and higher quality experience. However, for CDN, the hardware and infrastructure cost would be enormous.

Although current streaming technologies are considerably cheaper than either satellite or terrestrial TV broadcasting, they can still prove to be prohibitive for individuals or small companies. Companies need the necessary infrastructure to make sure the contents can be streamed to all of their users in good quality and uninterrupted [8]. Even popular sites and providers can be overwhelmed by unexpected surges in demand and thus have to deny service to end-users. With millions of potential users, the simultaneous streams of data will easily congest the Internet [3]. This scaling constraint becomes especially relevant as users and content providers demand higher quality video, that is, implying higher operating costs for the CDN.

In contrast to traditional client-server based streaming, a decentralized P2P live streaming system tries to eliminate the need for expensive central infrastructure and reduce latency by distributing the workload amongst peers viewing the same stream [1]. Instead of streaming content from one site to all connected users, P2P systems solve the scalability issue by leveraging the resources of the participating peers. Because the peer-based architecture takes advantage of the computing, storage and network resources of each user [6], the more users or peers a P2P network has, the better the quality of content distributed becomes. The video playbacks on all users are synchronized, i.e., every user in the swarm is viewing the same content at the same time.

P2P systems can achieve higher scalability while keeping the server requirements low. With this network, user can hear radio and watch television without any server involved. This reduces latency and network load and increases video quality for those watching the stream. P2P is able to provide efficient and low-cost delivery of professional and user created content. However, the decentralized, un-coordinated operation implies that this scaling comes with undesirable side effects. Especially P2P live streaming is still a big challenge that people have been trying to solve.

This paper is not about trying to solve the challenges faced by P2P live streaming but using the exist algorithms to build a live video and audio content injecting system. The rest of the paper is structured as follows: section 2 outlines the P2P live streaming protocol used in this paper. Section 3 describes the system components and section 4 presents system deployment and testing. Finally we draw a conclusion from our work.

2. P2P LIVE STREAMING PROTOCOL
Live streaming faces several challenges that are not encountered in other p2p applications such as file sharing. Live has to solve four problems simultaneously: low latency, high reliability, high offload, and congestion control [1]. The streaming content is required to be received with respect to hard real-time constraints and data blocks that are not obtained in time are dropped, resulting in a reduced playback quality. Additionally, a live broadcast ought to be received by all users simultaneously and with minimal delay. Another crucial property of any successful live streaming system is its robustness to peer dynamics. High offload is the fraction of data coming from peers instead of the source. This is what creates a low-latency, high-reliability stream,

and it only requires an upload capacity of five times the original bit rate on the original uploading machine [8].

The core P2P live streaming algorithms and pyhton code we used were from P2P-next project [5] [6] (used shadowed blocks shown in Figure 1). One of the key features is its custom overlay network (cf. Figure 1) implemented as a BitTorrent swarm with a predefined identifier [4]. It allows peers to exchange information using custom overlay communication protocols. At the same time, the overlay does not require more TCP sockets being opened in addition to regular BitTorrent sockets [2] [4].

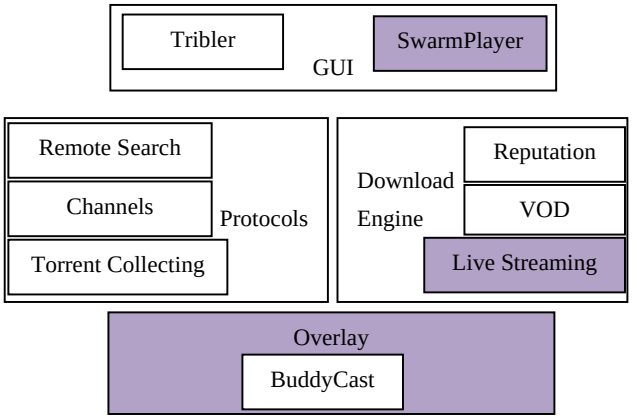


Figure 1. Tribler architecture [4].

A live broadcast can have an infinite duration that is the content to be broadcast is not known beforehand. The BitTorrent protocol assumes the number of pieces is known in advance. The P2P live streaming techniques from Tribler extended BitTorrent design to support infinite duration by means of a sliding window which rotates over a fixed number of pieces [2].

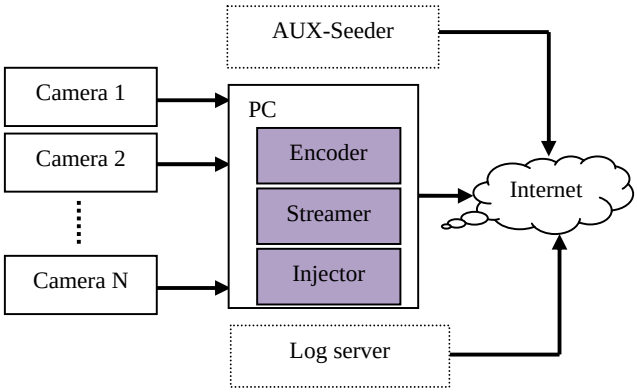


Figure 2. Injecting system architecture.

On source authentication aspect, in the BitTorrent protocol, hashes for pieces must be encoded and embedded in a metafile called .torrent file. However for live streaming the data is not known in advance and the hashes can therefore not be computed when the torrent file is created. The lack of hashes makes it impossible for the peers to validate the data they download, leaving the system prone to injection attacks and data corruption [2]. To prevent data corruption, they use asymmetric cryptography to sign the data, which is superior to several other methods for data validation in live P2P video streams [2].

The interested readers can find more detailed information from the specification documentation [2].

3. SYSTEM DESIGN

The system is comprised of the two components: an injector (cf. Figure 3) which sends the live video and audio to the Internet and a player (cf. Figure 4) which users can watch the live broadcast sent from the injector. The Injector includes cameras, desktop PC (including encoder, streaming and seeder) (see Figure 2). In Figure 2 we also setup two optional modules AUX – Seeder and Log server which we will discuss the utility later in this section. The cameras and a PC which hosts encoder and seeder software module are called “My P2P Live TV Station”.

3.1 Camera Module

Many live devices can be connected to the system as long as there is enough computer power. In our system, we attached two cameras to the seeding PC (cf. Figure 2) in order to switch between different views. Each camera is connected to the Blackmagic design’s Intensity Pro PCI card embedded in the injector PC via a HDMI cable. Intensity Pro PCI card features HDMI components and NTSC/PAL capture. It can capture every bit of every pixel in the full resolution HDTV uncompressed video signal.

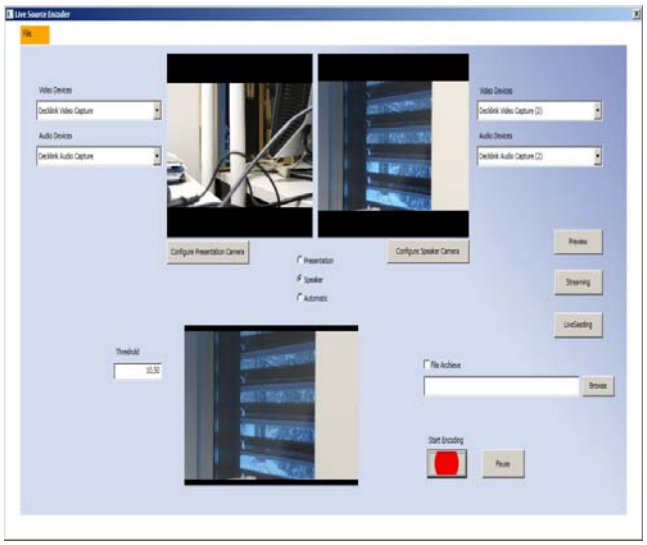


Figure 3. User interface of injector.

3.2 Encoder Module

The first step is to make our video look as great as possible. With the Microsoft Expression Encoder and provided SDK we developed a customized encoder which can access to all the tools and features to create encoded video and audio that will meet our streaming and live broadcasting needs. The Expression Encoder object model (OM), which is based on the Microsoft .NET Object Model Framework, is designed to make the functionality of Expression Encoder available without having to write much code. Figure 3 is a screenshot of the user interface of our P2P live TV station. The upper part is the customized encoder (to see clearly enlarge the PDF file to 150%).

3.3 Streamer Module

The encoded video and audio are then published by local host to the streamer (in our case VLC player in the same PC as the Encoder) where it is transcoded and encapsulated in MPEG transport stream as required by the Tribler P2P live streaming protocol. The live video is encoded in VC-1 and streamed in the H.264 codec. Adopting H.264 codec is in large part because of its broad support base. The streamer then send http streamed video and audio the injector.

3.4 Injector Module

The injector (or seeder) software module catches http streamed live source from the streamer and creates a metafile which gives the peer the IP address of the tracker (the injector PC) to verify downloaded parts of the live source and the peer then contacts the tracker to obtain a list of peers currently watching. Then it injects the pieces of the live source to the network indefinitely. It creates a metafile called for example live.ts.tstream which is used for the player as an input to playback the live broadcast.

The AUX-Seeder shown in Figure 2 is an optional module. It means that the system is still working very well without it. The reason to have this auxiliary server is that in BitTorrent based swarms, the presence of seeders significantly improves the download performance of the other peers (the leechers) [2]. However, such peers do not exist in a live streaming environment as no peer has ever finished downloading the video stream. The AUX-Seeder in a live streaming environment acts as a peer which is always unchoked by the injector and is guaranteed enough bandwidth in order to obtain the full video stream.

The injector controls the AUX-Seeder. The AUX-Seeder can provide its upload capacity to the network, taking load off the injector. It behaves like a small CDN which boosts the download performance of other peers as in regular BitTorrent swarms. In addition, the AUX-Seeder increases the availability of the individual pieces. All pieces are pushed to all seeders, which reduces the probability of an individual piece not being pushed into the rest of the swarm.

The AUX-Seeder is connected to the injector which has the IP address and port number of the AUX-Seeder representing trusted peer which are allowed to act as a seeder. The identity of the seeder is not known to the other peers to prevent malicious behavior targeted at the AUX-Seeder.

3.5 Live Playback

For playing back a live stream sent from the injector, users need to download and install an executable file to their system for either Internet Explorer or Firefox browser. The software simply works in the background to facilitate data transfer and doesn't allow any configuration. Video streams display in the browser via customized VLC player plugin. Neither user's computer nor user's browser will have to be rebooted.

Before watching streaming video in a browser, a peer first needs to obtain the live.ts.tstream metafile which is created by the injector. The P2P live streaming player is embedded in a HTML file via javascript. For Internet explorer browser, for example, we use the following code:

```
<script type="text/javascript">
document.write('<object classid="clsid:1800B8AF-4E3343C0-
AFC7-894433C13538" ');

document.write('width="640" height="480" id="vlc" name =
"vlc" events="True" target="">');

document.write('<param name="Src" value =
"file:///C:/live.ts.tstream" />');
</script>
```

Figure 4 is a screenshot of the live playback in the Internet explorer browser. We have customized and implemented specific statistic functionality into the Log server (cf. Figure 2).

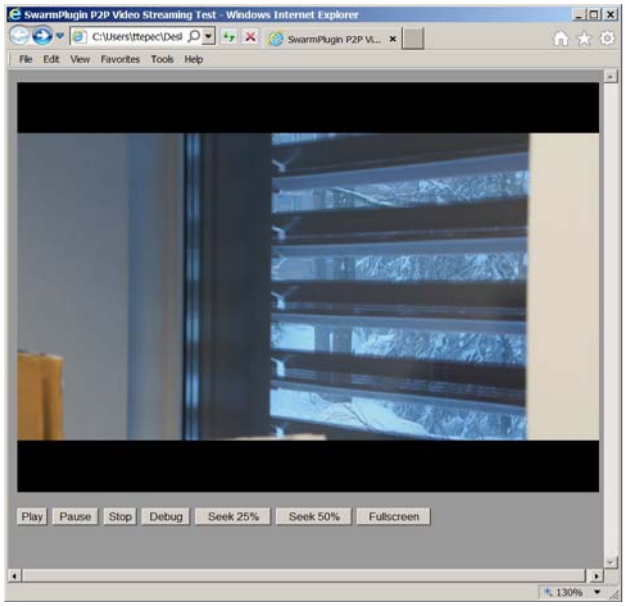


Figure 4. Screenshot of live playback.

The live stream downloading and uploading was run as a background process. Seeding is supported through a limited cache which is normally transparent to the user. Manual control over the seeding is available using a separate GUI built in the background process, where users can stop and start the torrent as well as modify advanced settings like the upload speed limit. It depends on the bitrate and the average network connection of a peer, viewers with high bandwidth connections and modern computers can experience up to full HD 1080p quality streaming in terms of latency and bandwidth.

3.6 Log Server

We have also setup a simple log server which is actually a normal web server in order to collect some information from users' computers and to know the viewing experience with the live system. The most important information we collected included maximum download and upload speeds, how many viewers watching the live stream, etc. According to these data we can judge the scalability of the system.

4. DEPLOYMENT AND TESTING

We have successfully deployed and tested the system in many occasions, ranging from companies' internal seminars, demos and meetings to outside formal academic events. It was capable of streaming live video across thousands of computers without any central infrastructure.

Watching live seminar was made available with our Intranet connection for all company's colleagues who have various terminals (PC, laptop) by sharing their bandwidth through enabled web browsers. The users experienced a live HD quality video and good sound without interruptions (cf. Figure 4). The tests were with very high audience satisfaction and resulting in significant server infrastructure savings. Streaming P2P has the potential to serve audiences at the scale of broadcast TV. Using our system, we can serve 10,000 online users a home broadband connection. However, we haven't found a big event which can catch up thousands of users.

In order to present our results to see how good our results are, Table 1 summarizes some key injecting and encoding parameters as well as software and hardware configurations during our live tests for reader reference.

Table 1. Parameters of the injecting system

Bit Rate	Resolution	Injector	AV Codec	Swarm Player Plugin
2190 kbps	1280x720	Tribler Source Code	H.264 mp4a	Tribler Source Code
Camera	**Capture Card**	**Encoder SDK**	**Piece Size**	**PC**
Sony NEX-VG10	Blackmagic design Intensity Pro PCI	Microsoft Expression Encoder 4 Pro SP2	64k	Intel Core i7CPU 3.33GHz

5. CONCLUSIONS

This is a very practical and useful system where live events need to be broadcast from our deployment. And it is physically portable which means that it is easy to setup the system on site. Referencing our system design, you can build your own P2P based TV or radio stations with minimal resources and enabling higher bit rates and millions of potential users.

We have our sights set on using the system for many events, for example, education, seminars, home parties, conferences, festivals and sports, etc. The approach is not limited to computers. They can be ported to various terminals such as Android mobile devices [7] and P2P set-top box which are already available [5].

There is a need to push the limits and see what happens when large numbers of users join a live swarm. It is not easy to find massive live events and willing users to test the scalability of the system. The hope is that we will find some events and large users in the future and do a large-scale test of this P2P streaming technology.

Another unsatisfied result is that the system introduced more delay. The total delay on the whole broadcast chain (from capturing, encoding, transcoding and streaming, to injecting) was in about 15 seconds range. This is not acceptable live sports event where fans cheer up the victory.

Reducing the latency introduced by the system is the most important research direction and challenging area especially on P2P live streaming protocol and algorithms.

6. ACKNOWLEDGMENTS

We wish to acknowledge the support from P2P-Next project partners especially Arno Bakker from TUT team.

7. REFERENCES

[1] Alekhya, M., Vamsidhar, Y., and Vinodbabu, P. 2011. Large scale Peer-To-Peer Live Video Streaming. *International Journal of Computer Science and Technology*. Vol. 2, SP1 (December. 2011), 190-192.

[2] Bakker, A. et al. 2009. *Tribler Protocol Specification.* Technical Report. http://svn.tribler.org/bt2-design/proto-spec-unified/trunk/proto-spec-current.pdf.

[3] Hao, Y., Xuening, L., Tongyu, Z., Vyas, S., Feng, Q., Chuang, L., Hui, Z., and Bo, L. 2009. Design and Deployment of a Hybrid CDN-P2P System for Live Video Streaming: Experiences with LiveSky. *Proceedings of the 17th ACM international conference on Multimedia* (Beijing, China, October 19–24, 2009). MM'09. ACM New York, NY, 25-34. DOI= http://dx.doi.org/10.1145/1631272.1631279.

[4] Niels, Z., Mihai, C., Arno, B., and Johan, P. 2011. Tribler: P2P media search and sharing. *Proceedings of the 19th ACM International Conference on Multimedia* (Scottsdale, AZ, USA, November 28 - December 01, 2011). MM'11. ACM, New York, NY, 739-742.

[5] P2P-next project. http://www.p2p-next.org/

[6] Thomas, L., Remo, M., Stefan, S., and Roger, W. 2007. *Push-to-Pull Peer-to-Peer Live Streaming*. Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg, Vol. 4731 (2007), 388-402.

[7] Tribler. Source Code. http://svn.tribler.org/

[8] Yifeng, H. and Ling, G. 2011. Peer-to-peer streaming systems. *Intelligent Multimedia Communication: Techniques and Applications*. Springer-Verlag, Berlin Heidelberg, SCI 280 (May 2011), 195–215.

TV for a Billion Interactive Screens: Cloud-based Head-ends for Media Processing and Application Delivery to Mobile Devices

Sachin Agarwal Andreas Beyer Krisantus Sembiring
NEC Laboratories Europe
Kurfursten Anlage 36
Heidelberg 69115
{sachin.agarwal, andreas.beyer, krisantus.sembiring}@neclab.eu

ABSTRACT

The mass adoption of smart phones and tablets has created a vast potential TV audience with access to interactivity-capable mobile devices. Yet, the heterogeneity, limitations, frequent updates and sheer numbers that characterize mobile devices pose serious challenges to porting TV content to these devices. This paper proposes a scalable cloud-based architecture that extracts these complexities from the device and offloads them into the cloud, thereby enabling content providers to reach mobile devices in a scalable and future-proof manner. First results from a preliminary implementation of a media application executed over the cloud, controlled via a mobile-device, and rendered on a TV screen are also presented. Finally, the challenges and solutions in developing scalable cloud-offloading and content adaptation for mobile devices in the context of the HBB-Next European Research project are discussed in this paper.

Categories and Subject Descriptors

C.3 [**Computer Systems Organization**]: Special-Purpose and Application-Based Systems

General Terms

Design, Performance

Keywords

Hybrid broadcast broadband, cloud, media, processing, scalability

1. INTRODUCTION

The number of mobile devices is set to overtake the human population this year [10]. The key challenge for digital TV is to reach these screens in content formats that are suitable for delivery and viewing on these devices. Content providers need tools to tailor TV content to take into account interactive touchscreens and bi-directional Internet connections afforded by mobile devices such as smart-phones and tablets. Users now expect a unified experience across the TV and a second mobile device screen, bringing further complexity of distributing and synchronizing content and applications in real time. Moreover, many heterogeneous mobile devices are frequently updated (via software and via replacement), compounding the complexity of offering such unified experiences across every device to every possible user on a sustained basis.

Even for linear TV viewing (watching common and free channels), TV content needs to be transcoded before it can be transmitted to bandwidth and battery-constrained mobile devices. At the very least, video resolution needs to be adjusted and media streams need to be transcoded into formats suitable for mobile devices. For example, it would be efficient to reduce the resolution of a full HD (1920x1200) stream if the target mobile device client has a QVGA (160x120) screen. Moreover, transmission of multiplexed transport streams containing multiple video channels, such as those transmitted over DVB-T [5], is clearly not advisable for a mobile device connected via a bandwidth-limited cellular 3G connection.

Given the limited battery and processing constraints of mobile devices, offloading such processor-intensive transcoding tasks to the cloud is a natural alternative. The vision is to allow content providers to deploy cloud-offloading tasks on the cloud in a manner that the cloud platform scales with user demand (e.g. number of end users). Media processing can be accomplished on cloud servers (bare-metal or virtualized) and the results transmitted to clients in the device-optimum format. Developers need not to craft multiple client-side applications for different end-devices/operating systems. Instead, by delivering processed media to different devices using standardized protocols and media formats, developers can reach heterogeneous device users and make their applications relatively future-proof. Application logic will also tend to become simpler as most of the complicated processor-intensive tasks are offloaded to the cloud from the device.

1.1 Contributions

This paper presents an architecture for building a scalable cloud-based media processing offloading service over commodity IaaS cloud systems (such as Amazon AWS or Openstack). In addition to transcoding TV content into formats suited to mobile devices, the designed system provides the flexibility to integrate interactive features into TV and web video content delivered to mobile devices. It solves the heterogeneity issue of tablets and smart-phones by abstracting out device-specific media attributes (codecs, containers, network connections, etc.). First results from a partial implementation of the proposed architecture are also presented in the paper. Finally, the paper dwells on the engineering and research challenges of designing a scalable and dynamic

version of the proposed system.

2. SYSTEM ARCHITECTURE

In this section we discuss the underlying considerations while designing a cloud-based offloading architecture for the HBB-Next European Union project. Then, we propose a system architecture to implement a scalable cloud-offloading system that can support multiple client devices and diverse media processing applications. For implementation, we consider Gstreamer [8] for media processing and Openstack [9] IaaS (Infrastructure as a Service) as our infrastructure cloud platform. We note that Openstack can be swapped for other IaaS cloud services such as Amazon's AWS; we ensure this interoperability by choosing the well known EC2 cloud API while orchestrating the IaaS in our implementation as described in Section 2.3. The suitability of virtual machines for highly processor intensive tasks may be questionable, but in fact the underlying cloud can simply be swapped for MaaS (Bare metal servers as a service) or Amazon AWS high performance cluster computing services controlled using the standard EC2 IaaS cloud control API.

2.1 Underlying Considerations

Compared to a typical web application, which might only entail occasional delivery of HTML documents and database queries per user, media processing applications are much more processor intensive on a sustained basis. Delays introduced while waiting for cloud worker machines to boot up would affect many more users on average (since each worker machine can only serve a few users and many more have to be spawned on short notice as compared to a typical web application). Moreover, a marginally overloaded web server may not immediately be noticed by users in contrast to an overloaded media server that is unable to serve video at the required frame-rate. Clearly, architecting scalable media applications in the cloud has different performance requirements as compared to a web service.

The above discussion also brings forth an important economic factor in system design, namely, that a commodity cloud server cannot process real-time media streams for too many users. Clearly, assigning a dedicated server resource (e.g. a Gstreamer media processing pipeline) to one user is not cost-effective, and the web application assumption of "one user, one isolated session" cannot be directly applied in cloud media processing applications.

Fortunately, most users tend to gravitate toward consuming similar content at any given point of time, as observed in web video popularity studies [2–4] which report strong adherence to the Pareto principle (e.g. the 10% videos attract 90% user requests [2]). The key scalability driver in the proposed architecture is the detection of duplicate content processing requests across users and devices which occur due to most users requesting the same content. The system then groups such client requests to be served from a single media processing pipeline in order to eliminate duplicate media processing in the cloud.

2.2 Proposed Architecture

Fig. 1 shows the high level functional architecture of the HBB-Next cloud offloading subsystem. The system is built over an IaaS cloud, through which worker machines can be started or stopped on demand by the HBB-Next cloud-offloading middle-ware. The worker machines are customized with media-processing-specific software such as Gstreamer or Microsoft direct-show (DS), making them capable of executing any application media processing pipeline specified by a developer. Further customizations allow fine grained control and monitoring of these worker machines.

The HBB-Next cloud offloading middle-ware comprises of a application and content head-end, which is the portal for developers and content providers to create, store, and schedule content-rich applications on the system. For example a developer may want two video streams to be mixed for a picture-on-picture functionality. The application to describe this processing may be described as a Gstreamer media processing pipeline and stored in the application database (app DB) while the content DB can be used to store the associated elementary video streams whose picture-in-picture mixing is desired. The content and application manager communicates to the web front end to present the user with application and content information.

The key controlling component of the system is the task scheduler and request router. Several optimizations, such as fine-grained process-level computational requirements, are taken into account while making scheduling and resource allocation decisions in this component. This component is responsible for routing application processing requests to appropriate worker machines, detecting duplicate processing requests by clients and and patching these clients to already-running processing tasks, automatically scaling up (or down) the underling IaaS cloud and monitoring the cloud health to maintain quality of service.

The client capability and offloading decision component is responsible for identifying connecting client capabilities and accordingly offloading tasks to the cloud from the client device. For example, this component uses the user-agent string from the client's web browser to ascertain the screen size (hence, video resolution) to target while transcoding video for the client. Different device profiles and characteristics are continuously stored and updated in the client caps database to assist with this decision making.

Finally, the web front end acts as the entry portal to the HBB-Next cloud offloading subsystem and controls aspects such as user authentication and integration with other WWW services such as social networks. Moreover, this HTTP server might offer HTTP-proxy functionality for content stream delivery to certain classes of devices. For example, Apple's HTTP live streaming protocol is implemented over the HTTP protocol, and the corresponding server may be part of the web front end. Another function of the web front end is to enable application control via second screen devices (e.g., a tablet computer to control an content application being played back on a TV) using HTTP protocols or their real-time derivatives such as web-sockets.

2.3 Example Application

A first prototype of the architecture is now being implemented. In this implementation an Openstack installation running Linux virtual machines installed with Gstreamer

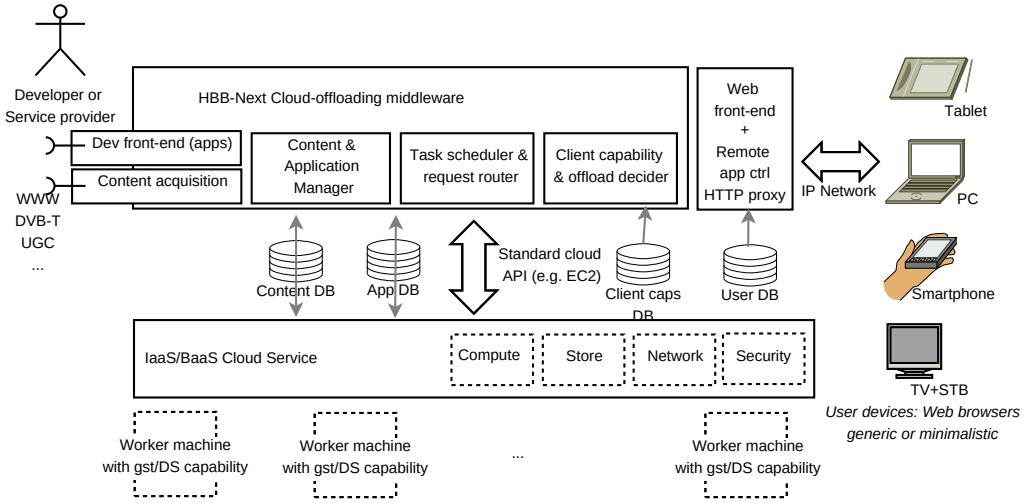


Figure 1: High level functional architecture of the HBB-Next Cloud Offloading Subsystem.

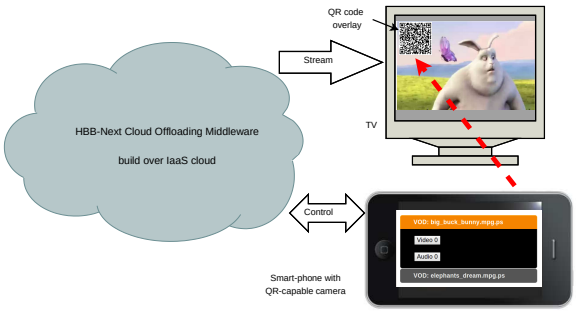


Figure 2: Use-case and components of a demonstration application

media processing software is employed to transcode high resolution video files and serve them to a smart TV at another resolution. The media processing and cloud control software is written in the Python programming language, Nodejs (web-server) [7], and HTML and Javascript [6, 12] (front-end).

Fig. 2 shows an example to highlight a use case implemented over the HBB-Next cloud offloading subsystem. Starting with a high-resolution video, the system scales it down to SDTV resolution, overlays the video stream with a QR code identifier, and transmits this video to a TV in the required codec and container. The user scans the QR code via any generic QR-scan app on her mobile device and gets redirected to a web-page in the mobile device browser that exposes a "remote control" web-page to the mobile device; this mobile device is subsequently used to select the video shown on the TV.

Work continues on developing the more sophisticated features such as cloud scalability, content and application databases, and a wider selection of applications to take advantage of all the features provided by the HBB-Next cloud offloading subsystem.

3. CHALLENGES

The HBB-Next cloud offloading subsystem as described in this paper presents several research and engineering challenges. These challenges are discussed below.

Media Adaptation to deal with heterogeneity across clients and devices, content, networks and protocols A key success indicator of the HBB-Next cloud offloading subsystem will be its ability to support end-device heterogeneity. For example, transcoding media in the cloud can convert media from one format to another, making it possible for devices with limited media decoders to receive videos originally offered in an incompatible video format and transcoded in the cloud. For example, some mobile devices are only able to decode video streams presented in specific codecs: Apple devices such as iPads and iPhones do not decode WebM encoded video. A cloud-based media transcoder that converts WebM video into H.264 video will solve this problem. Moreover, while some users' networks may be capable of delivering high resolution media other users' may not have access to such high quality networks at all times. This problem can be addressed by using the cloud to transcode media [11] into a series of resolutions and then offer these various resolutions to clients. The system's client capability manager will include intelligence to make the decisions about how such heterogeneous devices can be served appropriately adapted media from the cloud.

Cloud scalability as a service Cloud offloading must scale with the number of users in the system without inordinate increases in the cost to the service provider. Video and audio streams have to be decoded, processed, and then re-encoded in real time, making the process quite expensive computationally. One possibility to keep processing costs in check is to detect duplicate processing tasks requested on multiple clients and then replicate the output of these processing tasks to the multiple requesting clients. This poses challenges of duplicate detection and subsequent delivery of outputs across multiple clients without losing the one-to-one interactive capabilities of HBB-Next applications. Additionally, by accurately ascertaining the end-device capabilities

and sharing the processing burden between the cloud and the device accordingly, the processing costs per user can be reduced significantly. Algorithms implementing these features are planned as part of the capability manager, task scheduler and request router in the architecture proposed in this paper.

Real-time processing Implementing media processing use-cases is challenging from a real-time media processing and delivery perspective because of the computational intensity and network latency issues that increase media play-out delays on users' devices. Processing time varies depending on the processing capability of worker machines and the particular task at hand and also depends on the other tasks being executed in the (shared) cloud at a given time. Multiple approaches that may be adopted to address these challenges. Intelligent scheduling of tasks in the cloud, taking into account the processing capabilities of the underlying worker machines and cloud hardware is key to ascertain the approximate processing time for media-processing tasks in the system. In addition, network conditions measurements will be used to ascertain media stream properties (e.g. bit-rates and frame-rates, group-of-pictures intervals etc.) in order for the cloud to adapt the media stream based on network conditions.

Network latency Interactive applications face the hurdle of long round-trip-times between the controlling device (e.g. smart-phone in Fig. 2) and the cloud. Apart from the fundamental limits of propagation delay (proportional to the distance between the end user and the cloud data center), there are additional overheads pertaining to protocol overhead, jitter and queuing/buffering that increase the round-trip-time. Since HBB-Next seeks to deliver inclusive technology that can be used across different platforms (e.g. mobile devices running different operating systems) and heterogeneous best-effort networks (e.g. Internet), standard web protocols such as HTTP are desirable. HTTP-based delivery has the additional advantage of easy passage through public Internet firewalls and proxies. However several performance limitations of HTTP-based streams have to be overcome first [1].

4. CONCLUSION

The key to the success of any cloud off-loading platform is the quality of service (minimum network latency and real-time feel to applications running in the cloud) and low costs. The HBB-Next cloud offloading subsystem needs to overcome the challenges of heterogeneity, cloud scalability, real-time processing and network latency, while delivering a good quality of experience to the user at reasonable cost. The vision to support heterogeneous end devices is achievable by having the cloud support different devices to different extents based on their (lack of) capabilities and by using standard protocols such as HTTP for application and content delivery. Cloud offloading also creates a central control point for the service provider to control the quality of experience of end users and provide homogeneous service across the heterogeneous clients and devices, content, networks and protocols. It can also provide a control plane over which applications are distributed to the viewing screen, mobile device and the cloud in a seamless manner.

This paper has proposed an architecture to implement this architecture over an IaaS cloud and using open-source tools such as Gstreamer for media processing. An early prototype is presented to highlight the ideas and components needed for implementing a complete, flexible and generic cloud-offloading mechanism on an open-source cloud platform. Work continues in addressing the challenges enlisted in this work toward using the cloud as an enabler for building a next generation Hybrid Broadcast Broadband (HBB) standard to bring diverse TV and web media to every possible screen of the future.

Acknowledgment

The research leading to these results has received funding from the EC Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 287848.

5. REFERENCES

[1] Agarwal, S. Real-time web application readblock: Performance penalty of html sockets. In *Conference Proceedings of the IEEE ICC, 2012 (accepted)* (June 2012).

[2] Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., and Moon, S. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2007), IMC '07, ACM, pp. 1–14.

[3] Cheng, X., Dale, C., and Liu, J. Statistics and social network of youtube videos. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on* (june 2008), pp. 229 –238.

[4] Huang, C., Li, J., and Ross, K. W. Can internet video-on-demand be profitable? In *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications* (New York, NY, USA, 2007), SIGCOMM '07, ACM, pp. 133–144.

[5] Industry consortium. Digitial video broadcasting project. http://www.dvb.org.

[6] International, E. Ecmascript language specification. http://www.ecma-international.org.

[7] Joyent. Node.js: Evented i/o for v8 javascript. http://nodejs.org.

[8] Open source project. Gstreamer open source multimedia framework. http://gstreamer.freedesktop.org/.

[9] Openstack foundation. Openstack: Open source software for building private and public clouds. http://openstack.org/.

[10] Short, M. Mobile devices will overtake worldwide population by the end of 2012. http://www.fourthsource.com/news/mobile-devices-will-overtake-worldwide-population-by-the-end-of-2012-4556, 2 2012.

[11] Vetro, A., Christopoulos, C., and Sun, H. Video transcoding architectures and techniques: an overview. *Signal Processing Magazine, IEEE 20*, 2 (mar 2003), 18 – 29.

[12] W3C. HTML 5 specification. http://www.w3.org/TR/html5/.

Social TV System That establishes new relationship among sports game audience on TV Receiver

Yuko KONYA
NTT Communications
Minato-ku, Tokyo, JAPAN
y.konya@ntt.com

Mutsuhiro NAKASHIGE
NTT Cyber Solutions Laboratory
Yokosuka-shi, Kanagawa, JAPAN
nakashige.mutsuhiro@lab.ntt.co.jp

ABSTRACT
There are a lot of studies involving social interactive TV. Most of these services target families and friends. Your friends do not always watch the same program. Therefore, constructing a new relationship among audience is needed. We propose a system that considers a model of human relationship development similar to that of a sports game viewed in a sports bar. In gathering people with the same purpose are warmed up with a mood of users during watching a game simultaneously and their relationship steps up. We implemented a system that supports the format of a digital TV, evaluated it over IPTV platform, and confirmed the advantage of our approach. Our data shows that users were able to show that they enjoyed TV viewing using our approach.

Categories and Subject Descriptors
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Experimentation, Design.

Keywords
Social TV; Twitter; Digital TV; IPTV.

1. INTRODUCTION
Since the introduction of digital TV, audiences have been accustomed to interactivity in broadcasting. With the advent of broadband and network-connected TV sets, more and more interactivity is possible.

There is a lot of ongoing research and non-commercial services in development involving social interactive TV. A social interactive TV is a so-called Social TV. There are two types of social TV. One is a social communication on TV using exclusive application, and the other is watching TV with social media networks (SNS) like Facebook [4] and twitter [11]. In this paper, we use a social TV as watching TV program with SNS.

Just like a sports game, it is more enjoyable to watch TV together than watching it alone. We asked 10 persons about watching TV style. All of them answered watching TV together is more enjoyable than watching TV alone because they want to share emphatic points, unaware points and feedback points of the program they are watching. Akioka et al. [10] shows that TV and twitter have a special connection in Japan. A lot of audiences communicate via twitter while watching TV. Most people use PC or mobile phones to watch messages and comments from SNS with TV. We call this TV viewing style "double screen", because they watch two screens, i.e. TV and PC monitors, at the same time.

TV widgets of SNS like twitter and Facebook for TV sets have already been launched: for example, "Google TV" [6], "FiOS TV" [5], and "torne" [12]. It just displays a widget on the TV screen while watching TV. A typical user interface of widgets is set to a single display area. It is displayed on one side of the TV monitor.

2. PROBLEMS
TV widgets are not an optimal way for enjoying TV. These are not intended for warming up users' mood, because these widgets display users' time line, tweets searched by one hashtag, or tweets searched by fixed hashtag (e.g. TV channel's name). These comments displayed on TV widget are not fit for the program that is on air. There are a lot of unrelated messages on the screen. Previous watching style with SNS has the same problem as TV widget if users do not customize their time line. These viewing styles leave a way for users to enjoy and do not generate a synergistic effect of TV and SNS. To solve the problem, related messages and comments about the TV program are gathered and displayed on the interface.

On the other hand, there are a lot of research and prototype services for communication on TV with exclusive application. These will be useful for enjoying TV programs together with a small number of people like close friends or family members. "Amigo TV" [1] can be used to share consumers' opinions using a web camera. "Collabora TV"[2] provides an environment to annotate videos using temporal comments and allow users to share them. Geerts et al. [3] shows audiences' sharing opinions using a voice chat. These are more enjoyable than widgets on TV. However, these have a drawback in that they are designed for only close relationships. For example, a service using a web camera is not easy for people outside a close circle to enjoy if they are unfamiliar with the people in that circle.

There are various TV services currently available such as Free-to-air, Cable TV, and Internet protocol TV (IPTV). Because of this variety in options, your friends in the real world do not always use the same service as yours. If you want to enjoy watching TV together at your home, you need to find people who use the same TV service. A broadcast schedule depends on the TV service provider. So, the same program is not always broadcast with the same schedule when aired by different TV service providers. Furthermore, social TV with exclusive applications is an unrealistic way in the environment where various TV services are possible. There is a gap because although users want to enjoy social TV, they cannot find friends to watch TV with easily. Therefore, constructing a new relationship among audience is needed.

3. OUR APPROUCH
We suggest a system to solve the problem as mentioned above; "unable to find friends to watch TV with". Our system supports making friends among the audience even though friends or family members in the real world may not always watch the same programs. It is because they have other plans or there is a time difference. However, there are some people watching the same program, particularly popular events like a sports game. So,

previously mentioned dedicated services and research are not for ordinary use. Thus, it is necessary to design a user interface for social TV for users who may not know each other.

People feel comfort and enjoyment by becoming friendly with other people who have already joined the same television service or watched the same program. For example, when you watch an event on TV at the same location like a sports bar, you can create a fellowship feeling among the persons watching together. They may not have been friends before they met at the bar. They passed time watching an event on TV together with a common interest, and after that, they become friendly. On social TV, the situation is the same except that they can communicate with only verbal messages. Walther et al. [13] states that there are not so much difference between face-to-face communication and online communication.

In the case of using social TV, people who are watching the same program are not always friends. They watch TV with a common interest, but they cannot communicate easily because of the distance. So, we use the reference model of human relationship when people make friendsand support their communication. In this study, we refer to Levinger and Snoek's model [9] because they define human relationship steps between zero contact persons. Levinger and Snoek define human relationships at three levels: Level 1 (unilateral awareness), Level 2 (surface contact), and Level 3 (mutuality). Table 1 shows a correspondence table between the model and community formation around a TV program.

Table 1 Applying LEVINGER & SNOEK's model to community formation around a TV program

Levinger and Snoek's model		Extended model for watching TV situation
Level 0	Zero contact	To watch TV alone
Level 1	unilateral awareness	Feel awareness of other audience
Level 2	surface contact to find common features	Try to find common features with audience
Level 3	mutuality	Communicate constantly

4. Interface design

Our system has two interfaces, one has double columns for user comment (Figure 1) and the other has single column comment area (Figure 2). We call the double columns interface "Match mode", and the other interface is called "Normal mode". When a new tweet arrives, it scrolls automatically to indicate a tweet display area for the tweet's upper part. Our interfaces are optimized for watching sports games simultaneously. In addition, we stick with an easy and simple-to-use interface. Once users turn on the TV and launch the browser of our interface, they can enjoy social TV without interacting with it.



Figure 1 our interface for raising a fighting spirit among the audience. (Match mode)



Figure 2 our interface for easy to find people with a kindred spirit in the audience. (Normal mode)

4.1 Common design of the interface

These modes display user name and thumbnail with their message to help the audience feel aware of others. Each tweet message is displayed in a white rectangle and an updating tweet brings a star-shaped icon near the thumbnail (see Figure 3).

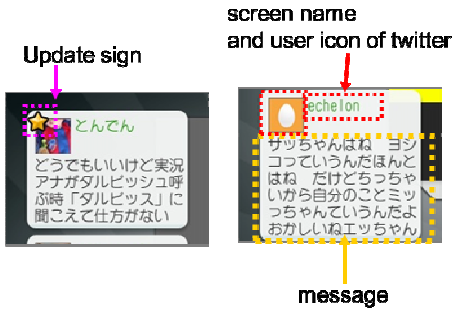


Figure 3 tweet displayed on tweet display area

4.2 Match mode interface

Match mode interface divides the screen into three parts. The tweet display area is divided in the quarter from left side and right

side of the screen. Center area displays broadcasting video. The bar chart representation area is established in the upper part of a broadcasting video area. The bar chart shows the number of tweets displayed on the left and right sides of area. A message categorized by the supporting team is displayed in each column to help audience find their common features, and to step up their relationship. In this interface, users can quickly find messages for their favorite team and can see messages in comparison with opponent team's messages. Baseball fans are summoned up with their combative spirit. It is designed to raise the relationship Level 1 to Level 2.

4.3 Normal mode interface

The normal mode interface divides the screen into two parts. The tweet display area is divided in the quarter from left side of the screen. The remaining area, three-fourths of the screen, displays broadcasting video. It displayes mixed tweets searched by more than one hashtag on the tweet display area. The normal mode user interface is similar to digital data broadcasting commonly seen in Japan,

5. IMPLEMENTATION

We implemented a social TV system with twitter messages on our two interfaces, on the basis of IPTV specification in Japan [7]. This content is created by Lightweight Interactive Multimedia Environment (LIME) [8]. LIME is a specification of ITU-T H.762. It is based on a subset of HTML, CSS, and JavaScript for use in IPTV terminals. LIME is rather strictly profiled so that it may be used on resource-limited devices like TV sets. The IPTV Forum, Japan, which formulated its technical specification for IPTV service in Japan, has contributed greatly toward the standardization of LIME. Therefore, users can enjoy this content without special equipment. Moreover, our interface is ready for digital TV sets in Japan because LIME is very similar to specification for digital TV named Broadcast Markup Language (BML).

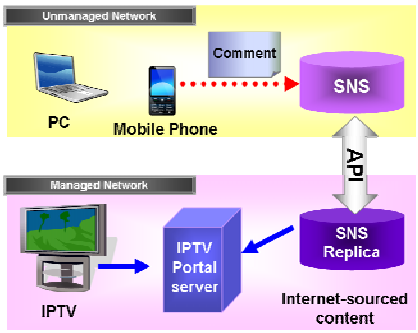


Figure 4 SNS linkage model of our system

Our implemented system cooperates with SNS (show figure 4). In this case, we use twitter. Our implemented system provides information from twitter users who are enjoying the game simultaneously on TV, at the stadium or at a sports bar. We design the interface for read-only viewers; if users want to post a message, they use a PC or mobile phone.

We provide SNS replica using twitter's streaming API. Tweets searched by selected hashtags are gathered in the replica. Our system counts tweets that are available to display on the tweet display area and the number makes a bar graph in the match mode.

6. Evaluation and Result

We conducted two evaluations of the implemented system.

6.1 A lab-based experiment

A lab-based experiment was conducted with 12 subjects (10 men, 2 women) in their twenties to fifties. They have ever watched a baseball game on TV. In this evaluation, we compared interfaces of our approach and previous styles. Each participant was shown a professional baseball game on TV with twitter, using four types of user interface: our interface Match mode, our interface Normal mode, TV widget that is installed in a digital TV shipped in Japan, and previous style of double screen: web browser on PC and TV. Our interfaces can display twitter messages searched by two hashtags of team A and team B. We displayed messages with team A's hashtag on the others interface, because the others can choose only one hashtag on their time lines. We asked the questions listed below.

< Questionnaire entries >

Q1: Do you feel the presence of the other people watching TV with you? (Level 1: awareness)

Q2: Do you find a common interest with them? (Level 2: to find common features)

Q3: Please rank these four interfaces in order to what you enjoy.

(1st … 4th, 1st is the best enjoyable)

Subjects selected their answer from 'yes' or 'no' about each interface in Q1 and Q2. In Q3, they set up an order of the interfaces that they enjoyed. Figure 5 shows the percentage of the number of people who answered "Yes" for each question. All subjects felt presence of audience watching simultaneously by Match mode interface and 11 of the subject (92%) felt presence by Normal mode interface. On the other hand, only half of the subjects (58%) felt presence by TV widget and double screen of the interface.

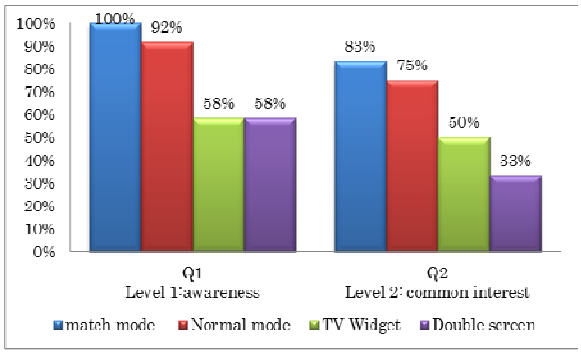


Figure 5 Result of questionnaires Q1 and Q2 by each interfaces (n=12)

Table 2 subjects ranked surveyed interfaces whether they enjoy watching TV with twitter comments. (n=12)

	match mode	Normal mode	TV widget	Double screen
1st	9	2	0	1
2nd	2	4	2	3
3rd	0	3	4	4
4th	0	2	5	4

6.2 In-depth interview

A person-to-person in-depth interview was carried out on 10 subjects in their twenties to fifties. The subjects are five men and five women. All of them like watching a baseball game and had an experience to use twitter while watching TV.

<Detail of the interview>

1. Watch a baseball game with our interface, Normal mode and Match mode. Make use free with the system by TV remote controller. (20 minutes)
2. Interview about the program

They watched TV sitting in a sofa set at 3 meter distance, and eating snacks and drinking Japanese tea if they wanted. The atmosphere of the interview is made to feel like a living room and they can feel relaxed.

We asked a question paying attention to the points listed below.

- Enjoyed watching TV program with our interface
- Feel awareness of other audience watching together
- Try to find something in common with audience

All subjects enjoyed the interface. It was answered that nine of them wanted to use actively and one of them wanted to use it additively. All of them felt awareness of audience from text comments on TV. Some of the subjects felt that they were enjoying a chat while watching TV.

7. Discussion

As shown in Figure 5, the interface of our approach has the advantage of all of the questions. We tested the result using Q-test; there are significant difference in Q1, and Q2. (Q1: chi-square value = 9.222 ($p<.05$), Q2: chi-square value =13.000 ($p<.01$))

The result of the in-depth interview shows that subjects enjoyed TV viewing using our system. It helps to feel the presence of other simultaneous viewers.

These results show the level 1 of the relationship model. This helps the users feel the presence of other simultaneous viewers (Level 1: Unilateral awareness) and a sense of oneness among viewers (Level 2: surface contact helps to find common features) while watching TV. We were worried whether users could enjoy watching TV, because there was no interaction function in our interface. However, we understood that users probably had a lot of fun with the atmosphere that they participated in. We think that users can enjoy watching TV more if there was interaction. Table 2 shows the ranking of the user interface by enjoyment watching TV with SNS. According to one-way ANOVA multiple comparisons by Holm, there are significant difference between Match mode and TV widget, and double screens.

We will consider whether a communication level becomes Level 3 if there is interaction. However, many TV sets and set-top boxes do not have keyboards. We are required to consider that input device and method.

In addition, users can not miss a highlight scene of a game because we implemented these interface on one screen at a time.

8. Conclusion

In this article, we described the system of our approach that considered the model of relationship development and a strengthened feeling of unity because constructing a new relationship among zero contact audience is needed. Our system's Match mode user interface is optimized for watching sports games simultaneously. It raises audience's fighting spirits and develops a feeling of fellowship.

We implemented the user interface using IPTV specifications, especially the user interface created by LIME (ITU-T Rec. H.762). Therefore, users could enjoy this interface without special devices. If the service providers are ready, our interface is ready.

We evaluated the system on the IPTV platform. In the evaluation, our interface had an advantage over the other previous viewing style with SNS; double screens, and TV widget, when viewers were watching a baseball game on TV. They enjoyed social watching with presence and a sense of camaraderie we found that our approach's interface Match mode could feel unilateral awareness (Level 1) and help to find common features (Level 2).

This time, we only evaluated a baseball game. We believe the system fits other sports games. Furthermore, we will consider whether this interface fits other genres of programs. Moreover, we would like to consider how to make interactions on TV with poor input device in order to realize deeper communication, from Level2 to Level 3.

9. REFERENCES

[1] Coppens, T., Trappeniers, L. and Godon, M., 2004. AmigoTV: Towardsa social TV experience. Proceedings of EuroITV 2004.

[2] Chris Harrison, C. and Amento, B., 2007. CollaboraTV – Making TV Social Again.

[3] D.Geerts, I.Vaishnavi, R.Mekuria, O.Deventer and P.Cesar, 2011, "Are we in sync?: synchronization requirements for watching online video together" . CHI '11 Proceedings. pp.311-314.

[4] Facebook, http://www.facebook.com/

[5] FiOS TV, http://www22.verizon.com/residential/fiostv?CMP=DMC-CV090057

[6] google TV, http://www.google.com/tv/

[7] IPTV-Forum Japan, http://www.iptvforum.jp/

[8] ITU-T Rec. H.762, 2009, "Lightweight interactive multimedia environment for IPTV services (LIME)

[9] Levinger, G., & Snoek, D. J. 1972 Attraction in relationship: A new look at interpersonal attraction. General Learning Press

[10] S.Akioka, N.Kato, Y.Muraoka and H.Yamana, "Cross-media Impact on Twitter in Japan", SMUC '10 Proceedings, pp.111-118.

[11] twitter, http://twitter.com

[12] torne, http://www.jp.playstation.com/ps3/torne/

[13] Walther, J. B., & Tidwell, L. C. (1995). Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. Journal of Organizational Computing, 5, 355-378.

Leveraging the Spatial Information of Viewers for Social Interactive Television Systems

Niloofar Dezfuli Manolis Pavlakis Florian Müller
Mohammadreza Khalilbeigi Max Mühlhäuser

Telecooperation Lab
Technische Universität Darmstadt – Germany
{niloo,pavlakis,f.müller,khalilbeigi,max}@tk.informatik.tu-darmstadt.de

ABSTRACT
TV watching has various stages beginning with the presence of viewers in a living room until they get fully engaged with TV content. In this paper, we explore how little spatial information about presence, head orientation and different body postures of viewers can be exploited to design implicit TV interactions. We present two user studies. In the first study, we empirically establish various levels of watching engagement. The results of this study ground the design of our prototype we call CouchTV: a social interactive TV that tracks spatial position, orientation and posture of viewers in front of TV and based on that, supports (re)engaging in TV content. CouchTV implicitly provides typical TV interactions such as volume adjustment and also awareness about TV content. We used our prototype in the second study to explore how implicit interactions may affect watching experience. Our findings showed that CouchTV eased joining and reengaging in TV content and enriched watching experience.

Categories and Subject Descriptors
H5.2. [Information interfaces and presentation]: User Interfaces-Input devices and strategies.

General Terms
Interactive Television, Design, Human Factors, User Experience

Keywords
Spatial Information, Implicit Interaction, Proxemics, Awareness

1. INTRODUCTION
Research on television has been more focused on viewers who are attentively watching TV in a sit-down manner. However, watching activity has various stages based on people's spatial situation such as presence, orientation and body postures. This begins with the presence of a viewer in a living room in front of TV until she gets fully engaged with the TV content. Although fully engaged with the program, viewers may change their posture or position for different purposes [2,6]. For instance, they may lie down to relax, lean forward to concentrate more or even temporarily leave the room to do something quickly. Moreover, there are also environment-related interruptions such as a phone call or starting a conversation with others, which may distract the viewer from watching and eventually, diminish the user experience. In this work we investigate how spatial information of viewers can be exploited to support implicit interactions with television systems and later, we want to study if they may affect user experience while watching TV. Our work is motivated by the following research questions, which to the best of our knowledge have not yet been explored: (1) Can additional awareness on the current TV content support (re)engagement into video watching activity? (2) Can implicit TV interactions promote viewer's comfort as a complementary input channel?

There are a number of prior studies investigating how spatial relationship can enrich the interaction with large interactive surfaces [1,7] or small displays [4]. Harrison and Dey presented Lean and Zoom [4] that detects user's proximity to a notebook and proportionally magnifies the on-screen content. Ballendat et al. [1] presented a complete ecology in a small space Ubicomp environment in relationship between people, devices and non-digital objects. They illustrated how proxemics can regulate implicit and explicit interactions running through a media player. However, research on the usage of spatial information for social interactive televisions is still very scarce. Geerts and Groof [3] have pointed the necessity of awareness tools for indicating the presence of other users to support remote interactions. Despite the common fallacy that TV viewers are always concentrated on the TV content, Hawkins et al. [5] explored attention style of viewers based on the length and frequency of looks at TV displays. Chorianopoulos [2] also informed the design of Interactive TVs by considering multiple levels of attention to deepen viewers' relationship with the world of program. There is also a variety of commercially available services which called up different sets of widgets depending on who's in the room. However, they focus more on supporting meta-data related to running video rather than supporting (re-)engagement in TV content [12, 13].

In this paper, we first conduct an explorative study to systematically understand video watching activity in terms of exploring different levels of engagement. The results of this study empirically ground the design of a set of awareness and interaction techniques. We also developed a prototype which coherently integrates all the techniques. We evaluate CouchTV to further investigate how implicit awareness and interactions may enhance TV watching experience and particularly, answer the aforementioned research questions. At the end, we discuss our findings and conclude with future work.

2. PILOT EXPERIMENT
In order to gain a better understanding on how TV viewers get engaged in watching activity and what are the intermediate steps, we carried out an exploratory field study. We observed 15 volunteers (7female, 8male) watching TV individually or collocatedly in their living rooms. We recorded participants watching activity for the whole two weeks. Each household watched TV about 10 hours in average (SD: 2.7). Multi video coders coded the video using an open and selective coding approach [11]. At the end, a semi-structured interview was conducted with each participant.

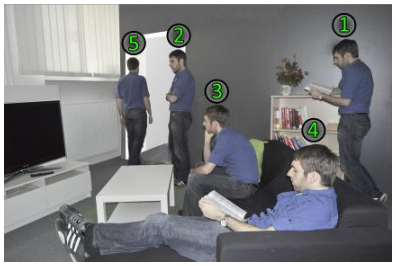


Figure1: Levels of engagement in TV watching.

Based on the analysis, the results reflect (1) viewers' levels of engagement in watching activities and (2) the requirements for the design of CouchTV prototype system.

2.1 Levels of Engagement in Watching Activity

The videos analyses revealed five different levels of engagement for video watching based on presence, location, orientation and posture of people (cf. Fig. 1):

1) Passive is assigned to people who are present in the TV room but for other reasons rather than watching TV e.g. to water plants. In this level, either people are not located yet in front of TV or they do not face towards the TV.

2) Inclined is a level in which people stand in front of the TV and their attention is also towards the TV display. Based on our observations, in this level it is very likely that people start zapping through TV channels in order to find a favorite program.

3) Involved is when people sit in front of TV and engage in the content and his attention is mainly oriented to the TV.

4) Oriented is a level in which people are already involved in watching however, their attention is partially influences by other activities (e.g. reading a newspaper) or by presence of other viewers (e.g. conversation).

5) Away is when people leave the TV room.

2.2 Requirements

Based on the qualitative analysis of videos and also participants' salient quotes during interviews, we derive three main requirements for our system.

Efficient Overview and Awareness: Participants frequently mentioned that they require awareness for something special shown on TV and proposed to do this through e.g. "visual cues on TV screen". P6 said "*I usually miss my favorite TV shows. It would be great if TV screen could display which of my favorite program is currently running when I pass its screen*". Moreover, participants underlined that these cues should also contain information about the "program timeline" and also "friends' activities". This will help them to know whether it is still worthy to join watching the program and if they can watch the program together with their friends. Participants also added that it the TV should support group watching by displaying the common friends or favorite programs of present users in the TV room.



Figure2: Couch TV supports awareness about TV content.

Easy and Immediate Reengagement: Based on our observation, participants were interrupted from watching, either caused them to leave the room e.g. smoking cigarette or shortly distracted them from watching such as receiving a phone call. In order to re-engage to the watch activity, they asked a co-viewer about missing scenes or did not follow the program afterwards and started zapping or surfing the EPG for a new program. The participants stressed the fact that re-engaging to the content is almost difficult and also can be highly time-critical. P11 added "*It can often happen that I do not understand what the actors are talking about once I am not in the room because of any reasons* ", therefore, they required easy and immediate re-engagement.

Implicit Interactions: The participants generally liked the idea of "being able to control TV implicitly". P14 said *"I'd like if TV will be paused whenever I'm watching alone and needs to leave the room."* P2 stated *"It's not always as easy to relax when you know that at any moment you may need to look for the remote control to give a command to the TV*". However, allowing users to interact explicitly was also considered important as p15 said *"I'd like to be able to choose, either to interact implicitly because it's convenient or command to the TV explicitly as it is necessary".*

3. SYSTEM PROTOTYPE

Based on findings of the study, we developed a prototype that supports (re)engaging in TV watching activity by providing appropriate awareness about the TV content. We first showcase the system functionalities via a walkthrough scenario and then present some technical details.

3.1 Walkthrough

We start by illustrating our concept and interaction techniques as shown by the following scenario:

Initially, the TV is inactive and no one is in the TV room. Karl enters the room (*Passive*). The TV activates its screen and displays the Karl's favorite programs that is currently running on the TV and also provides an overview of Karl's online friends who are watching this program (cf. Fig. 2.a) [4,10]. Karl approaches and stands in front of TV while looking at its screen (*Inclined*). TV displays more detail about the program such as a timeline which shows when it is started. Karl can sit on the couch which causes TV to play his favorite program in full screen, directly zaps through other programs by the remote control or pay no more attention to the TV. Finally, he decides to sit and start watching his favorite boxing match with remote friends *(Involved).* Later, he started watching a soccer match but in the meanwhile, his wife asked him to come to the kitchen *(Away).* When Fred reenters the room while his attention is back to the TV, a notification appears on the corner of TV screen about the goal that was scored while he was away (cf. Fig. 2.b) *(Involved).* TV also notifies that it records the missing scene for Karl. Later, while watching movie, Karl's wife enters the room to grab her book from bookshelves. She looks at the TV display to see what Karl is watching

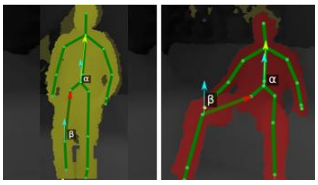


Figure 3: User's upper, lower and normal vectors between skeleton joints and their alignments in the 3D space.

(Inclined). So TV displays information about the programs which he is watching such as its title, short description and actor's name as it is shown in figure 2.c. (Its noticeable that this information also appears if Karl reenters the TV room and the program changes while his absence) As she finds the program not interesting, she leaves the room *(Away).* At this time, Karl receives a phone call *(Oriented).* As he picks up his phone, the TV decreases the volume till the call is finished.

The proposed interaction techniques do not necessarily form an exhaustive set of interactions but they provide solution to typical problems caused in different levels of attention while watching.

3.2 Technical Details

We have developed our proof-of-concept CouchTV prototype using Microsoft Kinect [10] since it does not require any instrumentation on the viewers' body. We mount the depth camera on top of TV screen. Thus, the built-in depth sensors detects user in a distance between ~1.2m and ~3.5m to the TV display.

Postures are recognized in a multi-step process based on vectors between skeleton joints and their alignments in the 3D space. We calculate three vectors, one from user's upper body (hip to neck, one from user's lower body (knee to hip) and the other as a normal vector straight up from the floor to the ceiling depicted as yellow, red and blue vectors in Fig 3 respectively. To determine different postures of the user, we compare the angle (α and β) between these two vectors with the normal vector. In order to calculate user's attention towards the TV, we observe the normal vector of the plane span by the joint positions of the user's spine and the two shoulders (cf. Fig. 3). If this hits the TV position, then attention of the user is oriented towards the TV screen. Using this spatial tracking data, CouchTV reacts to the different situations (explained in the walkthrough) and also displays appropriate awareness level on TV screen (cf. Fig. 3).

4. USER STUDY

We conducted a study to understand if and how CouchTV can enhance TV watching experiences. The study took place in a lab, furnished like a living room setting to simulate real life situations. We recruited 12 groups of 2 friends (10m, 14f, avg. 25 years; SD: 7.2 years). Fifteen participants were students, six of them were faculty members or university staff and the rest did not list their affiliation. All of them but two were regular TV viewers. They were introduced to CouchTV upfront.

The study contained two watching conditions: (1) with CouchTV system reactions and (2) without any system support (normal TV system). The study took about 3 hours for each group. The order of conditions was counter-balanced across groups. We picked 2 shows from 5 different genres (sitcom, cooking show, drama, sport and news) to study the effect of genre on our concept.

Since the result of the pilot study showed that interruptions from watching activity causes people to transit between different levels of engagement in watching activity, one of the authors of this paper participated in the study to simulate the real-life interruptions during the study. He did not attend in watching activity but externally imposed participants some interruptions such as asking for a conversation, making a phone call or offering some drinks and snacks in both conditions of the study. The goal of these interruptions was to cause participants either to leave the room for a short while or to lose the focus from TV content. Moreover, we let participants to answer their phone calls as well as leaving the TV room to go to rest room or to smoke cigarettes during the study. These interruptions and also different spatial situations of participants such as their body postures, head orientation, etc in front of TV enabled us to showcase our system's interaction techniques accordingly.

Data Gathering and Analysis

We used interaction logs, video/audio recordings, AttrakDiff questionnaire [6] and six focus group sessions conducted after each study with two groups of participants. The AttrakDiff test is used to evaluate hedonic (HQ) as well as pragmatic qualities (PQ) of CouchTV and normal TV watching experience. Participants completed two AttrakDiff questionnaires immediately after each session to avoid difficulties in recall. The focus groups and videos were transcribed and analyzed using an open coding approach [11].

5. RESULTS AND DISCUSSION

In this section, we report and discuss our research findings from coding process regarding the general reactions of users to our prototype and the overall TV watching experience while using both normal TV and CouchTV.

5.1 AttrakDiff

For questionnaires, a 2D graphic is generally generated, which places the CouchTV and typical TV systems in terms of PQ (i.e. usability of the system) and HQ (i.e. novelty, interest, etc.). The typical TV system has been qualified as Task-oriented. It means that the typical TV is almost usable but is average in terms of HQ. The confidence rectangle is particularly large in the hedonic dimension, which means that participants' opinions were mixed. In terms of attractiveness, the overall impression is that normal TV is moderately attractive. According to our participants, watching typical TV is "manageable" and "simple", but also very "ordinary"and "isolating".

The watching experience using the CouchTV system has been qualified as desired. It means that the CouchTV optimally assist and motivated the users. The confidence rectangle is small for both HQ and PQ. In terms of attractiveness, the overall impression is that this system is highly attractive. Participants found it very "novel", "pleasant" and "presentable" but the bad point is that it is not challenging for the users.

5.2 Issues Identified

In sync and immediacy is a must: Our primary research question was whether content awareness during program is seen as a useful addition to the TV viewing experience. Participants confirmed that with their comments during the focus group sessions. They clearly indicated that their experience is positive only when the relevant information is in sync with the TV show and appears as soon as user looks back to the TV screen to reengage to the program content. P23 said "*The awareness was definitely helpful as it was immediately there, right when I came back to the room and was curious about what happened during my absence!*" P8 put it more emphatically saying "*The system should make sure*

that the awareness is in sync with the content. I do not want to receive any info about an event which did not happen yet or is happening currently in the program".

Social vs. content connectedness: The participants emphasized that they enjoyed watching experience while CouchTV provided awareness for whom else reentered the room and thus, they were not distracted from watching. P18 commented " *I definitely liked it because my partner read what has happened when he was not at room and did not ask me when I was focusing on a movie dialogue*". However, CouchTV also induced a limpness between co-viewers: the more they received content awareness through CouchTV, the more they were disconnected from co-viewers. Nevertheless, the actual disconnection is not to be regarded as pure negative, since intentional disconnection such as P18 statement can enhance the overall experience. Additionally, several participants mentioned that they frequently came across situations in which they entered the TV room and wanted to look up about and probably join the TV shows which another collocated person was watching. However, they failed to do so because asking them could be too distracting for viewers or the viewers would response only one or two questions with incomplete answers as they were deeply involved in watching activity.

Genre plays an important role on awareness consumption during program: The participants mentioned that content awareness should be adapted to genre with different number of plots, video length and content quality. P2 said "*The importance of awareness can be really dependant on what I'm watching. For example, if I have to leave the TV room for two minutes, I may miss much more information while watching breaking news comparing to watching movies.*" Some participants also noted that they want to receive awareness for rich quality video which asks for viewer's continues attention and they do not care about what they miss about poor programs which are not on interest.

5.3 Particular things that participants liked

TV as an ambient display at home: Participants appreciated activating TV screen and receiving awareness about their favorite programs and online friends. P6 commented: "*That (Initial Awareness)is great because I enter the TV room several times a day for various reasons. So I can be aware if any of my favorite program is showing on TV and don't miss them*" and P19 said " *If I had this feature on my own TV, I explicitly pass the TV room to just know which program is currently going on*" and P1 added "*I hope I could also check my Facebook and news in addition to TV program in the morning just by looking at TV screen*". There already exists technology, which supports this to some extent [8]. However, it does not consider TV interaction based on viewer's spatial information.

Effortless information gathering while watching: The majority of participants indicated that the content awareness which appears on TV screen after interruptions, enabled them to easily and immediately reengage in the TV program without looking up themselves. P2 said *"The awareness about new program that began during my absence is really useful as it helped me to decide to watch it or zap to another program".*

Enrich interactivity with TV systems: According to our participants, CouchTV prototype enriched the TV watching experience in terms of interactivity. Participants commented that they frequently came across situations in which they desired using no 3rd party mediator device to operate TV as it was lost (e.g. P1 stated " *I'd like recording feature while I was watching the football match when I had to open the door for my friend when he was knocking it*") or it was misplaced (e.g. P24 stated " *I was lying on sofa and my phone was ringing. I wanted to ask my friend to reduce the volume via remote control as it was far but TV just did it when I picked my phone*").

6. CONSLUSION

In this paper, we explored various levels of TV watching activity and contributed how continues monitoring of people's spatial behavior enhances interaction with TV systems. The observed phenomena provides evidence that CouchTV offers several insights into the role, that implicit TV interaction and content awareness can play in (re)engaging in the TV content. It also provides valuable lessons for the design of iTV systems where the user's spatial situation can regulate implicit TV interactions.

We found a lot of trends that we would like to explore in the future. The largest unsolved issue is how one can interpret and create meaningful implicit behaviors and repairing mistakes and avoid unwanted situations of such systems. Even with this caveat, we believe that implicit interactions based on the user's spatial information will be a powerful way for TV interaction or even in the everyday environment. Future work also includes investigating other proximity sensing technologies and exploring other spatial information while watching different genres.

7. REFERENCES

[1] Ballendat, T., Marquardt, N. and Greenberg, S. Proxemic interaction: Designing for a proximity and orientation-aware environment. In *Proc. of ITS '10*.

[2] Chorianopoulos, K. User Interface Design Principles for Interactive Television Application. In *Journal of HCI '08*.

[3] Geerts, D., Grooff, D. Supporting the social uses of TV: sociability heuristics for social TV. In *Proc. of CHI '09*.

[4] Harrison, C. and Dey, A.K. Lean and zoom: proximity-aware user interface and content magnification. In *Proc. of CHI '08*.

[5] Howkins, R.P., Pingree, S., Hitchon,J., Radler,B., et al., What produces television attention and attention style: Genre, situation, person, and media perceptions as predictors. *Human Communication Research, 2005*.

[6] Hassenzahl, M., Platz, A., Burmester, M. and Lehner, K. Hedonic and ergonomic quality aspects determine a software's appeal. In *Proc. of CHI'00*.

[7] Vogel, D. and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. of UIST '04.*

[8] Harboe, G., Metcalf, C., Bentley, F., Tullio, J., Massey, N., and Romano, G., Ambient social TV: drawing people into a shared experience. In *Proc. Of CHI '08*.

[9] Oehlberg, L., Ducheneaut, N., Thornton, D. Moore, R., Nickell, E., Social TV: Designing for distributed, sociable television viewing. In *Proc. of EuroITV'06*.

[10] Microsoft Kinect, http://www.xbox.com/kinect.

[11] Strauss, A., Corbin, J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. *Sage Publications, 2008*.

[12] www.connectedtv.eu.

[13] http://www.v-net.tv/nds-surfaces-the-next-revolution-in-tv/.

Creating and sharing personal TV channels in iTV

Bruno Teles
Aveiro University
3810 – 193 Aveiro
Portugal
bmteles@ua.pt

Pedro Almeida
Aveiro University
3810 – 193 Aveiro
Portugal
almeida@ua.pt

Jorge Ferraz de Abreu
Aveiro University
3810 – 193 Aveiro
Portugal
jfa@ua.pt

ABSTRACT
Platforms for supporting UGC got very popular during the last decade. However, due to the large amount of content available solutions to support the categorization and personalization of the content by means of channels have been introduced. These improvements allow not only organizing the UCG but also, in some platforms, to promote live broadcasts of audiovisual content. Interactive Television (iTV) platforms are also providing means to access UGC, but the capabilities for self-broadcasting are still lacking. In this context, this paper presents a proposal for an interactive application to create and share personal channels in an iTV service. The proposed model was prototyped and tested in a typical TV usage environment. *MyChannel* allows its users to create personalized channels right from their TV sets gathering and combining audiovisual content from different sources: web video platforms; regular TV channels and; live content, created by the user, through a webcam. It also includes a set of social features around TV content. This model was evaluated in the lab by two groups of users with different levels of experience with iTV platforms. The results show that users are willing to have the ability to create and share their own channels. Based on these results, one can consider that an iTV application that reflects the proposed model may be relevant when a more personalized TV offer is at stake.

Categories and Subject Descriptors
H5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Prototyping, interaction styles*.

General Terms
Design, Experimentation, Human Factors

Keywords
IPTV; interactive television; user-generated content; personalization; user experience.

1. INTRODUCTION
Today the Internet provides a huge offer of web platforms for sharing and viewing audiovisual content. These platforms range from corporate platforms like Hulu to open platforms of a professional or amateur nature. In addition to providing professional video channels, this increasing number of "TV on Internet" platforms, also allows regular users to create their own channels using their own content (User Generated Content - UGC) which is enriched with sharing and social features. However, from the perspective of a traditional TV viewer (by traditional we mean those who prefer to enjoy TV in a more relaxed way) a more TV-centered approach can result in a better fit than the existing "TV on Internet" offer. In this sense, there is an interesting opportunity window for the design and development of interactive applications that would allow TV possibilities similar to those of UGC platforms. Applying the UGC paradigms to television, specifically as regards the creation and sharing of personal TV channels, demands a high level of personalization and interactivity. IPTV appears to be one of the best technical solutions available to implement this.

2. TOWARDS PERSONALIZATION
Online video content is currently far from being provided only by filmmakers, television producers or distribution companies. An increasing number of consumers/contributors (prosumers, as Alvin Toffler introduced [4]) are also at stake. In this context, the passive attitude of the consumer is transformed into a proactive approach, creating a kind of "*collaborative, collective, personalized and shared*" society [6] of a UGC era. Research data shows that in the U.S., during November 2011, more than 183 million unique viewers watched online videos [3] thus representing a considerably increase, taking into account the same period of the previous year, in which the total unique viewers were less than 173 million [2]. This phenomenon on the web, where YouTube stands out, is also moving to the TV arena, being leveraged by over the top solutions that allow UGC to be watched in regular TV sets.

From regular iTV platforms, to Smart TV systems, videogame consoles like Playstation3 or Nintendo Wii or Media Center Systems, users are able to access content from YouTube and other major video platforms. As an example of these systems, Boxee [1] allows access via browser or dedicated apps to various web platforms and, in addition, the "watch later" feature enables the user to benefit from a lineup of videos found on the web.

The positive incursion of web features on TV is clear; nevertheless this is mostly oriented to a viewing perspective. However, it may be expected that the offer of interactive television applications will carry a similar evolution as we witnessed in the Web, providing their users with the ability to publish, share and organize a wide range of audiovisual content in personalized channels.

Building upon this context, the work of Pronk et al. introduced the concept of *Personal Television Channel* [10] as an example of the creation and personalization of interactive TV channels. The authors' proposal allows the creation of channels for a specific thematic being the content defined through a set of filters defined by the channel owners. Although the users are able to choose from *strict channels* or *fuzzy channels* (differentiating by this the scope of the filters) the content is chosen by a recommender system without the control of users. The *Unified EPG* [9] also proposes the ability of users to create a personal channel with multimedia content from the EPG, DVR or web through a web portal. To

share the created channel with friends users are able to publish a clip of what they are seeing at a moment. Both projects bring an approach to creating personal channels but they have some limitations in what concerns with content personalization and scheduling. The social and community features are also limited.

In this context *MyChannel* proposes additional features like a customizable programme grid, the integration of several sharing, communication and recommending features to promote the social adoption of such a system.

3. THE MYCHANNEL MODEL

Considering the context described in the previous topic, this paper presents a proposal for a platform that allows users to freely create and watch personal broadcasted channels. This proposal is targeted at iTV platforms. It is the researchers' belief that the self and personal broadcasting may assume a significant role in the near future of Social TV applications. The proposal was conceptualized and a medium fidelity prototype was built with the main functionalities in order to allow the development of an evaluation process that could reveal the real expectations and level of acceptance of iTV viewers to the proposal.

3.1 Goals

The process underneath the specification and prototyping of the *MyChannel* application took in consideration the features already proposed in related projects complemented with features already available on on-line platforms aiming to get a fully integration in a TV context. Thus, the *MyChannel* proposal was based on the following goals: i) to grant support for the creation of personal channels organized in a dynamic EPG built upon multiple sources of content; ii) to allow the definition of a fully customized grid of broadcasting and; iii) to support the integration of social features enabling the user to know which are the most viewed channels, to share and recommend channels and to interact with other users through different methods of synchronous and asynchronous communication.

Considering these goals and the researchers' desire to get the users opinions and expectations for such a system, the work was structured in four major steps: i) the functional specification, allowing the identification of the relevant features to include; ii) the development of mockups, helping the research team to better predict the needs and specifications of each area; iii) the implementation of a prototype that could integrate in a semi-functional manner the proposed concepts and; iv) the evaluation of the prototype along with the proposed concepts.

Based on the work in each of these steps a proposal was built and its features are described in the following section.

3.2 Key features

This section describes and illustrates the defined features, categorized by the areas to which they belong.

A. Channel creation: the main area provides support for each user to create his personal channels in 5 simple operations:

- **Registration**: a regular login and a Facebook authentication is proposed;
- **Description**: each channel is categorized and completed with profile information: title, tags and description;
- **Visuals**: visual templates contribute to differentiate channels;
- **Scheduling**: defining the broadcast schedule for the channel (all day, only on weekdays, only in the weekend);
- **Programming**: definition of the programme grid. It allows users to select for each time slot the content to be broadcasted (derived from the DVR, on-line platforms - UGC, Webcam or from the operators' EPG).

B. List of channels: accessed through the homepage (see Figure 1), this area provides a list of all the channels available in *MyChannel*. The users are also able to read a synopsis about each channel; to include it in a favorites list; to know which of his friends are seeing the same channel and; to invite other users and creators to the friends list. To promote the channels, the list is filtered by "Top rated", "Most viewed", "Most Recent", "Category", "My favorites" and "My friends' channels". Some of these features are only available to registered users.



Figure 1. The homepage of *MyChannel*.

C. Friends and recommendations: an important characteristic of *MyChannel* is its social features. Therefore, the homepage includes important features that let users track the activity of friends, and use that information as recommenders for their own activity. Additionally *MyChannel* allows users to manage the friends list, their latest activities and favourite channels, as well as read direct recommendations (programmes/channels) received.

D. *When viewing a channel*: the user can watch the contents of his own channel or the channels created by other users. There are different features available while watching personal channels organized in the following areas:

- **Info area**: where all the information on the selected channel is available. It allows: channel evaluation (Figure 2) and bookmark; channel and programme recommendation; checking the channel grid and; making programme recommendations. The user may also share with others what he is watching;
- **Watch area**: it allows knowing who is watching the same channel and for how long;
- **Chat area**: users can communicate with each other, through text or emoticons. When a user sends a message, a notification pops out;
- **Help area**: with information to assist users with doubts about the available features.



Figure 2. Performing a channel evaluation.

E. Channel configuration: along with the usual channel settings, it is important to highlight an additional feature, "Transmission via Webcam", which enables users to perform a direct broadcast from a webcam connected to his STB.

3.3 Building a prototype

As referred the proposed model was developed into a prototype. This step revealed to be very demanding due to the willingness of the research team to include most of the conceptualized features. The first task consisted of choosing the platform in which the prototype would be implemented. Considering the limitations of time and resources and the technical limitations of the iTV platforms, the prototype was developed in Adobe Flash (with the support of Actionscript). One key aspect taken in consideration was the ability to support interaction by means of a remote control and thus allowing the creation of a more lifelike user experience, simulating the interaction with a STB. For that, a device that intercepted the infrared signals of the remote control was integrated.

4. EVALUATING MYCHANNEL

After the development of the prototype an evaluation of the proposed concepts structured in 3 phases was carried out: 1 - Selection and characterization of the sample of participants; 2 - Completion of the evaluation sessions with the prototype; 3 - Analysis of quantitative and qualitative data gathered in the 1st and 2nd phase. The following sections detail these phases.

4.1 Selecting a sample

An intentional sample [11] was established with 20 participants (all students starting a master course in multimedia). The sample was later divided into two groups of 10, taking in consideration the reported (in a pre-test questionnaire) experience of the participants in the use of iTV platforms. Although the sample is not representative of the population, it includes a type of users that could become early adopters of such a system.

The data collection occurred in three distinct moments along the evaluation: 1) a pre-test questionnaire with the group of participants; 2) tests with the prototype, carried in a lab room and; 3) a final questionnaire targeted at retrieving the participants' opinions about the experience. To support the evaluation sessions the researchers adopted a direct observation method supported by an observation grid. The sessions were also recorded in a non-intrusive way by means of sound and screen capturing (using CamStudio software).

The pre-test questionnaire was aimed at identifying the profile of the participants and their level of technological literacy, particularly in terms of their experience with iTV platforms. The post-test questionnaire was aimed at understanding the participants' level of opinion and satisfaction and the positive and negative aspects of their experience with the prototype.

4.2 The evaluation sessions

The evaluation of the prototype took place in a working space adapted as a living room and each session started with a brief presentation of the prototype and the explanation about how the test was going to be performed. It was also made clear that the participants were allowed to stop the test whenever they wanted and (based in a variation of the *Think Aloud Protocol* [5]) that they should express themselves in any situation. Participants were also informed that the tests were guided with the support of a list of tasks, although they had total freedom to explore the application. The list of tasks included creating a new channel, searching for content to that channel, viewing other channels and performing some social activities related with the content.

5. RESULTS AND DATA ANALYSIS

5.1 Characterization of the sample

The evaluators group was composed by a majority of male participants (55%) and the average evaluator had an age of 25 years. All the participants had a daily routine, which included watching TV at home, with 45% referring a viewing time between 1 and 3 hours. 50% watched TV alone, 45% together with the family and the remaining with other people. Most respondents had already used iTV platforms (75%) but their level of expertise with these platforms was wide-ranging. Despite the 75% that already tested an iTV platform, there was a weak or moderate frequency of use of the features available on such platforms. One interesting issue to point out was the high number of participants that said to simultaneously use distance communication mechanisms while watching TV (e.g. chat, instant messaging) - this *telenaute* practice (as referred by Lafrance [7]) is more and more frequent among people in the age of these participants. The questionnaire also showed that the participants were regular users of online video platforms (60% say they visit it "more than once a day"). However, they revealed low usage of the rating mechanisms available in the video platforms (75% with a frequency of use between 0-25% when watching a video). This information is relevant considering that this type of mechanism is used to evaluate the different personalized channels available in the prototype. Lastly, the majority of the respondents showed a prior interest in an application that allowed "creating a personal television channel, from the TV set (75%). Therefore the participants were motivated and willing to interact with the prototype.

5.2 The results about MyChannel

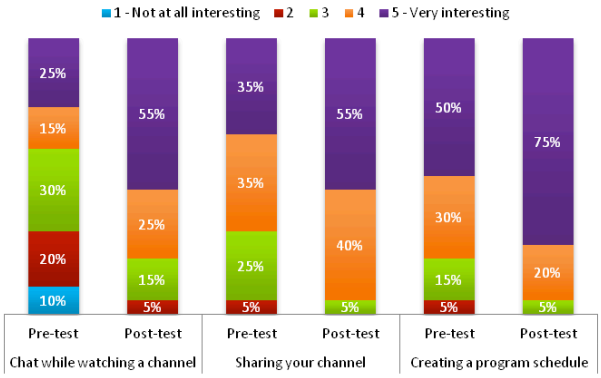
The results related with the *MyChannel* proposal are based on the analysis of the evaluation sessions and the post-test questionnaire taking into account the two groups of 10 users each, experienced [EX] and inexperienced [IN] with interactive television platforms.

Each evaluation session had an average duration of 12 minutes and 56 seconds, with the IN group spending 10% more time than the EX group. The difficulties experienced by participants with the prototype were few, being always possible to solve and did not hamper the participants from completing their tasks. Only a minority requested assistance from the observer that was in the room available to answer to any question or doubt.

5.2.1 *Post-test questionnaire*

This questionnaire was designed to gather feedback mainly regarding the concepts and functionalities tested and the User Experience (UX). Considering interface aspects the participants did not express significant problems. The icons/buttons were easily understood (80% EX and 70% IN) and correctly identified (70% EX and 50% IN). This trend applies also to the adopted colors and the aesthetic of the elements. Regarding the level of functionality of several *MyChannel* features, most of them were considered "functional" or "very functional" with a slight higher trend in the IN group. As an example: "*Creation of the channel*" was considered "very functional" by 70% EX and 80% IN

Several concepts proposed by the prototype were evaluated by means of asking the participants to specify their level of agreement with sentences about it. Despite some differences, the global tendency of the sample was positive towards the features described in those sentences. It is peculiar to refer that the less experienced group had more positive feedback towards the features. Most participants said they all had some or even great interest for the proposed features. 60% of the IN group answered to most features with the highest level of interest, but the EX group became more dispersed among the highest levels. We can highlight the: "*Rating and/or adding channels as favourites*" as one of the most interesting feature, with 60% of the EX group and 50% of the IN group answering "very interesting".



Graphic 1. Level of interest in pre-test and pos-test

Graphic 1 shows the level of interest towards three of the proposed features. It is important to highlight that the interest is higher than the reported in the pre-test questionnaire.

Considering their expectations the participants showed that they would like to be able to enjoy this application at their homes, since 50% of the IN group said it would be 'interesting' and 40% that it would be 'very interesting'. Only 10% responded with the average level. The EX group answers were between the highest (60%) and second highest (40%) levels of the scale. Participants were also favorable to the implementation of other features like the integration of a more "intelligent" search engine or the ability to add content to the library from the DVR of other users.

6. CONCLUSIONS

This research was targeted at conceptualizing a model for an application that allows users to create and share personal TV channels (using various types of content, to be viewed, recommended and evaluated by others) and evaluate this through the implementation of a prototype. Considering the very positive *MyChannel* results we might expect that the proposed model would provide important clues to the development of iTV applications that support the creation of personalized broadcasting channels. Regarding possible users with different levels of expertise with iTV platforms, the results show that an application like *MyChannel* would fit both types of users. *Mychannel* may now be reviewed in the light of the opinions and results obtained in the evaluation. Nevertheless, it is important to develop further studies to determine the opinion of other types of users, particularly the ones without a technological background.

For the near future, and given the highly positive evaluator opinions and their expectations of having *MyChannel* at home, an important next step would be the implementation of *MyChannel* in a commercial IPTV service. Very recently, a Portuguese IPTV operator launched a new service allowing their viewers to do precisely this by creating their own TV channels with content retrieved from on-line platforms, thus reinforcing the expectations for such a service [8].

7. REFERENCES

[1] Boxee. (n.d.). Boxee Bookmarklet. Retrieved January 20, 2012, from http://www.boxee.tv/watchlater.

[2] comScore. (2010). comScore Releases November 2010 U.S. Online Video Rankings. Retrieved January 19, 2012, from http://www.comscore.com/Press_Events/Press_Releases/2010/12/comScore_Releases_November_2010_U.S._Online_Video_Rankings.

[3] comScore. (2011). comScore Releases November 2011 U.S. Online Video Rankings. Retrieved January 19, 2012, from http://www.comscore.com/Press_Events/Press_Releases/2011/12/comScore_Releases_November_2011_U.S._Online_Video_Rankings.

[4] Fisher, L. M. (2006). Alvin Toffler: The Thought Leader Interview. *strategy+business*. Retrieved January 15, 2012, from http://www.strategy-business.com/media/file/sb45_06408.pdf.

[5] Hom, J. (1998). The usability methods toolbox handbook. San Jose State University. Retrieved January 19, 2012 from http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/UsabilityMethodsToolboxHandbook.pdf.

[6] IAB Platform. (2008). User Generated Content, Social Media, and Advertising — An Overview. Retrieved January 19, 2012 from http://www.iab.net/media/file/2008_ugc_platform.pdf.

[7] Lafrance J. (2005). Les télénautes, les pratiques mixtes de jeunes en télévision et sur Internet". In: Réseaux. Paris, 2005.

[8] PTC (2011). Meo Kanal. Retrieved March 3, 2012, from http://kanal.meo.pt.

[9] Obrist, M., Moser, C., Tscheligi, M., Alliez, D., Desmoulins, C., Teresa, H., & Miletich, S. M. (2009). Connecting TV & PC : An In-Situ Field Evaluation of an Unified Electronic Program Guide Concept. *EuroITV '09 Proceedings of the seventh european conference on European interactive television conference*.

[10] Pronk, V., Korst, J., Barbieri, M., & Proidl, A. (2009). Personal Television Channels: Simply Zapping through Your PVR Content.

[11] Tullis, T., & Albert, W. (2008). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Boston: Morgan Kauffman/Elsevier, 2008.

Interactive TV: Interaction and Control in Second-screen TV Consumption

Alexandre Fleury[1] Jakob Schou Pedersen[1] Mai Baunstrup[2] Lars Bo Larsen[1]

Aalborg University
Niels Jernes vej 12, 9220 Aalborg, Denmark
[1]{amf,jsp,lbl}@es.aau.dk; [2]mbauns06@student.aau.dk

ABSTRACT
The integration of television and mobile technologies are becoming a reality in today's home media environments. In order to facilitate the development of future cross-platform broadcast TV services, this study investigated prompting and control strategies for a secondary device in front of the TV. Four workshops provided about 1000 statements from TV consumers trying out working prototypes and engaging in discussions following a semi-structured interview approach. We explored if test participants liked to interact with TV content through a secondary device and which kinds of interaction types they preferred with which content. Overall, we found a clear preference for keeping interactive contents and prompting on the secondary device and broadcast TV content on the primary screen. The workshops generated numerous ideas concerning possible personalization of such service.

Categories and Subject Descriptors
H.5.2 Information Interfaces and Presentation (e.g., HCI) – User Interfaces, *Evaluation/Methodology*

General Terms
Experimentation, Human Factors, Measurement

Keywords
Second screen, iTV, Qualitative Study, Interaction, Prompting, Cross media.

INTRODUCTION
Television-related technologies have evolved vastly lately. TV is changing its form, with the consumer moving from passive reception of one way broadcasts to being a part of an interactive media experience. The TV audience is starting to get used to having a much larger degree of control over the TV content. Simultaneously, smart phones and tablets are making their way into the home. As a result more TV viewers engage in media multitasking activities such as browsing the web while watching TV: in Denmark, 59% report browsing the web on a smartphone of tablet while watching TV, and 45% of those focus on their Internet activity when doing so [8]. Today broadcasters are striving to support this evolution and provide cross-platform solutions to deliver content to their audience [2][13]. Communication between content providers and the end viewers increasingly becomes two-way instead of one way.

From a research perspective, it is therefore interesting to investigate how to successfully combine television and mobile technologies in a cross media, or second screen environment. By "second screen" or "secondary device" we refer to any device (smart phones, tablets, laptops) that allows TV consumers to interact with TV content displayed on a 'primary' screen (typically a home television set). It is fundamental to find out what types of interaction the TV consumers would like to engage in through a second screen application and how this should be designed.

The user study reported in this paper explored at a conceptual level how people would envision interacting with a second screen device while consuming broadcast TV. This paper focuses on how to engage viewers in interacting with live TV content in second screen setups. Furthermore, the question of whether the broadcast content should be separate from the interactive is investigated. These issues are studied through four workshops involving small groups of participants trying out prototypes specially developed to illustrate and challenge typical second screen scenarios.

As a brief outline of the paper, the next section presents related work conducted in the field of second screen setups; this is followed by an elaboration of the methods used to conduct the study; the study setup and data analysis method are presented before the results are reported and discussed. Finally, the conclusion summarizes the main findings and opens for future work.

RELATED WORK
Second screens have been on the agenda of interactive TV researchers since the mid-1990s, focusing on various aspects of the integration of the two devices, from content control to T-learning [4].

In order to investigate if viewers actually want to have the opportunity to interact while watching a TV show, and if this provides added value to the TV experience an experiment with 11 households was conducted in [1]. In this study the households were provided with a second screen prototype with which they were to interact while watching various TV shows for a period of three weeks. The qualitative data collected put forward twelve main topics of discussion, including general comments, liked features and issues. The enhancement of TV experience was found to be due to two factors: (1) the possibility of accessing extra relevant information immediately and after the show; (2) the broadening of the experience to outside the TV room and to an extended social circle. Synchronization and relevance of content, variety in

information sources, filtering of user generated content, and personalization of information were found necessary to ensure the success of such service.

In [3], four major usages of the secondary screen in an interactive digital television environment are investigated: control, enrichment, sharing, and transfer of television content. The latter, also referred to as presentation continuity, has been covered in a recent work investigating four specific methods for transferring video content from a mobile phone to a TV set in a ubiquitous home media environment [5]. The remaining three concepts of controlling, enriching and sharing content will be discussed in the present study.

METHOD

Goals

This study focuses on the second screen set-up, while watching TV in a domestic environment. In particular, we are interested in how to encourage viewers to use the second screen to interact with live TV content, or in other words what are their preferences in terms of prompting strategies? Furthermore, should the broadcast content be integrated with or separated from the interactive content and if so, how?

Methodology

The workshop approach was chosen as it allows presenting complex concepts with relatively simple prototypes. Furthermore a workshop allows gathering a group of test participants in a controlled environment, in which they may try out prototypes under the researchers' control. A workshop can be shaped in a variety of ways thus taking different directions depending on the variables included: duration, number of participants, content, and can include various techniques: interviews, card sorting, discussions or other initiatives relevant for the specific workshop. In our case, a semi-structured interview approach is selected in order to consider any idea brought up by test participants while still covering a set of predefined topics. In order to ensure the reliability and validity of our approach, we employed the 5-step verification process recommended by Morse in [10]: Methodological coherence; sample appropriateness; concurrent data collection and analysis; theoretical thinking; and theory development.

Participants

For qualitative studies of this kind, the recommended number of participants is 5-8, see e.g. [12]. However, to ensure sufficient coverage and validity of the results we repeated the workshop four times with different participants and carefully examined any lack of coherence between the four runs. In total 23 participants were included in the four workshops.

The age span needs to be large, due to the big differences in TV habits, interests, technical proficiency etc. that people have across generations. For the two first sessions we recruited males and females between 35 and 60 years old. University students in their early twenties participated in the last two sessions and were profiled as early adopters. All 23 participants received two cinema tickets as a thank you gift.

Content

The process of choosing the content for the workshop encompassed a total of 11 iterations and included brainstorming, discussions, selection, eliminations and testing of potential content combinations.

The final set of selected TV shows and typical interaction to be experienced by participants through prototypes is:

1. *Who Wants to be a Millionaire* (quiz show) – Answer questions simultaneously with live participants
2. *So Ein Ding* (consumer show) – Participate in a poll about a product reviewed
3. *Aftenshowet* (talk show) – Submit comments related to the program
4. *TV Avisen* (news show) – Retrieve more info about the news items presented in the show

WORKSHOP

The workshop consisted of three main parts:

Part one: The participants tried out four scenarios using different prototypes in order to obtain a clear understanding of the second screen concept.

Part two: The participants tried out two specific prototypes designed to investigate prompting strategies.

Part three: The participants discussed content / control separation by trying out a last prototype.

Setup

The workshop was conducted four times, each with two hours duration. The location was in a laboratory at Aalborg University which contained a conference table with room for six participants. Two facilitators ran the workshops and one media researcher represented the Danish Broadcasting Corporation (DR) during the two first workshops. The equipment in the room consisted of a 55" TV screen mounted to the wall at the end of the table and a number of iOS devices held by the participants (iPads, iPods, iPhones). The TV was connected to a desktop PC giving access to the prototypes on a web server.

The prototypes were developed and implemented as full-screen interactive web-apps. The TV content were pre-recorded clips of the TV shows mentioned in the previous section, and the web-apps content was synchronised with the one shown on TV. Figure 1 illustrates the prototype for the discussion regarding prompting strategies, and Figure 2 the prototype used to discuss content/control separation.

Data Capture and Analysis

Data collection was done with a voice recorder placed on the conference table and two webcams. The data collected comprise about eight hours of audio and video recordings – the latter mainly used for backup.

The first step of the analysis was "Meaning Condensation", i.e. only the essence of the answers and opinions is extracted and written in short precise phrases [7]. An external person carried out this task to insure against



Figure 2. Prototype illustrating prompting strategies: Via a ticker on the primary screen (top) or a popup message on the second screen (bottom).



Figure 1. Prototype illustrating separation of content and control: The live program plays on the TV (top), while the second screen displays either the interactive functions alone (bottom left), or with the video stream (bottom right).

experimenter bias. The condensed data contained approximately 1000 statements.

The second step consisted in coding the statements in order to decompose the data and rearrange it into categories that facilitate comparison between findings in the same category or between categories [9]. The categorization thematized the participants' viewpoints which uncovered essential issues. A coding scheme comprising six categories was developed and implemented for the study.

RESULTS

Statements about long lasting battery time, big screen size, fast system response and general ease of use were coherent with previous findings concerning second screens [1], mobile TV [6], and general usability principles [11]. Results specific to prompting strategies and content/control separation are presented in the next sections.

Prompting: Discreet On the Primary Screen

At first, most participants agreed that prompting should not happen on the primary screen. One said: "*It is annoying to look at prompting on the primary TV screen, and an advantage of being prompted on the secondary is that users have the control and can choose for themselves if they want to look at prompting or not*". However, participants also wondered how one would be made aware of the opportunity to interact with a TV show. E.g. "*I would prefer that the prompting occurred on the primary screen as I don't want to sit with my secondary device all the time waiting for this to happen*". As a result of the discussions, a very discreet prompting, e.g. an icon in the corner of the primary TV screen was suggested. This discussion puts forward an ambiguity inherent to the second screen paradigm: How to involve viewers in a secondary activity that takes away their attention from the primary screen while keeping their focus on the broadcast program?

Connected to this issue, one participant commented: "*I did not look up at the TV at any time, while interacting and watching the TV show on the secondary device.*"

The facilitator then asked if the second screen setup rendered the primary screen superfluous, and there was wide agreement on that this was not the case. This ambiguity is clearly due to the fact that TV consists of both audio and video – and in many cases the audio is quite sufficient for viewers to continue following a TV show even when engaged in other activities, e.g. on the second screen. This is expected to be especially true with low engagement TV shows such as entertainment or sport, which are particularly suited for second screen services.

Content and Control Should Be Separated

Discussions about having the TV show running on the secondary device generated much debate. However, a very clear conclusion emerged; participants agreed that the TV content belongs on the primary screen and interactive

content and controls on the secondary device. Furthermore having both content and control on the secondary device may render the primary screen irrelevant: "*When having the content along with the controls on the secondary screen the primary screen becomes unnecessary"*. In other words, this corresponds to just watching mobile TV.

Nevertheless, many of the participants stated that watching TV shows on the secondary device should be an option as it could be convenient under certain circumstances. Scenarios where this was mentioned to be relevant include when leaving the room where the primary screen resides without missing out on the TV experience or when the primary screen is occupied by other viewers watching a conflicting program: *"I would prefer to watch content on the primary screen, but would also like to have the opportunity to watch it on the secondary screen for situations where there is no primary screen available"*. The participants wanted to be able to control if the show should be running on the secondary device, in sync with the content broadcast on the TV screen.

CONCLUSIONS

We have presented a qualitative study of second screen use for TV. We designed and implemented a number of illustrative prototypes and presented those to two user groups in four workshops. Specifically, we addressed prompting strategies and the separation of TV content and interactive content. The conclusions from the study raise a number of design questions regarding the optimal division between primary and second screens.

Specifically, the outcome of the discussion clearly shows that prompting for interaction with live TV content should be very discreetly advertised on the primary screen to redirect viewers' attention to the secondary screen. It is anticipated that the practice of displaying a small icon or slightly animating the TV name's logo would be an efficient yet non-intrusive prompting strategy. Furthermore, the study participants demonstrated little interest in mixing live TV content and interactive functionalities on the same screen, neither the primary nor the secondary. For both age groups, the TV receiver is dedicated to content playback, while value adding interactive services belong to the second screen.

We believe designers of future second screen services linked to live TV content should consider these guidelines carefully in order to integrate the interactive content smoothly into the TV experience. Interactivity would thus become part of the show's flow, increasing not only the entertaining value for audiences, but also their involvement with the show and perhaps their loyalty to the show.

In general, the workshop approach proved to be well-suited as a data-gathering method and provided more in-depth information than e.g. a questionnaire survey, and consumed fewer resources than a corresponding longitudinal field trial. Conducting such study seems the logical next step to the work presented in this paper, providing invaluable insight in an ecologically valid test environment. Another interesting issue to further investigate would be to look closer into exploiting the "audio-only TV viewing" in a second screen context.

REFERENCES

[1] Basapur S., Harboe G., Mandalia H., Novak A., Vuong V., and Metcalf C. 2011. Field Trial of a Dual Device User Experience for iTV. *Proc. EuroITV 2011*, ACM, 127-135.

[2] Berthold, A., Schmidt, J. and Kirchknopf, R. 2010. The ARD and ZDF Mediathek portals. EBU Technical Review (2010 Q1). URL: http://bit.ly/f5HsdV (accessed January 12, 2012)

[3] Cesar P., Bulterman D.C.A., and Jansen A.J. 2008. Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content. Springer-Verlag Berlin Heidelberg.

[4] Cesar, P., Knoche, H. and Bulterman, C.A. 2010. From One to Many Boxes. Mobile Deevices as Primary and Secondary Screens. *Mobile TV: Customizing Content and Experience*. Springer, 327-348.

[5] Fleury, A., Pedersen, J. S. and Larsen, L. B. 2010. Evaluating Ubiquitous Media Usability Challenges: Content Transfer and Channel Switching Delays. *Design, User Experience, and Usability, Pt II*, HCII 2011, LNCS 6770, Springer-Verlag Berlin Heidelberg

[6] Knoche H. And Sasse, M.A. 2007. Getting the big picture on small screens: Quality of Experience in mobile TV. *Interactive Digital Television: technologies and applications*, IGI Publishing, 241-259.

[7] Kvale, S. and Brinkmann, S. 2008. *InterView – Introduktion til et håndværk. 2. udgave*, Hans Reitzels Forlag, 151-154.

[8] Larsen, A. L. & Svenningsen, U. H. (2011), Mange danskere surfer, når tv'et kører [Many Danes surf while the TV runs], in Medie udvikling 2011, DR Medieforskning, pp. 26--31.

[9] Maxwell J. A. 2008. Designing a Qualitative Study. *The handbook of applied social research methods, second edition*. Thousand Oaks CA: Sage Publications, 214-253.

[10] Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. 2002. Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods* 1(2), University of Alberta, 13-22.

[11] Nielsen, J. 1993. *Usability Engineering*. Morgan Kofmann.

[12] Nielsen, J. and Landauer, T. K. 1993. A Mathematical Model of the Finding of Usability Problems. *Proc INTERCHI '93* 206-213.

[13] Thompson, M. 2010. Putting quality first - The BBC and public space. BBC. URL: http://bbc.in/wTwWj6 (accessed January 12, 2011)

Elderly viewers identification: designing a decision matrix

Telmo Silva
CETAC.MEDIA/DeCA
University of Aveiro
Portugal
+351 234370200
tsilva@ua.pt

Jorge Ferraz de Abreu
CETAC.MEDIA/DeCA
University of Aveiro
Portugal
+351 234370200
jfa@ua.pt

Osvaldo Rocha Pacheco
DETI
University of Aveiro
Portugal
+351 234370200
orp@ua.pt

ABSTRACT
The central role that television-viewing plays in the daily life of elderly people motivate several research efforts towards the improvement of the benefits of companionship and information they seek from this media. Along with the advent of technical and interactive improvements, TV operators may provide the user with a more personalized television experience (with specific services that can, for example, increase the quality of life of the elderly) supported in a reliable viewer identification system. In this work, based on a prototype usage approach, guided interviews and direct observation of elderly viewers; we attempt to find the most suited viewer identification (vID) method for different combinations of social context, physical and psychological characteristics. A total of six vID methods were prototyped allowing the gathering of quantitative and qualitative data concerning the complex trade-off that influences the choice of the most suitable method. The analysis of these trade-offs was used to contribute for a decision matrix that, for a specific set of input values, returns the most appropriated vID method.

Categories and Subject Descriptors
H.1.2 [Models and Principles]: User/Machine Systems–*human factors*. H.5.2 [Information Interfaces and Presentation]: User Interfaces – *user-centred design.*

General Terms
Human Factors, Experimentation, Verification.

Keywords
iTV, Elderly, Viewer Identification, Privacy, User Experience, Personalization, Functional Classification, Decision Matrix.

1. INTRODUCTION
The World Health Organization (W.H.O.) reports an increasing rate of 2,7% per year in the group of people aged more than 60 years old (with a prediction of 20% of people in this group in 2050) [1]. This is an important issue, particularly in developing countries where actions to maintain older people healthy and active are considered a necessity and not a luxury [2]. As people get older, they tend to be large TV consumers with the potential advantage of companionship and valued source of information. The interactivity present in a raising number of pay-tv platforms can help this media play an even more important role when offering features such as: automatic audio volume adjustment, audio-description, personalised health care systems and community communication services. Currently, these features have the ability to enhance the elderly's viewing experience and their global quality of life. However, considering a typical usage scenario with multiple unidentified viewers using the same Set Top-Box (STB) there is an interesting advantage to have a reliable and automatic viewer identification method, allowing each user to benefit in a personalised way from the aforementioned features.

Defining the vID method more suitable for each specific set of user's characteristics was a complex task that relied on: i) understanding elderly needs, their limitations, motivations and behaviours when they are in front of the TV set; ii) prototyping a set of viewer identification methods (recognition of speaker voice, user finger print, personal card, tag in a dressing accessory and automatic or user controlled face recognition) and; iii) running a longitudinal study with the participation of elderly users. The outcome of this process allowed us to build a decision matrix that, for each specific usage context, suggests the most suited vID method.

This paper is organized in five main sections. In the next section the related work in viewer identification systems is presented. In section three the target audience is characterized. Section four describes the research methodology used to build the decision matrix; this section describes also some outputs that can be retrieved from the matrix. Finally, in section five, conclusions and future work are discussed.

2. RELATED WORK
One can say that the personalization of the viewing experience started in its simple form with the introduction of the remote control, easing the user ability to select the desired TV channel. After the remote control, Video Cassette Recorders, DVD players, along with the dissemination of other complementary devices, also improved the personalization of the user experience. However, nowadays with all existent interactive features, TV operators can offer a truly personalized iTV experience, namely if they provide their STBs with an automatic viewer identification system. However, operators must be aware that watching TV tends to be a relaxing activity and thus the identification system should be effortless and eventually transparent to the user.

Person identification relies on a physical characteristic or on something that people carry or know. The viewer identification systems proposed by Jabbar et al. [3] and by Brandenburg [4] are based on RFID (Radio Frequency ID) tags. Youn-Kyoung et al. [5] developed a system based on smart cards (the diference between these two types of systems is mostly technical and related with profile storage). Chang et al. [6] proposed a set of algorithms to infer user ID using data collected by accelerometers located in TV remote controls. According to the authors, this is a transparent method quite comfortable to the viewers. There are also academic studies centred on image processing algorithms that recognize the viewer's face [7, 8]. Despite the processing power this solution demands, it is also highly dependent of light and face position, which can have an impact on the user experience. Aligned with these experiments, TV set manufacters are incorporating cameras in their devices to detect if viewers leave the room for energy saving purposes. Recently, Samsung(TM) introduced a series of TV sets with cameras and sensors allowing viewers to interact using voice and gestures and to be identified through face

recognition [9]. Notwithstanding all these advances, there is a need to enhance iTV applications with the possibility to automatically identify the viewer or viewers in front of the TV set. This is specially relevant when iTV applications are targeted to elderly users, as is the case of iNeighbour TV project - a social TV application that aims to support social interaction and wellness [22].

3. ELDERLY VIEWERS

Elderly are strong TV consumers [10], although not very addicted and used to Information and Communication Technologies (ICT) [11]. When dealing with an iTV application, it is necessary to have a certain level of physical ability (for example to manipulate a remote control), visual acuity to read texts on screens, audio acuity, and cognitive abilities to understand information. Because of this, multiple difficulties arise in the elderly as they have limitations at these levels, e.g. deep navigation menus usually impose some difficulties due to short-term memory problems [12]; small and detailed icons are often illegible to the elderly; small remote controls could be impossible to use for persons with muscular tremor. These difficulties also need to be taken in consideration when deciding for a vID method suited to a specific user in a particular context.

4. DESIGNING A DECISION MATRIX

Regarding the available technical solutions to implement a vID system, we decided to study the following methods:

- SR - speaker recognition (with a built in microphone in the remote control);
- FP - finger print (with a reader over the remote);
- RC - RFID card (with a wireless reader near the user);
- WT - wireless tag (attached to a dressing accessory, e.g. a bracelet);
- CFR - controlled face recognition (the user needs to press a button in the remote to activate the recognition process);
- PFR - pervasive face recognition (with an always-on recognition camera).

With the help of 25 participants (13 female and 12 male, ranging from 55 to 93 years old) we followed an approach based in three phases:

i) exploratory interviews;
ii) matrix design (based on the results from the tests of a primary prototype);
iii) fill in the matrix (based on a full prototype).

It was decided to perform all the phases at the participants' homes as recommended e.g. by Obrist and Tscheligi [13] and Harboe et al. [14]. Some advantages of this methodology are related with the ability to: i) provide a relaxed atmosphere for the participants; ii) maintain them engaged in their natural environment; iii) collect data related with the usage context; iv) check the systems' advantages more easily; v) recruit new evaluators through a "friend of a friend" approach.

4.1 Phase 1: Exploratory interviews

We started our research with a set of five exploratory interviews, made to randomly selected participants, with the aim to tune the interviewing style and getting a better knowledge about the target users. The average age of this set of participants was 78.9 years old (with 11 years of standard deviation). In these interviews we had the opportunity to explain the concept behind iTV applications (with a focus on wellness) and about the benefits of having an associated vID system. With this preliminary approach we confirmed the need to have a relaxed and enthusiastic environment in the interviews; to create a tangible prototype and; to test it amongst participants with high digital literacy and younger.

4.2 Phase 2: Test a prototype to design the matrix

We started this phase with the development of a prototype to simulate a first set of two vID methods that allowed participants to be automatically recognised in an existent iTV application targeted to elderly users (the used application was developed under iNeighbour TV project). From the several offered features, we choose the medical reminder service of this application (that triggers alerts on top of the TV content reminding the viewer that he/she needs to take his/her medication) as an example of what they can benefit from being automatically identified by a vID system.

Nine new participants (5 women and 4 men) ranging from 55 to 66 years old (61.1 on average and 3.9 of standard deviation) had the chance to compare the already existent ID system (based on the input of a Personal Identification Number (PIN)) with the following two methods:

- WT - wireless tag inserted in a bracelet;
- RC - RFID card.

In this phase our main focus was to detect a trend amongst the several vID methods. Hence, after the tests the participants were interviewed to better understand their feeling about the prototyped vID techniques and opinions concerning the remaining identification methods: FP, SR, CFR and PFR.

Despite a light tendency to the PFR – Pervasive Face Recognition method, the spectrum of answers made it difficult to find a clear trend. However, it was clear that viewer characteristics' (physical and mental limitations, living context, etc.) influenced the choice.

Consequently it was decided to design a decision matrix (Figure 1) that considers two main groups of factors: i) technological factors related with technical efficiency, reliability, user-experience; ii) user factors which are related, for example, with living context and user's physical special needs. The matrix considers these factors for each vID method: SR, FP, RC, WT, CFR and PFR.

Regarding technical factors, that are independent of the usage context, the following were considered:

i) login and logout efficiency;
ii) implementation cost;
iii) privacy;
iv) user experience.

Concerning user factors, these were very complex to define and measure, since elders are heterogeneous users with a large spectrum of physical and psychological characteristics. Thus, it was necessary to select elderly physical and psychological functional classification concepts from gerontology. In this study it was considered a subset of the widely used (ICF) International Classification of Functioning, Disability and Health [15] to classify the participants. It allows the perception of multiple factors that influences the person's functional performance. ICF has four main areas to describe individual's functioning: body, activities, participation and contextual factors. Unlike other classification systems, in the ICF, all context (environmental and personal factors) details are factors that can enhance or limit the

performance of persons. Thus, using the ICF, it is possible to track how persons experiment limitations considering their possible weakness, illness or handicaps.

The ICF classification system uses a specific nomenclature to define the parameters used to classify persons. This nomenclature comprises a letter (B for body functions; S for body structures; D for activities and participation; and E for environmental factors) and a number to specify the detail evaluated.

Supported on the ICF, research aims and previous results; we defined the following parameters as matrix entries for user factors (Figure 1):

i) users' ICT knowledge (high skilled, medium; low) measured according to the European Commission Report [16]; this factor is important because an inexperienced ICT user will probably select a vID method that implies fewer operations.

ii) visual acuity; to measure this parameter we used the Jaeger Eye Chart (JEC) test [12]. The results were represented in a 3 level scale: high - no visual problem (reads J4 in JEC); medium - visual problems not affecting system usage (reads J6 in JEC); low - severe visual problems (reads J12 in JEC). This is an important issue due to the difficulty that people with low vision have in, for example, finding a regular button with a specific function in a remote control. In the ICF, this parameter is included in the subset of sensory functions and pain with the reference b210 – Vision function [17].

iii) hearing acuity, measured using the whisper test [18]; the results were represented in a 3 level scale: high - no hearing problem; medium - hearing problem not affecting system usage; low - severe hearing problem. In the ICF, this parameter is included in the subset of sensory functions and pain with the reference b230 – Audition function [15].

iv) orientation [17], which is a parameter related with persons' psychological capabilities. To gather data to evaluate this factor a set of questions about person's address and current year was used; we also measure this parameter with a three values scale. Orientation is also included in the ICF classification, in the subset of mental functions: b114 – Orientation function[15];

v) voice ability; also measured in a scale of three values: High, Medium and Low; we considered it because if a person has difficulties to speak this can influence the choice regarding a speaker recognition method. In the ICF, this parameter is included in the subset of voice and speaking functions with the reference b310 – Voice function [15].

vi) memory; as in the previous parameter, this was measured using a similar three value scale. To gather this information we used a set of questions about issues referred in a previous conversation between the researcher and the participant; this parameter directly influences the ability to use some vID methods. In the ICF, this parameter is included in the subset of mental specific functions with the reference b144 – Memory function [15].

vii) fine motor skills; to measure this type of participants' skills, the Nine Hole Peg Test [19] was used. The researcher simply points out the time each participant takes to complete the test with the dominant hand. This parameter does not have a direct comparable measurable factor in the ICF, but it is very important to our work because it influences the person's ability to manipulate objects such as the remote control. We also classified the participants in three levels [19]: High – without fine motor problems (the participant that take less than 30 seconds to complete the test); Medium (the participant that take less than a minute); Low - a severe problem (more than one minute to finish the test).

viii) mobility; to evaluate this parameter we used the timed "Up & Go" test [20], also measured in a scale: High - no problem (the person takes less than 10 seconds); Medium - probably the participant has a mobility problem (less than 30 seconds to complete); Low - severe problem (more than 30 seconds). In the ICF classification, this measure has a high correspondence to parameter d450- walk [15].

For each parameter, the ICF defines a qualifier to characterize the functional level. These qualifiers are presented in a very detailed scale. However, in this work, we decided to shrink this scale to three possible values for each parameter.

	Tec. factors					User factors																												
	Logout efic.	Login efic.	Cost	Privacy loss	User Experience	Dig. Literacy			Visual acuity			hearing acuity			Orientation			Voice ability			Mobility			Memory			Fine Motor Skills		
						low	med	high	low	med	high	low	med	high	low	med	high	low	med	high	low	med	high	low	med	high	low	med	high	low	med	high
RC																																		
FP																																		
SR																																		
PFR																																		
WT																																		
CFR																																		

Figure 1. Decision Matrix (in percentage)

4.3 Phase 3: Fill in the matrix

After building the matrix it is necessary to fill in the cells with the percentage of preference for each type of method from each type of user. Only after a full completion of this stage, the matrix will be settled to help decide which vID method is the most suited to a specific usage context. In order to populate the cells, a prototype based on the Wizard of Oz approach was developed [21]. With this approach participants were led to think they were interacting freely with the system, but actually there was a researcher manipulating it through the prototype, by remotely controlling the ID process. This approach was used to simplify the development phase.

Until now, to fill in the matrix, a set of eleven tests was conducted at participants' homes where they tried all vID methods. These sessions also included the functional tests described in previous section, in order to classify the participants according to all decision matrix parameters. At the end of the session participants were presented with a circular sketch representing all the tested methods; the form was designed considering that presenting technologies in a predefined order inducts elderly decision. Then, they were asked to assign a number (from 1 to 6) to each of the experienced methods ("1" representing the first most preferred).

Using the results of the small set of tests already conducted, it is possible to observe that if a person has a high literacy level he/she will probably choose a vID based on a finger print reader in the remote control (FP). The same happens, if the person has a high Fine Motor Skill. However, if the person has a medium "Fine

Motor Skill" he/she tends to prefer the Speaker Recognition (SR) with a microphone installed in the remote control.

Concerning technological factors, at this time, we only gathered data concerning "Privacy Loss" and "User Experience". To fill in these parameters, for each method, participants were asked about i) their felling of invasion of privacy and ii) how easy and how enjoyable it was. With the data collected until now, it is possible to verify that Finger Print reader tends to provide the best user experience and that Pervasive Facial Recognition induces the higher level of privacy loss.

5. CONCLUSIONS AND FUTURE WORK

This work aimed to design a decision matrix to ease the process of finding the vID method most suited to a specific user in a particular context. After a set of exploratory interviews, it was possible to tune the interviewing style and get a better knowledge about the target users. Then, a first prototype was developed and tested with elderly. The results showed that each user, in a specific context, tends to prefer a particular method. This lead to the design of a decision matrix that we wish to fill in with data gathered from the tests of a Wizard of Oz based prototype.

At the moment, it was only possible to fill in some matrix cells. For example, regarding the level of voice ability we only had participants scoring in the highest level so we cannot fill in the cells corresponding to medium and low level. However, concerning hearing acuity, our sample is a bit richer because it comprises persons with medium and high acuity. Despite that, the results gathered so far allow us to be strongly motivated to endeavour this type of methodology in a future research project where a large number of participants will be recruited.

Future work will also encompass the visual representation of the outcomes of the matrix in order that their potential users, e.g. iTV providers or caregivers' networks, can benefit from this work in a straightforward manner.

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from FEDER (through COMPETE) and National Funding (through FCT) under grant agreement no. PTDC/CCI-COM/100824/2008. Our special thanks to all participants for allowing us to take their time and patience.

7. REFERENCES

[1] Division, D.O.E.A.S.A.P., *World Population Ageing: 1950-2050*, U.N. Publications, Editor. 2001: New York.

[2] Organization, W.H. *World Health Organization launches new initiative to address the health needs of a rapidly ageing population*. 2004. Retreived 2-1-2012; Available from: http://www.who.int/mediacentre/news/releases/2004/pr60/en/.

[3] Jabbar, H., et al., *Viewer Identification and Authentication in IPTV using RFID Technique.* Consumer Electronics, IEEE Transactions on, 2008. 54(1): p. 105-109.

[4] Brandenburg, R.v., *Towards multi-user personalized TV services, introducing combined RFID Digest authentication,* 2009.

[5] Youn-Kyoung, P., et al., *User Authentication Mechanism Using Java Card for Personalized IPTV Services*, in *Proceedings of the 2008 International Conference on Convergence and Hybrid Information Technology*. 2008, IEEE Computer Society.

[6] Chang, K., J. Hightower, and B. Kveton, *Inferring Identity Using Accelerometers in Television Remote Controls*, in *Pervasive Computing*. 2009: Nara, Japão.

[7] Hwang, M.-C., et al., *Person Identification System for Future Digital TV with Intelligence.* IEEE Transactions on Consumer Electronics, 2007.

[8] Mlakar, T., J. Zaletelj, and J.F. Tasic. *Viewer authentication for personalized iTV services*. in *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS '07. Eighth International Workshop on*. 2007.

[9] Samsung Electronics Co., L. *Samsung Smart LED and Plasma TVs Usher in a New Era of Connectivity and Control*. 2012. Retreived 26-1-2012; Available from: http://www.samsung.com/ca/news/newsRead.do?news_seq=29921&gltype=localnews.

[10] Marktest, G. *Portugueses viram cerca de 3h30m de Tv em 2010*. 2011. Retreived 9-02-2012; Available from: http://www.marktest.com/wap/a/n/id~16e0.aspx.

[11] Taborda, M.J., *A Utilização de Internet em Portugal 2010*. 2010, Obercom: Lisbon.

[12] Pölönen, M. and J. Häkkinen, *Near-to-Eye Display—An Accessory for Handheld Multimedia Devices: Subjective Studies.* Journal of Display Technology, 2009. 5(9): p. 358-367.

[13] Obrist, M., R. Bernhaupt, and M. Tscheligi. *Users@Home: Implications from studying iTV*. in *20th International Symposium on Human Factors in Telecommunication*. 2006. Sophia-Antipolis, France.

[14] Harboe, G., et al., *Ambient social tv: drawing people into a shared experience*, in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. 2008, ACM: Florence, Italy. p. 1-10.

[15] Organization, W.H., *World Health Organization: The International Classification of Functioning, Disability and Health (ICF)*. 2001.

[16] Commission, E., *Digital Literacy European Commission Working Paper and Recommendations from Digital Literacy High-Level Expert Group*. 2007.

[17] Muras, J.A., V. Cahill, and E.K. Stokes. *A Taxonomy of Pervasive Healthcare Systems*. in *Pervasive Health Conference and Workshops, 2006*. 2006.

[18] Saudi, P., P. Tracey, and G. Paul, *Whispered voice test for screening for hearing impairment in adults and children: systematic review*. Vol. 327. 2003, London, ROYAUME-UNI: British Medical Association. 967-970.

[19] Oxford Grice, K., et al., *Adult norms for a commercially available Nine Hole Peg Test for finger dexterity.* Am J Occup Ther. 57(5): p. 570-3.

[20] Podsiadlo, D. and S. Richardson, *The timed "Up & Go": a test of basic functional mobility for frail elderly persons.* Journal of the American Geriatrics Society, 1991. 39(2): p. 142-150.

[21] Dahlbäck, N., A. Jönsson, and L. Ahrenberg, *Wizard of Oz studies: why and how*, in *Proceedings of the 1st international conference on Intelligent user interfaces*. 1993, ACM: Orlando, Florida, United States. p. 193-200.

[22] Abreu, J. F. d., P. Almeida, et al. (2011). *Participatory Design of a Social TV Application for Senior Citizens – The iNeighbour TV Project ENTERprise Information Systems*. M. M. Cruz-Cunha, J. Varajão, P. Powell and R. Martinho, Springer Berlin Heidelberg. 221: 49-58.

Text-TV meets Twitter – what does it mean to the TV watching experience?

Pauliina Tuomi
Tampere University of Technology
Pohjoisranta 11 A
P.O. Box 300, 28101 Pori, Finland
sptuom@utu.fi

Sami Koivisto
YLE
Radiokatu 5
00024 Yleisradio, Helsinki, Finland
sami.koivisto@yle.fi

ABSTRACT
In Finland the national broadcaster company YLE (Finland's national public service broadcasting company) has done experiments in creating social TV by combining text TV and social media for a couple of years now. YLE has offered viewers an opportunity to tweet on Twitter via text TV page in real time simultaneously with live TV broadcast. This has been done mainly around different large and nationalistic media events that are approached as media spectacles in this study. The communication done with tweets has been possible for example during the Eurovision Song Contest, Independence Day celebration and now currently as a part of presidential election of Finland. This focused on the Text TV & Twitter broadcasting of Eurovision Song Contest and Finnish Independence Day celebration in 2011. The purpose of this study was to firstly explore for what and how is this social media channel used by the audience –what does audience tweet about? For this, all the (overall 3005) tweets published on text TV were analyzed based on content analysis method.

Categories and Subject Descriptors
H.5.2 (**Information Interfaces and Presentation**): User Interfaces – evaluation/methodology, user-centered design, graphical user interfaces (GUI)

General Terms
Documentation, Experimentation, Human Factors

Keywords
Multiplatform TV-formats, media spectacles, Text-TV, Twitter

1. INTRODUCTION
The digitalization, internationalization and marketing of the media have created new competitive online and mobile communication forms that have changed television in particular. This development is approved by the increased competition on the one hand and cooperation on the other hand with online media [1]. Television, the so far "lonesome rider", has become a part of the big picture of interconnected devices, operating on several platforms. [11] Broadcasters are eager to provide new ways to drive viewer engagement.

In Finland there is a more innovating way TV content and watching experience can be supported with the assistance of the Web platform on the TV screen itself than just having the Twitter feed on different second screens for example smart phones. "Linnan Juhlat" is a Finnish yearly event that is held to celebrate and honour the independence of Finland. In 2010 YLE[1] enabled for the first time people to watch the event and at the same time follow the Twitter conversation on the Text TV in real time. The president of Finland invites Finnish people that have been noteworthy and have forwarded equality and good values during the present year. This event is one of the most followed TV broadcast in Finland. The Twitter meets Text-TV-experiment is also been done during the media spectacle Eurovision Song Contest that is a popular yearly song contest that all the active European Broadcasting Union (EBU) member countries are allowed to participate in.

2. BACKGROUND OF THE RESEARCH
2.1 Text TV & Twitter
Text TV & Twitter-experiment is technologically executed with the appropriate hashtag # for example #euroviisut by which the published tweets end up trough the RSS-feed into moderated publishing system. By choosing the text TV page 398, the tweeting of the audience absorbs onto the TV screen as a part of Eurovision TV broadcast viewing the same way as texting does.

Text TV as a phenomenon is mostly European and nowadays it is popular for example in Britain (BBC) and Sweden in addition to Finland [12]. The very first teletex tryouts took place in Britain in the 1970's. YLE started regular text TV broadcasts 7th of October in 1981 and as a curiosity at the end of 1990's, 70% were using text TV regularly. [12] For some reason text TV has not really been studied academically at all, despite the fact that text TV has survived as an existing feature all these years. [12] Social media, Twitter in this case, on the other hand is in the centre of a buzz. More detailed research on the linkage and its potential is necessary, as both industry and academia have become interested in the topic [11]. There are research done from several angles for example the Twitter itself and how and why people use it. [9,14] Nowadays also the connections between TV and social media start to be widely studied in order to analyze its enhancing features on the watching experience. [8] Although TV and the Internet are the most popular entertainment media people use in their everyday lives, the linkage between these two major platforms has so far been the focus of just a small number of scientific studies. The use of Twitter and its effect on the viewing rates are constantly been calculated in order to learn more and to boost watching rates of the TV companies. [5]

[1] YLE: http://www.yle.fi

2.2 Research methods and theory

The meeting between mass and digital media opens a number of suggestive avenues for research [4]. In general, it is clear that research both on the production and reception side needs to draw on and combine existing traditions of research on broadcasting, Web based, and mobile media [13]. In this paper, the term *multiplatform* is used to include formats that feature a cooperation of television and Web platforms as well as various peripheral platforms. Intermedialism is definitely one of the key issues in today's TV watching experience. According to Henry Jenkins the old and new media are communicating all the more complex ways [7]. Audiences are combining the use of different mediums in order to gather a coherent media experience.

In this research, media spectacle is seen as an event that lives in a wide range of different mediums. [10] However, media spectacles are events that already exist in the minds of the audiences. Usually because of the long history or national impact of the format. These spectacles have lasted a long period of time; they feed the sense nationality and create a feeling of togetherness and collectiveness. Media spectacle offers audiences both TV- and web platforms to attend and participate. Media spectacles happen mainly on the TV screens, but are also widely known among people also as non-TV events – Independence Day party, WC ice hockey tournaments and Eurovision Song Contest. [10] A similar term to media spectacle is the concept of the media event or televisual ceremony. [3] Media events demand and receive focused attention and intense involvement from the largest possible national and international audiences [2].

The research method employed is a qualitative content analysis. In this study 1423 tweets published on text TV during the Finnish Independence Day Celebration 2011 and 1582 tweets published during the Eurovision Song Contest 2011 were analyzed. Firstly, they were explored in order to come up with the most dominant themes that appear in both of the Finnish media spectacles. Secondly, the most frequently appeared themes were analyzed and categorized into seven categories. There were some difficulties in categorizing the tweets since the categories were often over lapping and the tweets were suitable for more categories than one. In these cases evaluating the most evident theme decided the final selection. Also the actual time of the tweet also affects the theme; the most nationalistic tweets were e.g. written when Finland's Eurovision presenter was performing.

The categories are *1) Eurovision Song Contest artists and performances/Independence Day quests & appearances, 2) Nationality, 3) Text-TV and Twitter- experience, 4) Interaction and dialogue, 5) Irrelevant comments and notions, 6) hosts & general arrangements* and 7) *Media spectacle & atmosphere.* The main interest of this study is the Twitter experiment itself and the use of it as a part of the media spectacle.

3. RESULTS

There are a large number of tweets being analyzed in this study, which is why only the most dominant themes are presented while others are briefly summed. The percentages of all themes are presented in Table 1.

Table 1. The amounts (%) of the analyzed Twitter themes

Categories	1)	2)	3)	4)	5)	6)	7)
Eurovision Song Contest 14-15.5	**28**	6.7	1,5	7,8	**48,9**	2	5,1 = 100%
Linnan Juhlat 6-7.12	**26,3**	2,7	12,9	10,6	**37,9**	4,3	5,3 = 100%

3.1 Eurovision Song Contest

A large amount of Twitter tweets (28%) sent during the Eurovision Song Contest finals 2011 dealt with the performing artists, their appearances and naturally the songs.

"Azerbaidzan's song is superficial. The singer (woman) does not have enough voice either." 14.5, 23.31

"The hair of the Danish singer looks like radish leaves.." 14.5, 22.25

There were also tweets (6,7%) that dealt with national aspects of Eurovision Song contest in this case from the Finnish point of view. Tweets supporting Finnish artist Paradise Oscar (the Finnish performer in 2011) and tweets referring how does it feel to be a Finn around Finland's performance.

"Us Finland! I feel like crying! 14.5, 22.16"

"Finland! I'm here at the sofa like rest of Finns! Go Finland, Go Oscar!" 14.5, 22.05

"I must say I'm so proud of the Finnish artist (Paradise Oscar)!!" 14.5, 23.54

The nationality and also the atmosphere of the media spectacle are also shown in the tweets (5,1%) that deal with the tweeter's habits to view and enjoy the atmosphere of Eurovision Song Contest. There are certain traditional ways that Finnish people spend this media event yearly. The expression of these habits to other co-watchers seems to play an important role.

"Champagne, snacks, TV, Twitter – The Eurovision-studio is ready!"14.5, 22.05

"I'm here like every other year – shall we be together?" 14.5, 21.59

"Well, this event sure has changed since ABBA won and we gathered in front of our black and white telly after sauna to watch the show. :)" 14.5, 23.25

This all emphasizes the idea of ESC being a media spectacle in its nature. It is more than just a TV program. It is a phenomenon that involves participating audience in its atmosphere year after year.

The actual use of the combination of TV and Twitter was also one of the themes (1,5%). The Twitter experiment was overall appreciated and the comments were all positive.

"The following of the #euroviisut hashtag has been the best part of the whole event!" 15.5, 00.23

"These songs have gone by quickly while tweeting!" 14.05, 23.37

"Yle! Leave the Twitter-text TV-feed on for awhile more for the aftermath discussions" 15.5, 01.26

3.2 Independence Day Celebration

A fair amount of tweets (2,7%) sent during the Independence Day Celebration 2011 dealt with Independence Day traditions and feelings of nationality. Just like in ESC tweets, which emphasizes that these both are strong media spectacles that are embedded in Finn's minds and are still based on TV as a main cluster of the event – text TV/the Twitter giving the platform for expressing feelings collectively (5,3%).

"Wow. I love Twitter. I've never felt this much connected and Finnish before. I also taught Kid how to waltz while we watched #linnanjuhlat" 6.12, 22.56

"We are getting ready here. Food is in the oven, cleaning is done. 6.12, 18.49

"Here we are having cheese, apple, pear, ginger bread. Drinking Jaffa (I'm sick, so no wine).." 6.12, 18.52

"Twitter, champagne and TV opened. The holy threesome of Independence Day!" 6.12, 19.08

The theme of Twitter in synch with TV broadcast was also discussed. (12,9%) The Twitter experiment was again widely positively adopted also during the Independence Day celebrations.
"Fun to tweet to television" 6.12, 22.24
"Twitter has done these festivities much funnier! And I don't feel all lonely being here by myself." 6.12, 18.59
"#Linnanjuhlat trends worldwide!!! Ah, I just love this moment!" 6.12, 19.18

One of the most appeared themes (26,3%) was the appearance of the Independence Day quests – their habitus and especially dresses.
"Linnan juhlat, here I am ready for the dresses!" 6.12, 18.46
"Hey, what is that awful sparkling vomit on that dress?" 6.12, 19.14

It is a long-lasted tradition in Finland to criticize and evaluate the outlook of the quests the president has invited. For example the prettiest and the most hideous women of the evening are chosen/voted the next day on online magazines every year.
"Here we ago #linnan juhlat, let's see who messes up, forgets the etiquette and has the ugliest dresses! :D" 6.12, 18.54

3.3 Summing up

The largest number of the tweets (48,9% & 37,9%) included e.g. general statements, inside jokes and other uncategorized issues. However, it would be challenging to present all of subthemes in this paper. To be brief, this category included e.g. discussion around the phenomena, individual opinions on several aspects, voting procedures on Eurovision Song Contest, who should have been invited into the ball and who's not and for example the ingredients of Independence Day's punch.

"The funniest thing is the foreigners asking what is this #linnanjuhlat and why are people just shaking hands" 6.12, 19.52
"I'm waiting for my friend to appear in the hand shaking queue. Nervous.." 6.12, 19.31
"There's a great atmosphere in the castle!" 6.12, 22.11
"Hababaha" 14.5, 00.37

Naturally people were communicating with each other with every tweet sent. It was however interesting to look at these ways of communication in more detail. There were some narrative and re-tweets sent between the audiences, which emphasizes that Twitter+text-TV actually brings us closer to the idea of social TV, especially since the interaction takes place on the TV screen itself. (7,8 & 10,6%) Communication does not require any other devices or second screens. The clear dialogues, in both cases, were often question/answer-type and quite many of the viewers actually took part in the conversations.

"Any idea how often the press choice has actually won the whole thing?" 15.5, 00.42
"The press choice has won two times: Lordi and Rybak. This is what they say on Yle's website." 15.5, 00.53

"Anyone? Have you seen Toivola (member of the Finnish parliament)? 6.12, 21.35
"He's been spotted on the dance floor shaking hands with someone!" 6.12, 21.38

However, the majority of the Finns were making a statement rather than actually waking up conversation. Still the dialogue was there, audience was talking collectively even when they did not address the comments and notions to some particular tweeter.

"I wonder how many will actually be tweeting inside the castle this year? 6.12, 18.35
"Good night and happy Independence Day! Thanks for the company to all of you! See you next year! 6.12, 23.10
"I'm here in spirit! I hope we won't get 0 points this year! 14.5, 22.05
"Any ides is there going to be any breaks? I need to go to the refrigerator?" 14.5, 23.10

The tweeters were also suggesting many new ideas how to exploit Twitter in the future, as a part of different TV formats and in which ways.
"Why couldn't you vote via Twitter? :D" 15.05, 00.03
"Next year we should tweet questions to the hosts that interview quests and suggests songs to be played while they are dancing, what do you think?" 6.12, 22.29
"Thanks for the Text-TV- Linnan Juhlat! Next time in Eurovision Song Contest. And how about again in UEFA Football ?" 6.12, 22.26

The use of Twitter seems to make the whole show/event worth watching in both cases.
"It's great to read the tweets since you can see the opinions of the Finnish people" 6.12, 20.25
"I couldn't watch Linnan juhlat without Twitter comments anymore, great entertainment!" 6.12, 20.29

Also the technique and bringing text TV and Twitter together were appreciated.
"This combination of poor man's internet (text TV) and web (twitter) is handy. Exciting! 6.12, 20.40
"I have never watched Linnan juhlat. Twitter makes miracles ☺" 6.12, 22.58
"YLE is mashing Twitter and text TV excellently! 6.12, 21.48
"Twitter and text TV has really boosted this event, difficult to keep up with the new pace!" 6.12, 19.27

4. DISCUSSION

TV broadcasters and –channels are communicating more and more with audience via Facebook and Twitter. The audience is being heard and at the same time they are able to interact in real-time with each other TV content related. Twitter & TV – combination is seen as a very functional way of creating the long wanted form of social TV. One reason is the low costs for tweeting. [14] Other one is the fact that Tweets do not need much time to write or read and one has an easy access via PC, notebook or smart phone. [8] Twitter Monitoring is functional method to do TV related audience research. There are however some limitations of the method. First, it only investigates active users of the Social Web. Second, there are still only very few Twitter users in Finland (Twitter itself does not publish the actual amount of the users but many countries have estimated the number of users) and third, the sociodemographic information about tweet authors often cannot be determined. In sum, Twitter Monitoring cannot replace people meters or surveys but it complements them. [8]

For example like in this study, the general themes people were discussing of and the overall opinions concerning the Text TV & Twitter – experiment were defined by analyzing the content. However, it was not possible to study the tweeters in more detail, at least not in this first round of the study. In all the opinions and comments are alone valuable since they do give insights on what is the general response among the Finns.

5. CONCLUSION

TV is in major transition where it is evolving from just scheduled broadcast to being always- on, participatory and social. Its traditional audience is now fully engaged with social media and here they effectively become "program" and content makers for each other but, in the back channel, they still drive conversation around traditional, compelling broadcast content [6]. The use of social media is re-shaping the linkage between TV and Web. Social aspects that it brings are affecting the collective TV experiences. In this paper the contents of tweets sent during the two popular Finnish media spectacles that engage large audiences were analyzed and presented. There were seven categories to be found and this paper tackled the most dominant categories in more detail. The people were using the Twitter opportunity for example to discuss the actual content of the spectacle (Eurovision artists and Independence Day quests appearances), the feeling of nationality, dialoguing with each other and the collectiveness the atmosphere brings to the watching experience and the Twitter experience itself. In both cases the experiment of combining social media into TV was appreciated. The text TV & Twitter-experiment works well also in more serious issues. YLE organized the same service with the presidential elections in Finland 2012. YLE published overall 4683[2] tweets during the elections and eventually the chosen president of Finland also tweeted his thanks to voters via Twitter. YLE is expanding these tryouts to several different TV formats for example around different talk shows with special themes etc. In the future the data of tweets will be analyzed in more detail – both qualitatively and quantitatively. At this point the purpose is to give an insight of what people are discussing of via Twitter and how is the experiment adopted. More tweets will be gathered in order to collect material that can be compared for example in yearly bases.

This study identifies that there is potential in remediating older technology (text-TV) with newer (social media). The analyzed tweets reveal that social communication that take place during the actual live broadcast is mainly used to express opinions, something one would like to say to the person sitting next to one on sofa, but there is actual dialogue to be found as well. The most valued features tweets on text-TV did brought to the audience were the idea of presence, a possibility to 'take part' in a new way and the possibility to tweet was seen overall as a positive extra value to more traditional ways of watching TV.



Picture 1. An example of the Twitter feed on TV - picture is from YLE's broadcast on Finnish presidential elections 2012.[3]

[2] http://yle.fi/uutiset/tuloslahetykseen_twiittailtiin_ahkerasti/3195932

[3] http://yle.fi/uutiset/teemat/presidentinvaalit/2012/02/twitter-kommentointi_vilkasta_tuloslahetyksissa__myos_vastavalittu_presidentti_twiittasi_kiitokset_3233724.html

6. REFERENCES

[1] Ala-Fossi, M., Herkman, J., and Keinonen, H. 2008. The methodology of radio- and television research. University Press – Juvenes Print, Tampere.

[2] Buonanno, M. 2008. The Age of Television. Experiences and theories. Intellect Bristol, UK/Chigaco, USA.

[3] Dayan, E. & Katz, E. 1992. Media Events. The life broadcasting of history. Harvard University Press, Cambridge.

[4] Deuze, M. 2006. Collaboration, participation and the media. New Media Society, 8(4): 691–698.

[5] Edison Research. 2010. Twitter Usage In America: 2010. The Edison Research/Arbitron Internet and Multimedia Study. Online: http://click.oo155.com/ViewLandingPage.aspx?pubids=7237|77|5&digest=1lrSuztl0vmcpiIREQpx0Q&sysid=1

[6] Hayes, G. 2009. Social IPTV: Interactive and Personal. SPAA 2009. http://www.slideshare.net/hayesg31/social-iptv-interactive-and-personal?from=ss_embed

[7] Jenkins, H. 2006. Convergence Culture. When old and new media collide. New York –university Press, New York & London 2006.

[8] Jungnickel, K. & Schweiger, W. 2011. Twitter as a television research method. General Online Research Conference GOR 11, March 14-16, 2011, Heinrich Heine University, Düsseldorf, Germany.

[9] Kwak, H., Lee, C., Park, H. & Moon, S. 2010. What is Twitter, a Social Network or a News Media? Proceedings of the 19th International World Wide Web (WWW) Conference, April 26-30, 2010, Raleigh NC (USA). Online: http://an.kaist.ac.kr/~haewoon/papers/2010-www-twitter.pdf (06.06.2012).

[10] Tuomi, P. 2010. The role of the Traditional TV in the Age of Intermedial Media Spectacles. In Proceedings of the 8th international interactive conference on Interactive TV& Video 2010 Tampere, Finland June 09 - 11, 2010. pp. 5-14.

[11] Tuomi, P & Bachmayer, S. 2011. The Convergence of TV and Web (2.0) in Austria and Finland. In Proceedings of the 9th international interactive conference on Interactive television. Lisbon, Portugal — June 29 - July 01, 2011. ACM New York, NY, USA 2011. Pp. 55-64.

[12] Turtiainen, R. 2010. Tulos ei päässyt edes teksti-TV:lle. Miksi vanhanaikainen teknologia on säilyttänyt asemansa digitalisoituneessa mediaurheiluympäristössä? Tekniikan Waiheita 4/2010. 32-48.

[13] Ytreberg, E. 2009. Extended liveness and eventfulness in multiplatform reality formats. New Media & Society, 11(4): 467–485.

[14] Zhao, D. & Rosson, M. B. 2009. How and Why People Twitter: The Role that Micro-blogging Plays in Informal Communication at Work: Proceedings of the ACM 2009 international conference on Supporting group work. Online: http://www.personal.psu.edu/duz108/blogs/publications/group09%20-%20twitter%20study.pdf (06.06.2012)

Contextualising the Innovation: How New Technology can Work in the Service of a Television Format

Sarah Martindale
University of Nottingham
Horizon Digital Economy Research
Triumph Road, NG7 2TU
+44 (0)115 8232558
sarah.martindale@nottingham.ac.uk

Stuart Reeves
University of Nottingham
Horizon Digital Economy Research
Triumph Road, NG7 2TU
+44(0)115 8232554
stuart.reeves@nottingham.ac.uk

Elizabeth Evans
University of Nottingham
Department of Culture, Film and Media
Trent Building, NG7 2RD
+44 (0)115 9514241
elizabeth.evans@nottingham.ac.uk

Paul Tennent
University of Nottingham
School of Computer Science
Jubilee Campus, NG8 1BB
+44 (0)115 8466532
paul.tennent@nottingham.ac.uk

Joe Marshall
University of Nottingham
School of Computer Science
Jubilee Campus, NG8 1BB
+44 (0)7905 696427
joe.marshall@nottingham.ac.uk

Brendan Walker
Aerial
258 Globe Road
London, E2 0JD
+44 (0)20 89830293
studio@aerial.fm

ABSTRACT

This paper introduces research investigating the potential benefits of integrating 'biodata' – captured physiological information – into television formats. The aim is to provide new resources to support viewers' interpretation of, and engagement with, the feelings of onscreen protagonists. To this end the project seeks to design vicarious television experiences for audiences. The example discussed here is *The Experiment Live*, a prototype show based on the live broadcast of an amateur paranormal investigation of a 'haunted house'. As part of a wider research agenda, this paper focuses specifically on how the format is constructed and understood, analysing it as part of the genre of supernatural reality TV, considering how the innovative added layer of biodata functions within the format and presenting some initial audience feedback in the context of work in progress.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems **–** *Evaluation/methodology*

General Terms

Design, Human Factors

Keywords

Biodata, television format, engagement, paranormal investigation, reality TV, vicarious

1. INTRODUCTION

The medium of television is changing at a startling rate in terms of technology, content and viewing context, and this has significant implications for both the creation of new formats and audience responses to those formats. Channels and platforms multiply, while the communal television set has been joined by other more portable and discrete screen devices (laptops, smartphones, tablets) that can be used to access both core programme and ancillary content. This content can be time-shifted and interacted with in increasingly complex ways that go far beyond the previous scope of the broadcast schedule. All of which means greater flexibility for the viewer in terms of where, when and how they consume television. At the same time as providing a more individuated and convenient experience, these changes also mean that television (as mechanism, text and activity) is fragmenting. Under these circumstances it is more important than ever that television programmes engage the viewer and engender a sense of sharing in a cultural encounter in order to maintain the loyal viewership still required by television's funding structures.

The Vicarious project sets out to go beyond the close-up and other formal features of television by developing new technology and techniques for communicating and connecting with the experiences of people on television. Capturing and displaying biomedical data (heart rate, respiration, galvanic skin response) has the potential to reveal information about someone that they may not even be aware of. This project builds on previous research that has technically and theoretically investigated ways of creatively incorporating biosensors and monitoring technologies into a number of rides, games and interactive experiences [7]. The current research objective is to explore the possible applications of biodata within different television formats, which would take different forms and meanings according to the nature of those formats. Therefore, in addition to addressing the production challenges associated with the practice, examined elsewhere [10], part of the work involves understanding prototype formats in the context of existing television content and conventions and qualitatively exploring how audiences respond to formats that include biodata. This paper represents an initial examination of these issues in relation to a particular example.

Of course physiological readings have featured in media before. The concept of a display presenting 'life force' is a longstanding feature of computer games (e.g. *The Legend of Zelda*) and visualisations of vital signs have similarly been used in dramatic narratives to communicate and create tension around characters' corporeal status (e.g. *Aliens*). Television programmes like *House* depict monitoring technologies within the medical context, while others like *The Chair* appropriate them for entertainment

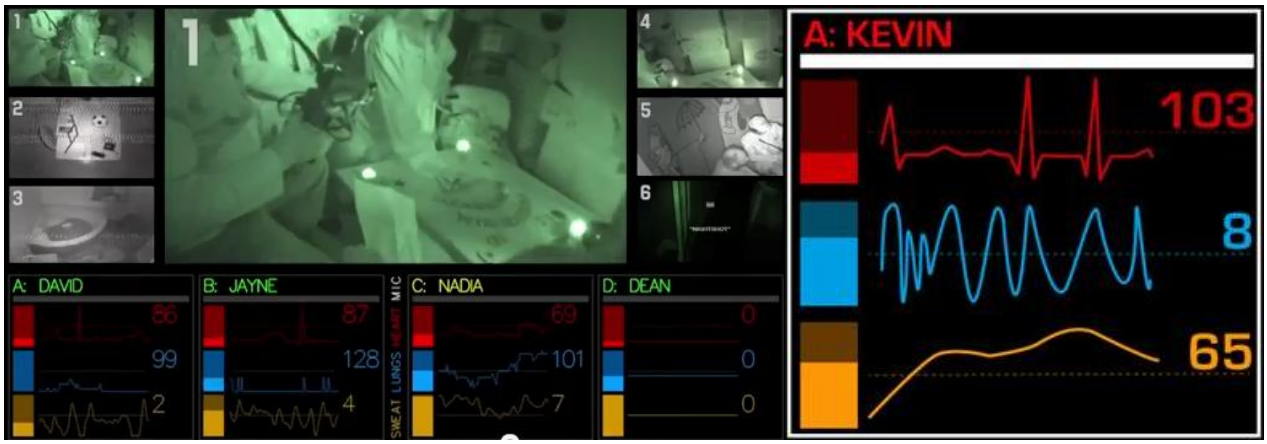


Figure 1. The arrangement of the onscreen display (left), and detail of a biodata visualisation (right).

purposes. As part of the presentation of competitive situations showing contestants' heart rates can demonstrate both physical exertion (e.g. *Tour de France*) and self-control (e.g. *Poker Dome Challenge*). This project seeks not only to develop techniques to capture and utilise other biomedical measurements in addition to heart rate on television, but also to examine how the 'reality' of that data is represented, perceived and valued. Doing so may contribute new elements to the viewing experience and even enable new forms of interaction with television in the form of second screen applications or haptic feedback devices.

2. FORMAT OUTLINE

By monitoring the physiological responses of participants engaged in a paranormal investigation, *The Experiment Live* (*TEL*) aimed to convey and give access to their heightened states. The scripted event required four volunteers assuming a role as 'paranormal investigators' in order to explore the basement of a Nottingham tearoom, which, according to the back-story created, was the site of a haunting by the infamous 'sobbing boy'. The success of *TEL* depended on not only a traditional expositional script, but also a strong 'affective' script, designed to elicit specific emotional responses from the participants. While relying almost entirely on suggestion, certain set pieces were used to drive forward the experience, in particular bringing the biodata to the fore at key moments. For example, the narrative included a key moment during which the biodata capture unit (BCU) of one participant would 'fail' (i.e. staged failure), and subsequently be removed from the participant and left in public view. The participants would then perform a séance, during which the discarded BCU would start presenting 'ghostly data' for the duration of the séance. The appeal of this format was intended to arise from the experience of witnessing a horror scenario alongside the effects of that scenario on the minds and bodies of the participants.

In order to achieve the aims of the format the production team behind *TEL* needed to explore strategies for delivering biodata in an understandable way to an audience. Each visualisation of an individual's physiological information consists of three streams of data, presented either directly, or with some filtering. Heart data (Figure 1, red) is displayed as a direct electrocardiogram (ECG) trace, as well as a numerical heart rate. Respiration (Figure 1, blue) is displayed as a direct trace of chest expansion as well as a numerical respiration rate. Finally, stress, which was actually represented by galvanic skin response (GSR) in this case (Figure 1, yellow), is displayed as a graph of rate of change in GSR, along with an immediate numerical value, in order to highlight changes in arousal, particularly sudden shocks or moments of calm experienced by participants. This data was relayed in real time, along with six live video streams, one of which was centralised by an editor to call attention to the action (Figure 1). This layout was designed to give the impression of viewing a CCTV feed, with the associated connotations of having direct, unmediated access to events.

The biodata (and video) was streamed to a production room above the 'haunted' basement, where live editing was performed. The editing was designed to highlight the ongoing narrative and included: choosing video streams to draw attention to in the central window and causing cameras to 'fail' (switching to static); mixing the audio to either level out participants' speech, or at times to subtly emphasise which participant the action was following; and the synthesis and mixing of the 'ghost's' biodata. In addition, a live biogenerative soundtrack was created, mixing predetermined melodies with variances automatically produced in response to the data captured (though mediated by a composer). This allowed the music to reflect the physiological states of the participants and further served to draw attention to the biodata.

Through manipulation and selection of component parts of the broadcast display, the production crew worked to create and manage a complex array of different combinations of formal elements. Production work involved making use of a variety of pre-arranged static camera perspectives as well as opportunistic ones provided by both participant and professional roaming camera operators (as opposed to a traditional arrangement of shots being 'offered up' by operators). These views were selectively arranged alongside visualisations of the biodata (as seen in Figure 1), and both could be manipulated in the ways outlined above (e.g. cameras failing and the introduction of fake biodata). It was a live broadcast event, which allowed the production team to confront the challenges involved in simultaneously recoding and transmitting video, audio and biodata. Subsequently a 40 minute version of the paranormal investigation based on the live broadcast material was produced, which presents it in the form of a prototype supernatural reality TV format. This television show is the focus here and it needs to be situated within a wider genre for the purposes of analysis.

3. REPRESENTING SUPERNATURAL 'REALITY' ON TELEVISION

The idea of constructing a situation in which drama and entertainment value can be derived from a paranormal investigation is not new. Building on the Victorian tradition of public séances, several television shows have sought to present the 'reality' of contact with the spirit world. In 1992 the BBC broadcast *Ghostwatch* on Halloween. This drama, written by Stephen Volk and starring well-known television personalities Michael Parkinson, Sarah Green, Mike Smith and Craig Charles, triggered controversy because many viewers believed that they were witnessing real, and really terrifying, events. The programme achieved this startling impact as a result of the frame that it built around the haunting. It employed conventions from factual television with an on-location presenter (Green) and camera crew monitored from a studio nerve-centre by an anchor (Parkinson), an expert commentator (Dr Pascoe, played by Gillian Bevan) and an audience. The impression of live interaction with ordinary members of the viewing public was created by having telephone operators ostensibly taking calls in the studio (reported on by Smith), while Charles interviewed bystanders on location.

Recently the acclaimed British documentary filmmaker Adam Curtis has argued that *Ghostwatch* succeeded in blurring the boundaries between fact and fiction so effectively because it encapsulates the paradox that while viewers are aware that things reported on television are not necessarily true, television coverage nevertheless makes events feel real for audiences. As a consequence television 'possesses our imagination more powerfully precisely because we don't know what is real and what is not' [1]. This ambiguity is played upon by reality TV programmes like *Most Haunted* (2002-) and emulators such as *Ghost Hunters* (2004-), which document paranormal investigators exploring and conducting experiments in purportedly haunted locations. These programmes are based on a claim of veracity: by presenting direct action, with the production crew in shot and night-vision cameras to reveal people's reactions, the format sets out to show truthful experiences rather than tell a ghost story [6]. Periodically a live version of *Most Haunted* is produced, which reproduces the studio/location configuration of *Ghostwatch* very closely, demonstrating again that the trappings of reality are compelling whether employed in the context of fact or fiction.

Complaints have been levelled against *Most Haunted* that it is fraudulent and deceptive and consequently potentially harmful to vulnerable or susceptible viewers [8]. Indeed the possibility of fakery has even caused friction among those actually involved in making the show [9]. The position taken by the British media regulator Ofcom is that *Most Haunted* is clearly produced for entertainment purposes and therefore the authenticity of events depicted is not a matter of concern [8]. It would seem then that what is primarily at stake in these supernatural television formats is not the reality of the paranormal activities depicted but the excitement of witnessing participants' responses when these suggest that they believe that they may be experiencing something uncanny. This is the glimpse of reality that audiences look for within the factual entertainment frame [5].

4. NOVEL FEATURES OF THIS FORMAT

TEL incorporates key features of both *Ghostwatch* and *Most Haunted* – the pretext of scientific investigation, multiple perspectives on the action, an apparently transparent setup – but seeks to intensify the emphasis on the feelings of the paranormal investigators and increase opportunities to relay these to the viewers. It does so by placing ordinary people in an extraordinary situation and signalling the physical reactions that this provokes by means of the biodata. By inviting viewers to come to their own conclusions about the status of what is happening onscreen paranormal entertainment formats cultivate emotional and psychological commitment [4]. Multiplying the streams of audio-visual and biodata information offered to the viewers (as detailed above) serves to promote an even greater sense of involvement in *TEL* through interpretive agency. The act of choosing to focus on and switch attention between different elements of the broadcast display gives the impression of direct participation in a process of piecing together and making sense of the unfolding action. Each viewer can make their own connections or identify tensions between the various pieces and types of information, and thereby take part in a simultaneous process of narrative construction. The split screen aesthetic is self-consciously artificial [2] and therefore fits with the unfamiliar biodata visualisations in the sense that meaning is not obvious and has to be deduced.

The search for and identification of points of interest or significance is given added impetus by the perception that anything might happen at any second. The frame of paranormal investigation heightens the dramatic tension by raising the possibility of (and audience desire for) unexpected occurrences. The fear generated by *Ghostwatch* (and many other horror texts, of course) stems from glimpses of indistinct but creepy things in the background or corner of shots. An awareness of these generic tropes prompts the anticipation that evidence of paranormal manifestations will be discernable in *TEL* to frightening effect. Having six separate video streams amplifies anxiety about where the seemingly inevitable shock will come from, and the biodata means that the nervousness of participants is communicated through raised heart rate, respiration rate and change in GSR as well as body language, facial expression and verbal interaction. Breakdowns in technology underline the impression of uncertainty and associated unease, and also invite speculation about their cause. The perception that paranormal activity is interfering with equipment fits with a longstanding belief that communication technology can provide a conduit connecting the living with the dead [3]. Thus technical glitches serve as groundwork for the climax of the programme, a simulation of technological possession.

There is a porous boundary between being in front of and behind the camera in *TEL*, which gives the impression that the programme offers unrestricted access to the events being shown. When the production room is illustratively presented at the start of the programme as a setting for on-camera interviews with participants, viewers are shown that the production interface is the same as the onscreen display, and can therefore infer that they are being provided with all the available information and authentic biodata. Immediately after the investigation has ended, the participants are shown being briefly interviewed in front of the audience who watched the live broadcast (on a big screen in a cinema). This part of the format underlined the importance of audience participation in the co-performance and co-production of the television event as a collective experience [4]. Establishing physical co-presence with an audience substantiates the reality of the participants and, by extension, asserts the truthfulness of their actions and reactions as depicted in the show. The impact of this moment in front of an onscreen audience was intensified by means of a clever detail of production design: the participants appeared

wrapped in emergency foil blankets, suggestive of them having been through an ordeal. These techniques cultivate a feeling of affinity with the participants and thereby support the new component of biodata in terms of promoting the impression of getting inside the experience.

5. SOME AUDIENCE RESPONSES AND IMPLICATIONS

Three members of the audience, opportunistically recruited, who chose to attend the live broadcast of *TEL* as part of the Mayhem Horror Film Festival in October 2011 shared their opinions about it with researchers in a focus group held immediately after the event. Their responses to the format suggest engagement with the formal features discussed above. It was possible to see this complex process of deriving entertainment value from a supernatural reality TV format at work in the case of *TEL*. A 'naturally sceptical' viewer asked about an incident in the show when a plate smashed. When told by researchers that this was rigged he did not question the programme's validity, but responded thus: 'well that was quite cool, I liked that. It added to the atmosphere'. Another of the focus group participants repeatedly described herself as being 'engrossed' in the viewing experience, which she described as 'a sensation based thing for me, experiencing the emotions that I was going through at the time'. The role of the biodata seemed to be problematic in relation to engagement, however, as this comment illustrates: 'some people were having -5 sweat. I don't even know what that means'.

What this points to are issues around the presentation and interpretation of biodata. These are being delineated more fully in the course of four focus groups discussing the 40 minute show, which have recently been conducted with locally recruited volunteers willing to give their time to watch and then share their opinions. Analysis of this qualitative material is underway and the key points that are emerging concern the difficulty of making sense of biodata, especially without any prior knowledge of the concept. Viewers seem to trust the veracity of the information but do not know what to do with it. They find the display complex and illegible and suggest that a more mediated application would be preferable. As an alternative to constant readings, intermittent simplified visual/aural representations or expert commentary is proposed to draw attention to abnormal or interesting responses. The overriding impression at this early stage is that explicit guidance and direction is seen to be necessary, indeed desirable, in order for viewers to feel they can identify the significance of biodata as an element of a television format.

The production and reception of television relies on individual and shared frameworks of understanding by which meaning is assigned to and derived from audio-visual characteristics. It is necessary to develop technological systems and creative methods by which to integrate biodata into these frameworks in ways that endow it with value and consequence. This is a challenge because biodata is an unknown quantity for viewers. If the meaning of biodata can be made to seem self-evident then it will be possible to incorporate the technology as a source of engagement into a range of television formats. While qualitative accounts of a viewing experience can provide information about the interpretive resources viewers are drawing on in the process of trying to make sense of biodata, it is very difficult to isolate the part played by the biodata within responses to the format. In order to address this, the next part of the project will be a controlled study applying signal detection theory within a game show format in order to test whether biodata impacts on people's ability to identify truthfulness or deception.

6. CONCLUSION

This analysis details the ways in which various formal features of *TEL* as a supernatural reality TV format encourage and support active and invested audience engagement. The integration of biodata in this case builds on established television conventions of depicting paranormal investigations and modes of response to those conventions. By utilising a range of technologies, techniques, perspectives and types of information *TEL* reveals more about the experiences of participants and thereby boosts the entertainment value of a format that relies on audience engagement to provide a vicarious thrill for viewers and disturb their position of comfort and safety. In doing so it negotiates with complex perceptions about the authenticity of reality TV and the believability of paranormal media [5, 4]. At the same time it asks viewers to develop new interpretive repertoires in order to attach visceral, emotional and narrative significance to biodata. As a preliminary iteration of the research aims, *TEL* and this paper point towards the dimensions being explored in the ongoing project.

7. ACKNOWLEDGEMENTS

This work is supported by Horizon Digital Economy Research, RCUK grant EP/G065802/1.

8. REFERNCES

[1] Curtis, A. 2011. The ghosts in the living room. *The Medium and the Message* (December 22, 2011). URL= http://www.bbc.co.uk/blogs/adamcurtis/2011/12/the_ghosts_in_the_living_room.html

[2] Evans, E. 2011. *Transmedia Television: Audiences, New Media, and Daily Life*. Routledge, New York and London.

[3] Goldstein, D. E., Grider, S. A. and Thomas, J. B. 2007. *Haunting Experiences: Ghosts in Contemporary Folklore*. Utah Statute University Press, Logan.

[4] Hill, A. 2011. *Paranormal Media: Audiences, Spirits and Magic in Popular Culture*. Routledge, London and New York.

[5] Hill, A. 2005. *Reality TV: Audiences and Popular Factual Television*. Routledge, London.

[6] Koven, M. 2007 Most Haunted and the convergence of traditional belief and popular television. *Folklore* 118, 2 (Aug. 2007), 183-202.

[7] Marshall, J., Rowland, D., Rennick Egglestone, S., Benford, S., Walker, B., McAuley, D. Breath control of amusement rides. In *Proc. CHI*, 2011, ACM.

[8] Ofcom. 2005. Most Haunted/Most Haunted Live. *Ofcom Broadcast Bulletin* (December 5 2005).

[9] Roper, M. 2005. Spooky truth: TV's Most Haunted con. *Mirror* (October 28, 2005).

[10] Tennent, P., Reeves, S., Benford, S., Walker, B., Marshall, J., Brundell, P., Meese, R., Harter, P. 2012. The machine in the ghost: augmenting broadcasting with biodata. In *Ex. Abs. of Proc. CHI*, 2012, ACM.

iTV in Industry

from Meaning to Sense

what two-way service providers might do in the content curation era

To-Do

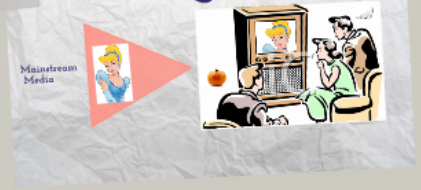
- Develop Social Framework
- Develop User Digital Identity Management Platform
- Enter the User Social Sphere
- Understand User Preference and Willingness to Pay based on User Behavior, Location, Presence
- Provide Consistent Content/Services/Ads



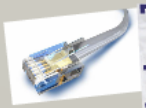




The Broadcast Paradigm



The Two-Way Paradigm





Pitch

bio



details

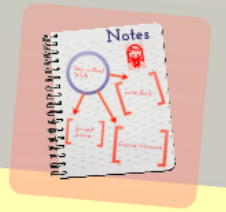






pics or didn't happen





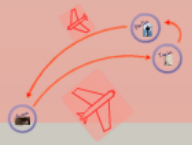
notes

Prezi

Pitch

bio

Antonio Pavolini has been working in the media sector for over 16 years. After a long standing experience in PR multinationals, like Burson-Marsteller and Brodeur (where he was Director of the Rome Bureau for 3 years, with EMEA-Wide assignments for several top-tier ICT corporations), he became Corporate Communication Director at Saritel, the company in charge of managing Internet Data Centers for the Telecom Italia Group. From 2009 he covers the role of Senior Analyst, Media Industry in the Innovation Department of Telecom Italia, with the responsibility of identifying key trends and the most promising start-ups in the media industry. He recently contributed to the book"Connecting Television" with an essay dedicated to the New Business Models of OTT-TV and Web TV.



details

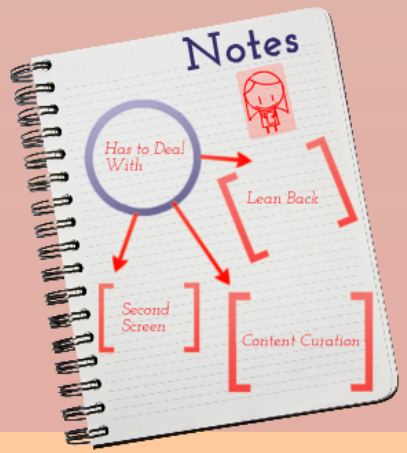


antonio.pavolini@telecomitalia.it



and remember that we can chat later on this issue :)





notes



pics or didn't happen

Prezi

TV Recommender System Field Trial Using Dynamic Collaborative Filtering

Rebecca Gregory-Clarke
BBC Research & Development
56 Wood Lane
London W12 7SB UK
+44 3030 409643
rebecca.gregory-clarke@bbc.co.uk

Chris Newell
BBC Research & Development
56 Wood Lane
London W12 7SB UK
+44 3030 409747
chris.newell@bbc.co.uk

ABSTRACT
Matrix Factorization techniques have performance and scalability advantages when used for personalised content recommendation. However, the requirement for offline training means that it is difficult for this approach to accommodate new users and new content items without significant delays. Recently, high-performance algorithms have become available that can accommodate new users and items using online update methods. This paper presents the results from a field trial which examined the effectiveness of these methods when used to recommend TV programmes. Our results show that this approach outperformed an alternative metadata-based approach in terms of the perceived quality and variety of the recommendations.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**] Information Search and Retrieval – *Information filtering, Relevance feedback, Selection process*.

General Terms
Algorithms, Measurement, Performance, Experimentation.

Keywords
Personalised Recommendations, Bayesian Personalised Ranking, Collaborative Filtering, Matrix Factorization.

1. INTRODUCTION
Collaborative filtering (CF) systems based on past user behavior are an effective tool for building content recommender systems. Experience with large datasets such as the Netflix Prize data [1] has shown that Matrix Factorization algorithms have performance and scalability advantages compared to alternative techniques. However, these algorithms require a time-consuming training phase to determine a set of latent factors for each user and content item. As a result there is a significant delay before the most recent user feedback can influence recommendations and before new users and items can be accommodated. This is not ideal for online TV services where there is a continuous flow of new programmes and where only a minority of users can be persistently identified.

High-performance collaborative filtering algorithms have recently become available that can accommodate new users and items using online update methods [2]. These methods create and add new latent factors to an existing model and are fast enough for online use. For example, the collaborative filtering algorithm evaluated in this trial could process the new user information using the online update methods in less than 20 milliseconds per user. This enables the possibility of a two-stage approach, where a model is first created using offline training and then updated dynamically in response to additional user feedback. This approach is very suitable where historical records of viewing activity exist only for a minority of users. Models can be trained using this historical data and then updated dynamically to provide recommendations for additional users for whom there is only more recent viewing information.

In this paper we report on a user-centric field trial where this approach is evaluated, and compared with a metadata-based alternative. We also compare these systems to a non-personalised recommender, which is used as a reference point.

2. RECOMMENDER SYSTEM
As explained, we wish to evaluate the performance of a collaborative filtering recommender which makes use of an online update mechanism. This and the two algorithms we use for comparison are detailed below.

The recommender system was implemented using the Java version of the MyMediaLite Recommender System library [3]. All three were based on positive-only feedback provided by the users i.e. a seed list of programmes they have watched, or would watch.

2.1 Collaborative Filtering Algorithm (BPR)
For the collaborative filtering system we used a Matrix Factorization algorithm optimized for Bayesian Personalised Ranking [4]. Offline training data was obtained from media server logs and consisted of 68256 viewings, involving 303 TV series and 4856 users who had persistent accounts. Please note that this algorithm will be referred to as BPR throughout the rest of this study – however we use this to refer to the whole algorithm including the online update mechanism that we wish to evaluate.

2.2 Metadata-based Algorithm (MB)
For metadata-based filtering we applied Cosine similarity to the hierarchical genres assigned editorially to each programme. We found that weighted hierarchical matching, taking into account partial matches was more effective than exact matching since some sub-genres match only one available programme. Including other attributes (e.g. actors) had little benefit because matches were rare within the available set of programmes.

2.3 Non-personalised Algorithm (MP)
A non-personalised algorithm based on programme popularity (total number of viewings) was included in the trial as a baseline. This algorithm is referred to as Most Popular.

3. FIELD TRIAL METHODOLOGY

The trial was implemented as a web application. The main sections of the trial are outlined below.

3.1 Capture of Online Training Data

To capture online training data the user is first presented with a list of programmes from which they can select ten items that they either *have* watched, or *would* watch. The programmes were split into genre categories to ease navigation. It should be noted that this capture interface was developed for experimental purposes rather than operational use.

The number of programmes was fixed at ten following some preliminary tests, as it was considered to be a high enough number to produce reasonable recommendations, without being so high as to become arduous for the user.

3.2 Presentation of Recommendations

The three subsequent pages of the trial each contained a set of ten programme recommendations (one for each of the three algorithms). The three algorithms were used in a random order for each user, to prevent time-biasing. Each recommendation included a title, description and screenshot from the programme.

On each of these pages the user was invited to give their feedback about the set of recommendations, via a questionnaire which is explained below.

The number of recommendations presented to the user (per algorithm) was fixed at ten, as it was considered a sufficient number for the user to judge the overall quality of the recommender, without the task becoming tedious; as three sets of recommendations were presented to the user in quick succession.

3.3 The Questionnaire

We considered three main indicators when assessing the overall quality of the recommendations:

- **Quality** - is the recommender producing 'good' recommendations?
- **Variety** - is the recommender suggesting a wide range of content, or just 'more of the same'?
- **Novelty** - is the recommender suggesting interesting content that the viewer might not previously have considered?

The questionnaire used a 5-level Likert scale, where the participant was asked to say how well they agree with a particular statement, according to the following scale:

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

In order to reveal any inconsistencies in the responses; the questions are posed in pairs, where the question is asked once from a negative, and once from a positive point of view. There are three sets of question pairs such as this, representing quality (questions 1 and 4), variety (questions 2 and 5) and novelty (questions 3 and 6). The questions are ordered such that the pairs do not appear together.

As another measure of how well the recommender had performed, the user was asked how many of the recommendations (out of ten) they considered to be 'good'. Lastly the participant was asked if they had any additional (optional) comments they would like to make about the recommender.

The full list of questions is given below in the order that they appeared to each of the 55 participants:

1. The recommender is providing good recommendations
2. The recommendations were all from the same genre
3. The recommender is recommending interesting content I hadn't previously considered
4. I didn't like any of the recommendations
5. The recommendations contained a lot of variety
6. The recommender is recommending interesting content that I already know about
7. Of the ten recommendations, how many would you consider to be good?
8. Any other comments (Optional)

4. FIELD TRIAL RESULTS

4.1 Analysing the Results

In analysing the results for a given algorithm, overall mean scores for quality, variety and novelty can be calculated. This is done by inverting the scores of the negatively posed questions (questions 2, 4 and 6), and then calculating a mean score for each of the question pairs, over all 55 users.

However, whilst there was good correlation between the question pairs for both quality and variety, it was found that the correlation was not as good between the two questions concerning novelty. After consideration it was decided that these questions were not well-matched and should be treated separately rather than as a pair. We have therefore only considered question 3 when analysing the novelty within the results.

4.2 Mean Quality, Variety and Novelty Scores

The following graphs show the mean scores awarded by the participants for each algorithm.

According to Figure 1, both the BPR (collaborative filtering with online update) and MB (metadata-based) algorithms perform significantly better than the MP algorithm in terms of overall quality, as we might expect.

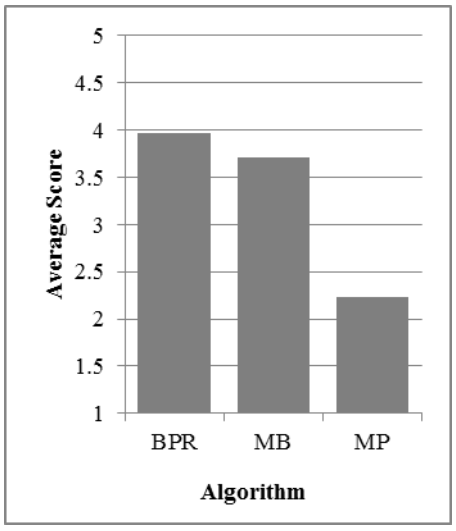


Figure 1. Recommendation quality scores

While BPR does appear to outperform MB, the results do appear relatively close when we consider the dramatic difference between these and the results for the MP algorithm. However, a paired, two-sample t-test of the hypothesis that BPR outperforms MB in terms of the mean quality score reveals a p-value of 0.027, indicating that this hypothesis is statistically significant.

Figure 2 indicates that BPR also performs significantly better than MB in terms of the perceived variety of the recommendations. This is perhaps not surprising because the MB algorithm is based on content similarity and will not recommend programmes outside the genres represented in an individual user's feedback. Collaborative filtering algorithms, such as BPR, do not suffer from this limitation. Variety appears to be a significant factor in the perceived quality of a set of recommendations. However, the results for MP show that a high score for variety alone is not enough to warrant a high score for quality.

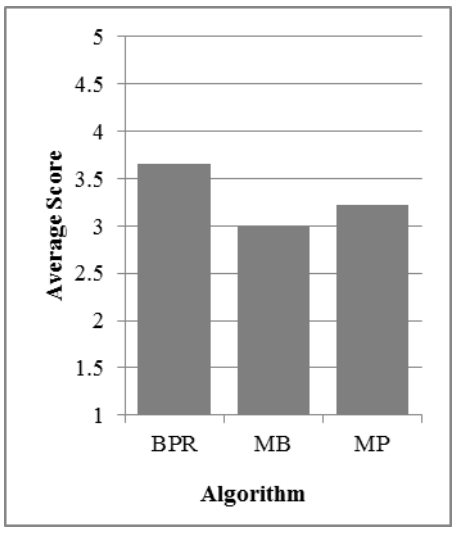


Figure 2. Recommendation variety scores

Figure 3 shows that BPR and MB perform similarly well in terms of the perceived novelty of the recommendations and both perform significantly better than the MP algorithm. This supports user feedback suggesting that novelty is also a significant factor in the perceived quality of recommendations.

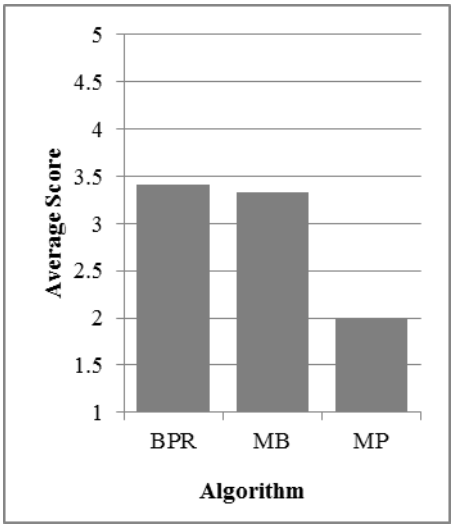


Figure 3. Recommendation novelty scores

4.3 Number of 'Good' Recommendations

Table 1 shows the average number of recommendations that the users considered to be good. We can see that both the collaborative filtering and metadata-based algorithms clearly outperform the non-personalised recommender, which supports the observations made above. Again, the BPR algorithm appears to produce the best results.

Table 1. Mean number of 'good' recommendations produced by each algorithm

Algorithm	Mean number of good recommendations (out of ten)
BPR (collaborative filtering)	5.7
MB (metadata-based)	5.0
MP (non-personalised)	2.0

4.4 Frequency Distribution of Quality Scores

Figure 4 shows the distribution of the quality scores awarded by the 55 participants. Smooth line plots have been overlaid on the graphs below for easy comparison between the algorithms. These distributions are useful as it allows us to see the range of user opinions rather than just an average, and allows us to see how well the users agree with each other. Note that the quality scores are able to take half-point values due to the overall quality score being the average of the answers to two questions.

The distribution for the MP algorithm shows a peak at 1.5 and a secondary peak at around 3.5. The distribution of scores is clearly weighted towards the lower end of the scale, showing that generally it did not perform very well, however the secondary peak shows that there were a number of people who also considered the recommendations to be reasonable.

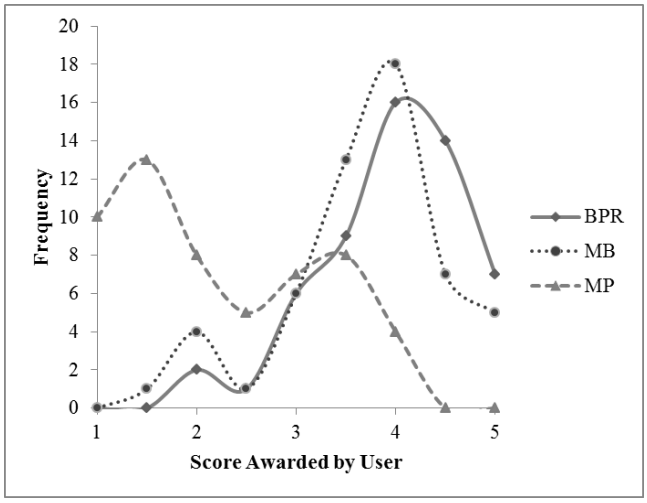


Figure 4. Frequency distribution of quality scores

The BPR and MB algorithms both have distributions with higher peaks at around 4, showing more consensus between the users that the algorithms were performing quite well. However, we can see that MB has a slightly wider distribution than the BPR algorithm. The narrower, BPR distribution does not show as many people giving low scores (in fact only 3 people awarded BPR a rating of less than 3, and nobody awarded it a rating of less than 2), which is clearly desirable.

4.5 User Comments

The two strongest themes that arose from the optional user comments were:

1. One particularly bad recommendation is off-putting, even amongst a generally good list of recommendations. Several comments suggesting that one extremely bad recommendation "might put me off being recommended anything else in the list".

2. Too many mainstream programmes are not considered to be very good recommendations. Several negative comments related to recommendations that were "rather mainstream", with one person noting that "What I'd like from a recommendation tool is the unusual stuff that escapes me".

There were not enough user comments available to reliably say whether any one algorithm was performing particularly poorly in either of these areas; however these general observations are still useful for identifying the characteristics of a good quality set of recommendations. In respect of the first observation, it is thought that, in future trials, the user could also be asked for the number of recommendations (out of ten) that they deem to be completely inappropriate, as this could provide a further insight into the reasons behind the overall quality scores awarded.

5. CONCLUSIONS

The Matrix Factorization based collaborative filtering algorithm with the online update mechanism clearly performs best in terms of overall quality and variety, and in terms of the mean number of 'good' recommendations. It also shows a more favourable frequency distribution of scores, with a strong consensus that the algorithm is providing generally good recommendations, and with very few people awarding very low scores.

The metadata-based algorithm also produces good results in terms of mean scores, as well as the frequency distribution of scores, which gives strong evidence to suggest that most users think it is generally providing good recommendations. However, it performs less well in the perceived variety of the recommendations.

As expected, both the collaborative filtering and metadata-based algorithms clearly outperform the non-personalised recommender.

The field trial confirmed that the online update methods of the Matrix Factorization algorithm can provide good personalised recommendations and solve some of the practical difficulties encountered when implementing collaborative filtering within the constraints of an online TV service.

6. ACKNOWLEDGMENTS

Our thanks to Christoph Freudenthaler and Zeno Gantner at the University of Hildesheim, and Steffen Rendle at the University of Konstanz and for helpful advice and discussions.

7. REFERENCES

[1] Koren, Y., Bell, R. and Volinsky, C. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer*, 42, 8 (2009), 30-37. DOI= http://doi.ieeecomputersociety.org/10.1109/MC.2009.263

[2] Rendle, S. and Schmidt-Thieme, L. 2008. Online-Updating Regularized Kernel Matrix Factorization Models for Large-Scale Recommender Systems. In *Proceedings of the 2nd ACM Conference on Recommender Systems* (Lausanne, Switzerland, October 23-25, 2008). RecSys '08. ACM, New York, NY, 251-258. DOI= http://doi.acm.org/10.1145/1454008.1454047

[3] MyMediaLite Recommender System Library. http://www.ismll.uni-hildesheim.de/mymedialite/index.html

[4] Rendle, S., Freudenthaler C., Gantner Z. and Schmidt-Thieme, L. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence* (Montreal, Canada, June 18-21, 2009). UAI 2009. AUAI Press, Corvallis, Oregon, 452-461. http://uai.sis.pitt.edu/papers/09/p452-rendle.pdf

Tobias Fröhlich (founder of QTom TV)

Tobias Fröhlich is an expert in concepts and visions. In the field of format development and digital formats, he created the first interactive movie on German television, Germany's first mobile soap as well as convergent Web TV products for Top of the Pops and The Dome. In addition to classical publishing house TV magazine programs (Max-TV, Cinema-TV, Bravo-TV), he supervised the launch of VH-1. As an executive producer and chief editor, he produced successful TV shows for many years, among others for RTL, ProSieben, ZDF, MTV and RTL2.
His strengths: devising products that enthrall end customers and secure market advantages for companies.

About QTom

For the first time, QTom gives viewers the power to easily adapt the music program to their own individual taste via remote control. QTom is available on internet-capable televisions and set-top boxes as well as on the Web at www.qtom.tv. QTom is also the first provider of free music television on the iPhone and has been further developed for use on the iPad and iPod touch. The app has since been downloaded more than 250,000 times. QTom is currently only available via WLAN for these three Apple products. Current partners include Philips Net TV, LG Smart TV, Panasonic VIERA Connect TVs, Loewe TVs with MediaNet and Sharp TVs with AQUOS NET+. QTom also runs on TechniSat and eviado televisions. In November 2010, QTom was awarded the Deutschen IPTV-Award in the category "Best Creation, Design and User-Friendliness", in 2011 with a silver and a bronze Lovie Award (Entertainment, Integrated Mobile). QTom is financed through advertising and is free of charge to viewers. An ad-free premium model is in development.

EuroITV industry case

QTom: Interactive TV leads to customer satisfaction

Intro (live demo)
The rulers: completely new experience in watching music television

- Interaction: adaption to personal preferences, moods and taste
- Variety: numerous genre-, special- and theme-channels
- Despite the new: a really great music programming

Case

- Distribution: smart TV, tablet, mobile and the rest of online
- User appreciation: viewing time, rankings, reviews
- Reach: the growth path
- Money: invest, survive, invest, survive, invest again and harvest

MyScreens: How audio-visual content will be used across new and traditional media in the future – *and how Smart TV can address these trends.*

Submission for the industry track of EuroITV 2012 in Berlin

The Study

Today's media users can choose from a broad variety of technologies to access audio-visual content. New devices like tablets and smartphones and new services like video on demand and social media have changed media usage dramatically. SevenOne Media and Facit Digital have conducted a comprehensive study among German early media adopters to find out how linear and non-linear TV is combined with emergent services on the *second screen*.

Our talk will outline our most relevant findings to the industry, e.g.

- why some devices are more suitable in social situations than others
- what triggers second screen usage
- how various screens interact and how an overarching user experience can be achieved
- how *Smart TV* can address user needs in the future

How did we find this out? More than 100 Smart TV users conducted online diaries about their cross-media usage and took part in various moderated forum discussions. Some of them were invited to subsequent in-depth explorations of their daily media usage behaviour. **The results will be illustrated with new interactive products of the ProSiebenSat.1 Group that already account for our research insights.**

The Authors

Martin Krautsieder is Senior Research Manager New Media Research at SevenOne Media. He is responsible for the research about the usage of Video-content on different Screens and the effects on traditional TV, and all new Gadgets in the evolving TV business.

Michael Wörmann is managing director at Facit Digital and has been doing research on user experience success factors for more than 12 years. As a psychologist he is convinced that listening to the users' needs is the key factor to success for interactive digital media.

The Companies

As Germany's leading marketing company **SevenOne Media** provides audio-visual media for advertising companies and agencies. As a subsidiary of the ProSiebenSat.1 Group , we promote the German network's entire portfolio across all types of media (TV, online, mobile, games, teletext, pay-TV, video on demand) - on a national as well as international level. Besides traditional and special advertising formats, we also focus on networked communication in collaboration with our sister company SevenOne AdFactory.

Facit Digital is one of Germany's leading user experience research agencies with a strong focus on interactive TV and other emerging entertainment platforms like tablet PCs and smartphones. We're helping clients like maxdome, Sky, Vodafone, Kabel BW, ProSieben, or the Swiss public TV providing joyful user experiences through innovative research.

EUROITV 2012

Submission for iTV in Industry

Title

Nine hypotheses for a sustainable TV platform based on HbbTV presented on the base of a on a Best-of example

Author & Company

Markus Kugler, 33, is a dedicated user experience expert and certified usability engineer that has been working in the online and TV field for over a decade. Together with his team the founder and CEO of coeno realizes positive user experiences for TV- and entertainment solutions for customers such as maxdome, nacamar, Kabel BW and Vodafone. The agency coeno is specialized in the creation of high-class concepts and designs for usable and joyful user interfaces. Projects have been shortlisted for the German eco award 2010, the international ConnectedWorld.TV Awards 2011 and won an award from the Digital Communications Awards in 2011.

Content of Talk

Mark Kugler drafted nine hypotheses about HbbTV and calls for more commitment to the implementation of HbbTV and Smart TV applications. Frustrated consumers lose interest if market participants do not observe important basic rules.

The talk aims at explaining the relevance of a positive user experience in the entire value chain from hardware to applications, about the unique opportunity to create solid standards, about naming, clever advertising, social media and again and again about how to put consumers really into the center of efforts.

Using the example of the HbbTV application of KinoweltTV (powered by onteve) the hypotheses around social media will be filled with life.

KinoweltTV enables access to and interaction with both mass media channels on modern TV sets. Television and internet users can interact over an intuitive HbbTV-based user-interface using their TV remote control, rate and recommend program events, check what their friends have rated and recommended and browse through archived media-content.

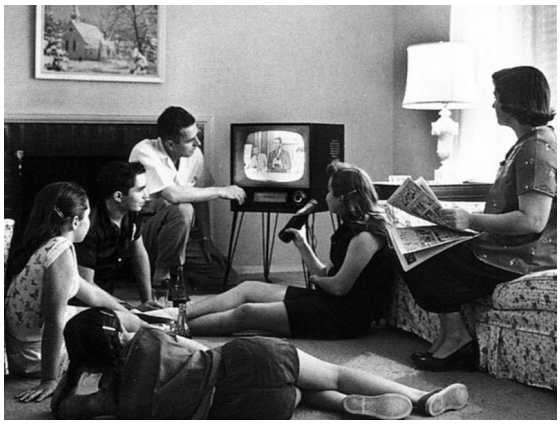
coeno GmbH & Co.KG
Westermühlstraße 23
D-80469 München

www.coeno.com
info@coeno.com
Fon: +49.89.85 63 898 11
Fax: +49.89.85 63 898 29

Geschäftsführer:
Bettina Streit & Markus Kugler
Amtsgericht München
hra 84295

Komplementärin:
coeno Management GmbH
Amtsgericht München
hrb 152952

Lean-back vs. lean-forward: Do television consumers really want to be increasingly in control?



Starting in the early 1930s, television has paved its way into the homes and living rooms of people throughout the world. Originally intended as a mean of information, it soon became THE medium for entertainment. Since then, television has been additionally praised for its relaxing effects (strong lean-back situation).

For many decades the typical usage scenario has been a very linear one: *"Come home from work – switch on television – zap through programs – find and watch something you like".* A second very common scenario is television as a ritual, like news at 8, followed by a movie at 8.15pm.

Although the reasons to watch television remained the same, the way people are watching seems to change. One elementary change is the introduction of secondary devices, like tablets, smart phones, laptops or stationary PCs into the consumption of television content. These are either used for multi-tasking while watching TV on the main screen (companion devices) or the content is entirely watched on one of the "second screens" (standalone devices).

At the same time, with the establishment of the web, the amount of available content is exploding, which makes search options and intelligent filters as well as personalization and recommendation features essential assets. More and more, consumers want television to follow their needs and routines and not the other way round, which can be described as an on-demand mentality. But are television users really willing to "lean forward" in order to reduce complexity and to customize everything there is? Do users want to be increasingly in control?

Sky is acknowledging these dynamics and while improving the "traditional" television is developing innovative products and features in order to cater the evolving needs of their customers. Our goal is to provide them with an "intelligent lean-back" or a "relaxing lean-forward" atmosphere when consuming television.

Two examples of our ambition into this direction are the products Sky Go and Sky Anytime, which will be broadly introduced and discussed in our session.

The author/presenter

As a cultural anthropologist, **Mirja Baechle-Gerstmann** is dedicated to understanding how people are affected by today's technology in their daily lives. For over 6 years she has consulted for clients on how to improve their web-based products and services. Over the years she has acquired strong skills in the media industry, especially in the area of future television, which she now uses in her new position at Sky Deutschland. Mirja holds an MA from Ludwig-Maximilians-University in Munich (Germany) and is member of the international Usability Professionals' Association (UPA).

The company

Sky Deutschland (sky.de) is a German media company with its core business in premium pay television. The company offers a wide range of programs including current films, new TV series and live sports, particularly the football Bundesliga, the DFB Cup and the UEFA Champions League. Sky has set standards with its HDTV offering that currently comprises more than 40 channels, including the first 3D channel in Germany and Austria.

NDS Surfaces



A presentation by

Simon Parnall
James Walker

© NDS Ltd 2011. All rights reserved.

Personalisation of Networked Video

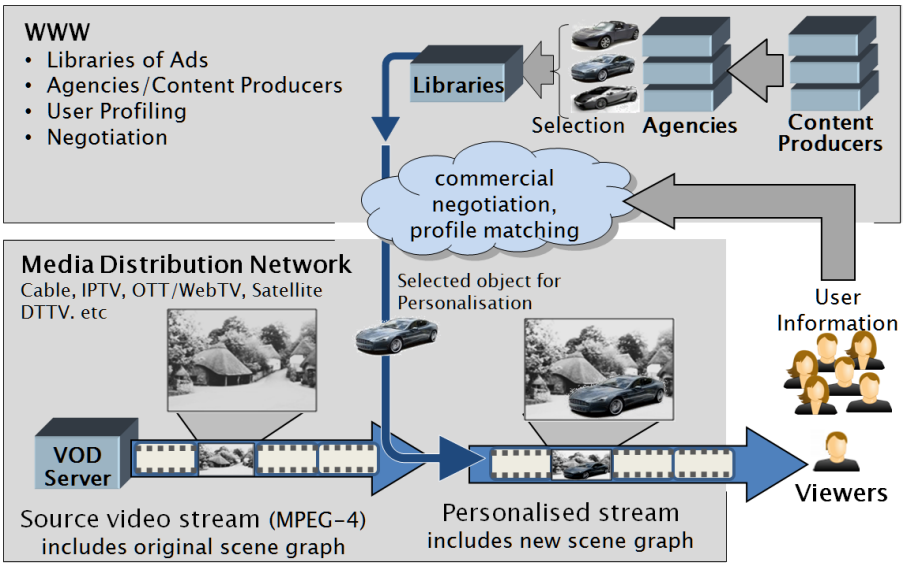
Jeremy Foss
Birmingham City University

Benedita Malheiro
Instituto Superior de Engenharia do Porto

Juan Carlos Burguillo
Universidad de Vigo

Personalised Product Placement, and more ... Following targeted advertising and product placement, TV and online media needs more personalised methods of engaging viewers by integrating advertising and informational messages into playout content, whether real-time broadcast or on-demand. Future advertising solutions need adaptivity to individuals or on-line groups to respond to the commercial requirements of clients and agencies.

We are currently developing an architecture for commercial personalised placement of externally acquired objects for insertion into source video content which may be specifically produced to accept future object placement. The solution is based on existing standards and supports all video distributions – cable, satellite, Telco IPTV, Web/OTT, DTTV.



Overview: The video platform is based on MPEG-4 object-based video utilising scene-graph representation of the video scene (other formats are being investigated). A standard video headend streams the source video to the user (STB, PC, pad, phone, *etc.*). Selected objects are served from third-party libraries and are sent via broadband to the user terminal where they are integrated with the source video to create the personalised program. Using this method the content editorial integrity remains under control of the production team, since they can stipulate (in the metadata) what ranges and characteristics of objects may be admitted for integration into the source video.

Objects are selected via agent-based brokerage which utilises the metadata (MPEG-7) of the source video, the metadata of available objects and the personalization data of the user to select the appropriate object to insert in to the source video. Negotiation procedures are brokered between agencies who aim to have clients' artefacts placed in the video. This can take the form of straight selection based on profile matching. In addition, agencies may compete in an auction. The process can theoretically take place in near-real-time.

Other Applications: The process can be applied to various personalization services including training, group-participation in broadcast video playouts, integration of gaming into video playouts. The personalization/brokerage process can also derive content to be streamed to additional user devices (i.e. multi-screening selected apps and content synchronised to the playout), integration of personal/group social network activities into video broadcasts.

Interested in collaboration in research or implementation in this area? Please contact us (emails below)

Jeremy Foss is a lecturer in the School of Digital Media Technology at Birmingham City University which offers BSc and MSc courses in Web Technology, TV and Film Technology. Jeremy has 30 years telecommunications industry research and development experience with interests in virtual worlds, agent technology and triple play networks. He is a Chartered Engineer and a Member of the Institute of Engineering and Technology. *email:* jeremy.foss@bcu.ac.uk

Benedita Malheiro is an Adjunct Professor in the Faculty of Engineering at the Instituto Superior de Engenharia do Porto, which has wide expertise in distributed systems, networking and web applications. Benedita holds a PhD and an MSc in Electrical Engineering and Computers. Her research interests are in multi-agent systems, intelligent agents, distributed systems and knowledge representation. She is active in the area of context-awareness, navigation and location based services. *email:* mbm@isep.ipp.pt

Juan Carlos Burguillo is an associate professor in the Department of Telematic Engineering at Universidad de Vigo. The University provides courses in technologies, social sciences and humanities. Juan Carlos holds an MSc in Telecommunication Engineering, and a PhD in Telematics. His research interests include intelligent agents and multi-agent systems, evolutionary algorithms, game theory and telematic services. *email:* J.C.Burguillo@det.uvigo.es

MARKET INSIGHTS FROM UK

WHAT ARE THE OPPORTUNITIES?

The UK is traditionally an indoors culture. Unlike warmer climates with opportunities for outdoor activities in the evening, life in the UK predominantly revolves around the living room and the television. Television is a large part of people's lives and central to many day-to-day conversations. As a soccer-mad nation, sports and reality shows are also big business regularly pulling in audiences of several million.

Competition between terrestrial broadcasters such as the BBC, ITV and Channel 4, satellite broadcaster the British Sky Broadcasting Group and cable provider Virgin Media, has ensured rapid research and development to identify new ways to engage audiences with programming content.

Amberlight has the privilege to work on exciting projects with UK broadcasters, content owners, Smart TV platforms and hardware manufacturers. Based on this research, we have identified a number of important factors that influence the way consumers engage with content. These include:

- The type of device used (screen size, type of interface) and whether a single screen is being used, a second screen or even three screens
- Context of use: in home and out of home; solo or group; and live, recorded or on-demand
- Genre of programming (drama, soap operas, reality shows, sport, news, films...)

We also see opportunities that broadcasters and advertisers are overlooking. There is currently a lot of talk about "social" TV especially with the launch of Zeebox at the end of 2011, but this is only a small part of the picture and there are massive opportunities to engage consumers with content in ways beyond just "social".



As well as the ability to discuss content, tell people what you're watching and what you think, read messages and tweets, there are opportunities to provide supporting content that is of sufficient quality, relevant and timely as well as product placement and 2nd screen point-of-sale (POS) opportunities.

It is also currently easier for users to consume additional or supporting content via 2nd or 3rd screens. In other words, Smart/Connected TVs don't make it easy for consumers due to limited input methods and "clunky" user interfaces, and consumers are turning to their mobile phones, tablets and laptops. For example, although there are half as many iPads as Connected TVs, there is four times more BBC iPlayer traffic via the iPad.

The way we watch television is changing in the UK, but this is due to developments in 2nd screen technology and proprietary hardware rather than the television set itself. It seems Smart TV still has some way to go. Amberlight will present and discuss its research and UK market insights and provide an opportunity for questions.

ABOUT THE AUTHOR

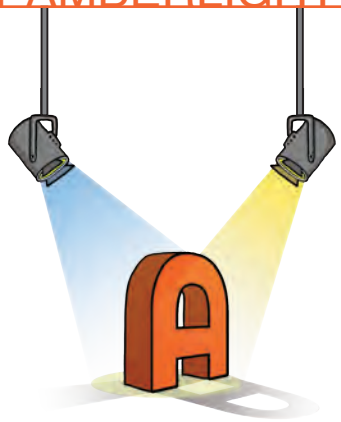


Sanjeet Bhachu, MPhil, BSc (Hons) Psychology & Computer Science, Principal Consultant

Sanjeet is one of Amberlight's most experienced consultants with a strong background in customer experience across digital platforms. He joined Amberlight in 2006, is responsible for a number of our most important clients and has created high-quality user experiences for high profile products.

Prior to Amberlight, he used his entrepreneurial instincts to set up a Web development consultancy in Scotland, working with various businesses to establish their web presence. His interests include digital and interactive TV, mobile technology and the connected home.

ABOUT AMBERLIGHT



Amberlight Partners Ltd (www.amber-light.co.uk) helps organisations design and develop better digital products and services and maximise the ROI from them. We help organisations understand what their customers experience when they come across their brands in digital channels. Do they find products usable? Do they find services useful? Do they find communications persuasive?

Amberlight delivers insight and advice at every stage of the product development lifecycle:

- Before products have been developed, we help our clients identify what they should be developing
- During product development, we help organisations understand the content and functionality required
- After products have been launched, we advise on how their performance can be improved

Amberlight

Tutorials

Tutorial I: Foundations of interactive multimedia content consumption

Among the events organized to celebrate the 10th edition of the EuroITV conference, this special tutorial will cover the foundations of interactive multimedia content consumption. Some of the most active members of the community will join forces to present fundamental aspects associated with the research topics, which are of interest to the EuroITV community.

The following topics will be covered:

- Standards: IPTV, Hybrid TV & WebTV (Stefan Pham)
- Interactive TV advertising models (George Lekakos)
- Ambient Media (Artur Lugmayr)
- Cross plattform applications and multiscreen scenarios (André Paul)
- Personalization & Recommendation (Paolo Cremonesi)
- Navigation and information architecture (Lyn Pemberton)
- UX/HCI methods (Regina Bernhaupt)
- Social TV: past, present and future (David Geerts & Pablo Cesar)

Designing Gestural Interfaces for Future Home Entertainment Environments

Radu-Daniel Vatavu
University Stefan cel Mare of Suceava
13, Universitatii
720229 Suceava
vatavu@eed.usv.ro

ABSTRACT

Home entertainment systems have known incredible technological advances in the last years by incorporating developments from sensing, processing, communications, and visualization technologies. As such systems become more and more complex, presenting more and more options to their viewers, careful design of interacting devices and interaction techniques are needed. The tutorial will explore the use of gestures for interacting within home entertainment scenarios by discussing gesture acquisition and recognition techniques and focusing on designing for the interactive TV. Case studies will be discussed and a practical design session on reusing the Wii Remote for interactive TV scenarios will be included.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: Multimedia Information Systems, User Interfaces; I.5 [**Pattern Recognition**]: Applications, Implementation.

General Terms

Algorithms, Design, Human Factors.

Keywords

Gesture-based interfaces, home entertainment, interactive TV, gesture recognition, human-computer interaction, novel interfaces, Wii, Kinect

1. BRIEF DESCRIPTION

Home entertainment environments, when properly designed, possess enormous potential for creating multisensory experiences engaging their users into complex enriching activities, all taking place in the comfort and familiarity of the home environment [1]. The new standards of such entertainment scenarios that mix the digital space with the real one can only be achieved when design architectures work together with engineers [2, 3]. In this vision of future entertainment experiences, the existing controlling devices in the form of standard remotes can't find their place any more. As such standard input devices can't handle the ever increasing number of functionalities without ruining and diminishing user experience, new interaction paradigms must be explored. Among these, gesture-based interfaces represent a viable candidate. This tutorial explores the use of gesture commands for interacting with home entertainment environments by focusing on practical design case studies.

2. DETAILED OUTLINE

Title: Designing Gestural Interfaces for Future Home Entertainment Environments

Instructor: Radu-Daniel Vatavu, http://www.eed.usv.ro/~vatavu

Presentation format: power point slides, video demonstrations, and interactive content (using the Wii Remote)

Tutorial level: between introductory and intermediate.

Length: half-day

Contents:

1. **An introduction to gesture-based interfaces: Why gestures?**

   An introduction will be provided on the existing research on gesture-based interaction by highlighting benefits but also pitfalls of using gesture commands in interfaces.

2. **Gesture acquisition: How to choose the right technology?**

   The main technologies used to acquire human body motion and gestures will be presented: worn and hand-held equipments (motion sensing game controllers, gloves, mobile phones) and non-invasive solutions (video processing). Advantages and weakness of each approach will be discussed.

3. **Gesture recognition: Recognize this!**

   Two widely-used pattern recognition algorithms for recognizing gesture motions will be presented (nearest neighbor classification using the dynamic time warping and Euclidean distance metrics). This section also introduces terminology from the pattern recognition domain such as classifiers, training and testing sets, within- and between-class variation, and performance measurements.

4. **Applications of gesture-based interfaces to home entertainment scenarios: Designs and misdesigns**

   Practical applications and case studies of how gestures have been integrated into interactive TV and home entertainment scenarios will be discussed.

5. **Practical work: Re-using the Wii Remote for interactive TV scenarios**

This will be an interactive session in which participants are going to be introduced to programming a low-cost widely-available motion sensing device (the Wii Remote). The device will be re-used in order to create a simple gesture-based interface for TV control. After a short introduction on how to program the device, attendees will be involved into a participatory design of a gesture-controlled TV interface.

6. **Conclusions**

   The challenges of the field as well as the vision of future gesture-based interfaces will be discussed.

Potential target audience: Researchers and practitioners working in human-computer interaction, user interfaces, and design of home entertainment experiences that want to include gestures into their interfaces. No previous knowledge of gesture-based interaction is required. A mixture of researchers from various fields will be ideal. Participants are going to be introduced into gesture acquisition (how to select the right technology?), recognition (how to select and implement the right recognition algorithm?), and design studies (what are the guidelines when including gestures in interfaces?). A practical programming and design case study is included in the tutorial.

3. PRESENTER RESUME

Research interest

Radu-Daniel Vatavu works in Human-Computer Interaction and focuses on novel interactions and gesture-based interfaces. Radu-Daniel Vatavu has developed gesture recognition algorithms [4,5], designed computer vision systems [7,9], and investigated human performance aspects of gesture production [6]. He is also interested in designing novel interfaces for home entertainment scenarios such as interactive TV systems, ambient media, and computer games, with a specific concern towards the human aspects of the interaction and how technology helps to enhance everyday Human-Human Interactions.

Research interest specific to EuroITV

Dr. Radu-Daniel Vatavu has been investigating the opportunity of using gesture-based interfaces for interactive TV scenarios. For example, the interactive coffee table prototype [9] represents a computer vision system that allows viewers to control their TV set using simple hand gestures. The coffee table prototype brought many advantages over existing research such as dealing with fatigue problems when using gesture commands and alleviating privacy issues for video-based home entertainment scenarios. Design guidelines for novel interfaces and future TV environments are discussed in [10]. Home Point & Click [7] represents a system for controlling a large video projection display in the home environment using the Wii Remote controller. The project introduces the concept of multiple media screens that are reconfigurable and can be customized by their viewers/owners. Standard WIMP interaction techniques are mixed with gestures for controlling such media screens.

Teaching experience

Dr. Radu-Daniel Vatavu is a Lecturer at the Computer Science department of the University Stefan cel Mare of Suceava where he teaches Algorithms Design, Pattern Recognition, Advanced Programming Concepts, and Advanced Artificial Intelligence Systems for undergraduate and master students.

4. ACKNOWLEDGMENTS

This research was supported by the project "Progress and development through post-doctoral research and innovation in engineering and applied sciences- PRiDE - Contract no. POSDRU/89/1.5/S/57083", project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

5. REFERENCES

[1] Baillie, L., Benyon, D. 2008. Place and Technology in the Home. Computer Supported Cooperative Work, 17, 2-3 (April 2008), 227-256, doi:10.1007/s10606-007-9063-2

[2] Canizares, A. G. (2006) Fun Rooms: Home Theaters, Music Studios, Game Rooms, and More. Harper Design, 2006

[3] Rushing, K. (2006) Home Theater Design: Planning and Decorating Media-Savvy Interiors. Quarry Books, 2006

[4] Radu-Daniel Vatavu, Laurent Grisoni, and Stefan-Gheorghe Pentiuc. 2009. Gesture Recognition Based on Elastic Deformation Energies. In Gesture-Based Human-Computer Interaction and Simulation. Lecture Notes In Artificial Intelligence, Vol. 5085. Springer Berlin / Heidelberg, 1-12

[5] Radu-Daniel Vatavu, Laurent Grisoni, Stefan-Gheorghe Pentiuc (2010). Multiscale Detection of Gesture Patterns in Continuous Motion Trajectories. In S. Kopp, I. Wachsmuth (Eds.) Gesture in Embodied Communication and Human-Computer Interaction. LNCS 5934, Springer Berlin / Heidelberg, 85-97

[6] Radu-Daniel Vatavu, Daniel Vogel, Géry Casiez, and Laurent Grisoni. 2011. Estimating the perceived difficulty of pen gestures. In Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part II (INTERACT'11). Springer Berlin, 89-106

[7] Radu-Daniel Vatavu. 2011. Point & click mediated interactions for large home entertainment displays. Multimedia Tools and Applications, Springer Netherlands, http://dx.doi.org/10.1007/s11042-010-0698-5

[8] Radu-Daniel Vatavu. 2011. Reusable Gestures for Interacting with Ambient Displays in Unfamiliar Environments. The 2nd International Symposium on Ambient Intelligence (ISAmI), Salamanca, April 2011. Advances in Intelligent and Soft Computing, vol. 92, Springer Berlin / Heidelberg, 157-164

[9] Radu-Daniel Vatavu, Stefan-Gheorghe Pentiuc. 2008. Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience. EuroITV 2008: the 6th European Interactive TV Conference, Salzburg, July 2008, LNCS 5066, Springer Berlin / Heidelberg, 183-187

[10] Radu-Daniel Vatavu. 2010. Creativity in Interactive TV: Personalize, Share, and Invent Interfaces. In A. Marcus, A. Cereijo Roibas, R. Sala (Eds.) Mobile TV: Customizing Content and Experience. Springer Human-Computer Interaction Series, Springer London, 121-139

Workshops

Workshop 1: Third International Workshop on Future Television: Making Television more integrated and interactive

ImTV: Towards an Immersive TV experience

João Magalhães[1], Sharon Strover[2], Teresa Chambel[3], Paula Viana[4], Teresa Andrade[5], Luis Francisco-Revilla[6], Flávio Martins[1], Nuno Correia[1]

[1] CITI / Dep. Informática, Fac. Ciências e Tecnologia, Uni. Nova Lisboa, Portugal

[2] School of Communication, University of Texas, Austin, US

[3] LaSIGE, Faculdade de Ciências, Univ. Lisboa, Portugal

[4] Instituto Superior de Engenharia do Porto and INESC TEC, Portugal

[5] Universidade do Porto, Faculdade de Engenharia and INESC TEC, Portugal

[6] School of Information, University of Texas, Austin, US

jm.magalhaes@fct.unl.pt, sharon.strover@austin.utexas.edu, tc@di.fc.ul.pt, pviana@inescporto.pt, mandrade@inescporto.pt, revilla@ischool.utexas.edu, flaviomartins@acm.org, nmc@fct.unl.pt

Abstract. The media marketplace has witnessed an increase in the amount and types of viewing devices available to consumers. Moreover, a lot of these are portable, and offer tremendous personalization opportunities. Technology, distribution, reception and content developments all influence new "television" viewing/using habits. In this paper, we report results and findings of a transnational three year research project on the Future of TV. Our main contributions are organized into three main dimensions: (1) a user survey concerning behaviors associated with media engagement; (2) technologies driving the social and personalized TV of the 21st century, e.g. crowdsourcing and recommendation systems; and (3) technologies enabling interactions and visualizations that are more natural, e.g. gestures and 360º video.

1 Introduction

Current media entertainment and traditional television are now being delivered through different communication vehicles (e.g. Wi-Fi, 3G, DVB, and cable) and are accessible through multiple devices. The Web has also enabled viewers to interact with content through revolutionary applications, and content itself is no longer exclusively delivered by broadcast or cable channels. Given the complexities of this new media environment, researchers have begun to grapple with the dynamics of contemporary media usage.

The ImTV project[1] is set in the middle of this media revolution. To realize the full potential of these numerous changes, we envision a future where TV delivers an immersive entertainment experience. Viewers will be socially immersed in the media chain by producing media and related content (i.e., crowdsourcing moves to TV), increasing their sense of belonging and enjoyment. Technologies will enable an immersive viewing experience for an increased sense of presence and engagement.

[1] http://imtv.di.fct.unl.pt/

ImTV is a socio-technological research project involving university and media partners from the US and Portugal. The new interactions afforded by current technologies are helping to change the role of TV from a traditional "spectatorial culture" to a "participatory culture", where users are not only consumers but also producers of information.

This paper summarizes the project's most relevant contributions.

Future viewers are very different from the "TV zapping or VCR" generation. A US/Portugal user survey with over 1,000 participating subjects (Section 2) provides a picture of the media consumption habits and trends of the future media market.

Viewers seek greater social interaction in the consumption of entertainment media. They comment, tag, rate, chat and create new content about other media. Technologies to better capture and explore this social feedback and interaction will be discussed in Section 3. Hardware technologies and TV sets with rich interactive features have changed traditional TV. From context-aware TVs to 360º experiences, our work concerning future interaction and viewing experiences will be discussed in Section 4.

In the past, the TV industry has been driven by technical standards, but the current trend disrupts the monolithic and incumbent industry bias that standards-setting processes favor. How will traditional media players cope with this fast changing scenario? Section 5 tries to answer this question.

2 Getting to know viewers: A user survey

Our study decodes the behaviors associated with media engagement among university student populations in two distinct locations: a large research-based university in the southern United States and three similar institutions in Portugal. Young people are in the vanguard of new ways of engaging entertainment, and we believe their responses are good e-predictors of patterns that will be present in the rest of the population eventually. Our intent is first, to describe some of the new "viewing" behaviors among this group of users/participants, and second, to understand the influence of industrial settings (e.g., services available), cost, and platform on patterns of engagement. The broader goal of this study is to unpack how interactivity operates at this moment in time, and to broaden discussions around the idea of platform and choice. The background of media industries and their ability to structure choices is highlighted by the opportunity to compare two different environments, those of the US and Portugal, where viewing/using opportunities vary.

Engagement: Devices

In response to our concern regarding how viewers/users engage with "television" or visual entertainment, we found strong evidence of a shift from the usage of television content on standard audiovisual devices such as the television and DVD player to portable multi-purpose platforms, especially the laptop computer. The student population by and large views media on their laptop devices, but also opts for television secondarily as a ready-made system for display. The Portugal sample is statistically significantly more likely to use a "normal" television as a display device

than is the U.S. sample (75% versus 47% report they frequently or very frequently use televisions as a display device). Approximately 71% of the overall sample indicated using a laptop frequently or very frequently as their primary medium for watching entertainment fare, with an average 59.2% giving comparable responses for "television" (multiple responses were accepted on this item since we assume people use multiple devices). Surprisingly, very low percentages reported using desktop computers or tablet computers for entertainment viewing; about 11-12% of U.S. and Portuguese students reported using mobile phones for this purpose.

Engagement: Content sources

In terms of what constitutes the source of visual material, cable television material is still highly popular, although more so among Portuguese students than U.S. students (78% versus 63.5%), while web-based films (such as available from Amazon or Netflix) were most popular for American students (78% use these services regularly). The most striking statistic, however, comes from a query regarding how frequently people use certain content sources: There, while "television" (cable or broadcast) is certainly cited (24.1% weekly and 48.5% daily use), downloaded or streaming fare is also very high (33.9% weekly and 32.3% daily use). Web-served content is obviously surging.

The young people in this sample create their own content to share with others. In response to "What services do you use to share your creative work," students focused largely on Facebook (87.6%), YouTube (57.9%) and email (64.9%). Other options such as Vimeo, Flickr, Photobucket, Picasa, personal blog/site, Google Docs, Twitter, MySpace, and Hi5 were mentioned by some, but in numbers far behind the top three.

Among U.S. students, who had better Internet access and also more individualized viewing settings, Facebook use for responding to others' content dwarfs all other services: 95% of the US sample report using Facebook to share or comment on content created by others, and over half the same (66%) do so several times a day. In other words, this secondary circulation of commentary is extremely common.

Mobility

Are people taking advantage of the mobility inherent in laptops and mobile phones for entertainment purposes? When we cross-tabulate use of the mobile phone with use of Facebook, YouTube and email - which are primary platforms used for entertainment for the sample - we see mixed results. First, the use of Facebook is so dominant that the device platform does not appear to matter: people are on Facebook a lot, wherever they happen to be and whatever technology is in their hands. YouTube is a somewhat different story. The people creating and sharing content via this video platform rely more heavily on mobile phones; their frequent use of the mobile is associated with YouTube content creation. The 100 or so people who report being heavily dependent on the mobile phone also frequently use Google Docs, Picasa, MySpace, and Hi5 to share content.

Facebook is central

For U.S. students, Facebook was cited by 95% of the sample as a service used to share or comment on content created by others, with U.S. students somewhat more intensive users than Portuguese students. We find 81.3% of the latter used it as a

sharing/commenting service. In terms of where people shared their own work, Facebook again was the primary venue. About 94% of the U.S. and 84% of the Portuguese students indicated it was their primary location for presenting work. Given this, it may not be surprising that in the combined sample, 52% reported that they use Facebook multiple times per day. Everyone uses it and nearly everywhere as well. Facebook is the dominant medium for sharing and commenting on others' content and for sharing one's own content as well.

Usability factors

When assessing comparative qualities of various content delivery systems, we asked about issues such as cost, ease of use, the graphical interface, the physical interface (if applicable), content availability, image and audio quality, and recommendation systems (if any). On these dimensions, Facebook, YouTube, and email received very high ratings compared to many other competing services, particularly in terms of ease of use. For instance, the entire sample agreed email was extremely easy to use (an "excellent" rating), compared to the 32% of the sample who said the same about gaming consoles, or the 27% who rated "television" as excellent. Facebook does not compare as well in terms of content availability, however, compared to streaming media or YouTube, although it receives a rating that is higher than "conventional" television.

Access and Downloading

Authorized downloading services like Netflix (27.2%) and Hulu (13.8%) are ranked highly by U.S. respondents as their most frequent method of accessing media, with nearly 40% of the U.S. surveyed population using one of these two services. Cable TV is the preferred medium of just 16.2% of the U.S. respondents. Only 6% of U.S. respondents indicate that other streaming services serve as their primary media choice.

In contrast, the Portugal media system does not feature access to Netflix or Hulu, nor do people have access to a popular authorized streaming service at the time of the survey. Instead, Portugal respondents turn to cable TV at a rate of 45.5%, followed by 28.4% who choose streaming media for their primary source of entertainment.

3 The *New* Social TV

The viewer/user results indicate that the convergence of TV and the Web has empowered TV viewers and made them active contributors towards a more inclusive TV. Not only do users chat online about TV shows, but they also rate them, tag, comment, or create new content through Wikis, YouTube, Facebook or other fan-dedicated sites. We are witnessing the emergence of crowdsourcing via social networking platforms, thus enabling new mash-up applications merging information from the crowds.

3.1 Augmented TV and Web in crossmedia environments

In the ImTV project, finding relevant and possibly unexpected video-based content created by traditional producers or the users can be of great value. User generated data (ratings, tags and content) can be created while media is being played. We have explored cross-media environments, ranging from TV to PCs, and mobile devices in different contexts [Prata 11a] [Prata 11b], where users access additional and related content or add data in cross-media environments in real-time. Our prototype system allows the creation, access and sharing of flexible web personalized informal learning environments via different devices, as additional information to the video being watched, and as an answer to the informal learning opportunities created by video and iTV. When watching a TV program or a video, viewers can select topics that are presented and catch their attention in the video content (e.g. what the characters are talking about), or they can select information about the program (e.g. actors, director, shooting locations). They can get a short piece of information right away or alternatively, a more detailed webpage, created with information about all the items selected during the viewing, later on. Users may then access, add to, and share these contents with other users, anytime and anywhere. For instance, they can add a picture captured from their mobile device at the same location of a specific scene of a movie they have watched, and comment about this related information.

User evaluation in [Prata 11a] and [Prata 11b] provided encouraging results in terms of usability and user acceptance, through indicators such as utility, ease of use and user satisfaction, in contexts of use where the main affordances of the different devices can be more adequate, in isolation or combined. Content classification and metadata, as well as human-centered design in HCI are central aspects in this work.

3.2 Listening to viewers: Sensing the audience mood

We set out to develop a Social-TV system prototype that allows users to chat with each other, in an integrated way. Our prototype system analyses chat-messages to detect the mood of viewers towards a given show (i.e., positive vs. negative). This data is plotted on the screen to inform the viewer about the show's popularity. Although the system provides a one-user / two-screen interaction approach, the chat privacy is assured by discriminating information sent to the shared screen or the personal screen.

In our system [Martins 12], we infer the sentiment of the message towards the show, illustrated in Fig. 1 and Fig 2. In the proposed system, user messages are processed in real-time by a sentiment analysis algorithm, allowing the plot of the emotions felt by viewers. An advantage of the proposed visualization method is that viewers can get a quick glimpse of the show's popularity in the last few minutes.

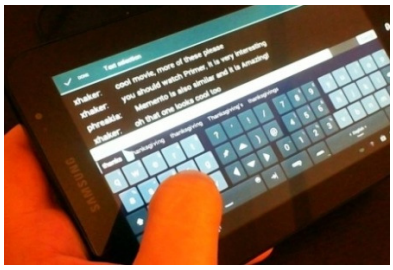


Fig. 1. Chat interaction in a tablet device.

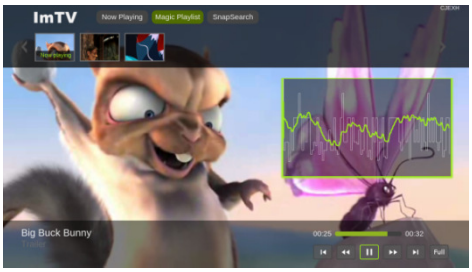


Fig.2. Chat sentiment analysis graph.

3.3 Mash-up-TV: Rating, tagging and commenting

As illustrated earlier, television is becoming more connected, more social and linked. Valuable information is available in external applications and services that make up a cloud of information around TV providers and users. Comments on social networks can be used to infer the mood of a user or the number of followers of a TV show [Martins 12]. IMDb, TV.com and other related sites can provide extra metadata to characterize content and enable better recommendations [Peleja 12]. A Mash-up-TV solution that is able to use the cloud of information to create new content is the final, integrated stage of our system. This amount of data can be provided to recommendation engines to enhance the viewing choice decision process. Users' feedback such as ratings, tags and comments, are scattered across many different Web applications.

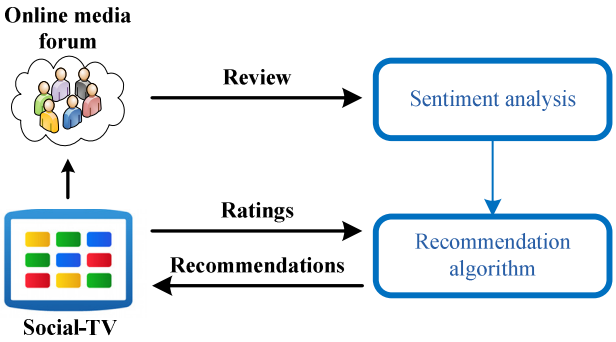


Fig. 3. Recommendations based on ratings and reviews.

Besides exploiting traditional approaches to Recommender Systems, ImTV project also explored recommendations to groups [Dias 11], diversification of recommendations [Dias 12], comments-based recommendations [Peleja 12] and evaluating the quality of recommendations [Soares 12]. In [Peleja 12] we developed a system, Figure 3, that uses machine learning techniques to analyze user reviews and infer their preferences from these comments/reviews. A recommendation algorithm is then responsible for exploring these comments to compute collaborative recommendations.

4 Immersive viewing and interaction technologies

As display devices such as "televisions" start to come fitted with multi-core processors and camera devices, their processing power opens the way for new interaction technologies. Gestures and face recognition, and tablet-based control are already in the market. We have researched other engagement mechanisms including 360º video, awareness interaction and multiple-screen entertainment, towards more immersive interactive experiences.

4.1 Multiple screens: How to merge them?

Designing an integrated user experience for watching TV across multiple screens requires considering that the number of participants, screens, and locations can vary and change. For instance, users can start watching a football match at home alone using their laptop, then continue watching in their mobile phones in the subway, and finally meeting their friends at a sports bar where they all watch multiple large shared screens in combination with their personal devices. In this scenario, the characteristics that define a "session" are fluid. Separating sessions by screen location or number of participants can break the flow of the activity.

In order to manage different combinations of screens, contexts and participants, it is necessary to abstract the concept of "session" and make it available to different devices as necessary. As people might engage with different screens in different ways, each device needs to be aware of the state of the multiple activities in the overall session (or sessions). At the same time, screen changes are important as they often signify important transitions in terms of context and participants, and the session needs to be aware of the affordances, constraints and context of the involved devices. In [Martins 12] we developed an entertainment system spanning multiple screens for the same session. All users share one main screen and each user can have a private screen (to browse extra info, communicate with others, and control the main screen).

Certain devices are better suited for interaction paradigms such as push-content (e.g., traditional TVs work well with channels and pre-scheduled programming). Other devices are better suited for pull-content paradigms (e.g., online video providers support interactions based on recommender systems as the ones developed in [Peleja 12]).

Interacting with TV goes beyond engaging with a particular program or video clip. It involves activities such as browsing, searching, sharing, rating, and commenting. In circumstances where multiple screens are available, it is possible to match the activities to the screens. For instance, consider modern smart TVs that allow using tablets as smart remote controls such that viewers can see the program schedule and select channels using a separate screen, less intrusive of the TV viewing experience.

4.2 Immersive viewing and interactivity

Highly immersive cinema and video have stronger impact on viewers' emotions

[Visch 10] and their sense of presence and engagement. 360º video has the potential to create highly immersive video environments since the user is no longer locked to the angle used during the capture of the video, and 360º video capturing devices are becoming more common and affordable to the public. Hypervideo[2] stretches boundaries even further, allowing the viewer to interact with the video, to explore and navigate it, and to get related information through links defined in space and time, towards stronger interactive and engaging experiences [Douglas 00].

As a first approach, we designed and developed an interactive interface for the visualization and navigation of 360º hypervideos based on the web. The interface allows users to pan around to view the interactive video content from different angles and to effectively access related information through the hyperlinks. One challenge for presenting this type of hypervideo includes providing users with an appropriate interface capable of exploring 360º contents, where the video should change perspective so that the users actually get the feeling of looking around. Another challenge is to provide the appropriate affordances to understand the hypervideo structure and to navigate it effectively in a 360º hypervideo space. The main navigation mechanisms include drag interface (to pan around the 360º video), view area and minimap (for 360º orientation), link awareness cues (to spot the links even when out of sight), image maps (to summarize and index the video), and a memory bar (to know where the user has been). User evaluation [Neng 11] has provided very encouraging results in terms of usability, orientation and cognitive load, which are related with the main challenges. Perceived usefulness, satisfaction, and ease of use were also highly rated in the evaluation [Neng 11].



Fig. 4. Navigation in the 360º Hypervideo Player. Left: details about points of interest on the left, image map below, with link to another part in the video; Right: View with minimap below, link hotspots on the video.

As a next step, we are designing and developing immersive interactive videos as extensions of the first 360º hypervideo. Moving out of the box, by stretching or breaking the window boundaries, allows for a higher level of immersion, potentially full 360º [Neng 12]. We are exploring increasing levels of immersion, from settings closer to the current TV viewing environments, towards more immersive ones. In terms of output: 1) accessing HV 360º in full screen and in larger screens; 2) extending the user workspace across more than one screen side by side towards a

[2] Hypervideo is video containing clickable objects linking to extra content.

circular alignment; 3) using video projections in large walls; and later on 4) developing video projection in an immersive Cave environments (e.g. the Portuguese science center: www.lousal.cienciaviva.pt), with varying angles of projection, possibly towards full immersion, in cylinder or spherical (e.g. allosphere.ucsb.edu) environments will be pursued. In any case, HD videos and surround sound can increase the sense of immersion. In terms of input, moving away from the desktop towards more natural interactions may also increase the feeling of immersion. As a first step, we are using a cordless mouse with gyroscope that works in the air (Logitech Max Air), mainly in scenarios 1) and 3), with promising results that allow for unrestricted and flexible interactions with a sense of control and increased immersion. Other modalities might be explored, including gestures, or gloves, especially in 3D videos.

5 Standards: How to keep-up the pace?

Metadata and video formats pose major challenges to service providers. In what concerns 3D metadata, major difficulties arise from the lack of a unique approach for video coding, which has a direct impact on metadata. Moreover, although 3D has been around for quite a long time (on video games and computer graphics), it is still a child in what concerns immersive cinema and television and de-facto standards are expected to have a strong impact. The question is "who will survive?" or "is there a killer solution?"

In 3D immersive environments, multiple views of the same scene are often captured using multiple cameras, to enable wide field-of-view (FOV) and 3D reconstruction, requiring considerable resources for storing and transmission. The main approaches for 3D video coding include conventional stereo, video plus depth, multi-view or combinations of each of the basic techniques [Ozbek07] [Schreer10]. Most promising formats are MVC (Multi View Coding), 3DVC (3D Video Coding) and FVV (Free Viewpoint Video). The diversity of formats greatly jeopardizes interoperability, and content adaptation techniques may assume vital importance. In multi-view 3D video, visual and audio attention modeling approaches can be considered interesting alternatives for designing scalable and adaptable systems. The viewing angle and/or audio may provide important information driving visual attention, and hence should be considered for adaptation purposes. Approaches based on color texture, or the multi-view depth information can also be considered. Adaptation can also be employed to increase the user quality of experience and to establish interoperability across different compression formats and ensuring quality presentations in diverse client terminals [Kim 03].

New aspects introduced by 3D have a direct impact on metadata that enables describing, using and adapting content efficiently. Re-usability of metadata and consequently of the content itself, strongly depends on the availability of standardized schemes. In the recent past, different initiatives have produced rich standards, usually covering different areas of application. For instance, the MPEG7 standard offers tools to describe both high-level concepts, as well as intrinsic characteristics of content trying to address the multimedia community at large. Its complexity compromised its

acceptance within the broadcasting industry. Organizations like SMPTE, TV-Anytime and EBU, proposed their own solutions, and no solution can be singled out as the winner at this time. Mapping and gateways are often the solution to integrate different applications or different areas of the workflow.

6 Pruning: Which services will survive?

ImTV research suggests that the audience is evolving into one that expects to create and use content in various forms and in various places. While students may not typify the adult audience since they typically have fewer economic resources and possibly limited time to spend with media, they are early adopters of technology, and they may set the agenda for how other generations engage with entertainment programming of various sorts. We found evidence that user-generated content is popular, and the two to three hours per day of use, common among our sample, is spent in a "lean forward" fashion, rather than a "lean back" fashion: leaning forward to create and interact, rather than leaning back to just consume.

As the interplay of media interaction mechanisms and multiple device configurations evolves, new services emerge. For instance, social-TV interaction applications, such as SentiTVchat [Martins 12], will be part of every entertainment service. Other excellent examples of the "lean forward" attitude are the fan-groups Web sites, where users leave long traces of their thoughts towards a given show. This rich user feedback is extremely valuable for many innovative applications [Peleja 12].

An increased number of screens available to users is another emerging trend. New user interfaces and interaction paradigms will rely on distributed architectures. New types of media will emerge, such as 360º video [Neng11], to provide users with immersive and realistic experiences. However, the most interesting aspect is that these screens are not just screens – the manufacturers of these multiple screens use it as an extension to reach end-users and offer them new content, services and applications. Thus, TV manufacturers (an old passive player of the TV industry) are now more active and suddenly closer to users than most TV broadcasters. They can actually drive viewers into their online Web entertainment services in a much smoother way.

Although standards are an essential step towards interoperability, the time required to produce a solution is not compatible with the urgency for new types of content and experiences that the new generation of users requires. Industry-imposed standards and new products coming directly from the universities are expected to have a great impact in the near future.

We are currently witnessing a massive change in one of the most successful media industries. TV manufacturers are pushing technology into the market at a faster pace than it can absorb: they all want to offer the "killer app", but it will be the social interaction among viewers that will decide the winner.

Acknowledgments. This work is financed by the Portuguese Foundation for Science and Technology within project FCT/UTA-Est/MAI/0010/ 2009 and PEst-OE/EEI/UI0527/2011, Centro de Informática e Tecnologias da Informação (CITI/ FCT/ UNL) - 2011/2012.

7 References

[Dias 11] P. Dias, “Recommending media content based on machine learning methods,” M.Sc. Thesis, Universidade Nova de Lisboa, 2011.

[Dias 12] P. Dias, J. Magalhães, “Greedy vertex-angle maximization for multi-user diverse recommendations,” Technical report, Universidade Nova de Lisboa, 2012, *submitted for publication*.

[Douglas 00] Y. Douglas, and A. Hargadon, “The Pleasure Principle: Immersion, Engagement, Flow,” ACM Hypertext’00, 153-160, 2000.

[Harboe 08] G. Harboe, C. J. Metcalf, F. Bentley, J. Tullio, N. Massey, G. Romano, “Ambient social tv: drawing people into a shared experience,” SIGCHI conference on Human factors in computing systems (CHI '08).

[Lane 11] N. D. Lane, Ye Xu , H. Lu , S. Hu, T. Choudhury and A. T. Campbell, "Enabling Large-scale Human Activity Inference on Smartphones using Community Similarity Networks (CSN)", Intl. Conf Ubicomp, Sept. 2011.

[Martins 12] F. Martins, F. Peleja, J. Magalhães, “SentiTVchat: Sensing the mood of Social-TV viewers”, ”, EuroiTV’ 2012, 10th European Interactive TV Conference, Berlin, Germany, July 4-6, 2012.

[Narasimhan 09] N. Narasimhan, T. Horozov, J. Wodka, J. Wickramasuriya, and V. Vasudevan. 2009. “TV clips: using social bookmarking for content discovery in a fragmented TV ecosystem,” Intl. Con. on Mobile and Ubiquitous Multimedia (MUM '09).

[Neng 12] L. Neng, T. Chambel, “Get Around 360º Hypervideo: It’s Design and Evaluation,” in "Ambient and Social Media Business and Applications”, Special issue of IJACI, International Journal of Ambient Computing and Intelligence 2012.

[Otebolaku 11] A. Otebolaku, M. T. Andrade, “Context Representation for Context-Aware Mobile Multimedia Content Recommendation”. In proceedings of IMSA 2011 – The 15th IASTED International Conference on Internet and Multimedia Systems and Applications, May 2011, Washington DC, USA.

[Peleja 12] F. Peleja, P. Dias, F. Martins, J. Magalhães, “A recommender system for the TV on the Web: the power of user reviews”, Technical report, Universidade Nova de Lisboa, *submitted for publication*, 2012.

[Prata 11a] A. Prata, and T. Chambel, “Going Beyond iTV: Designing Flexible Video-Based Crossmedia Interactive Services as Informal Learning Contexts”. EuroiTV’ 2011, Lisbon, Portugal, June 29-July 1, 2011.

[Prata 11b] A. Prata, and T. Chambel, “Mobility in a Personalized and Flexible Video Based Transmedia Environment”. International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Lisbon, Portugal, Nov 20-25, 2011.

[Soares 12] M. Micael Soares, P. Viana, “TV Recommendation and Personalization Systems: integrating broadcast and video on-demand services,” European Conf. on the Use of Modern Information and Communication Technologies, Gent, Belgium, March, 2012.

[Strover 12] S. Strover, W. Moner and N. Muntean, “Immersive television and the on-demand audience,” Intl. Communication Association Conference, May 2012, Phoenix, USA.

[Kim 03] M.B. Kim, J. Nam, W. Baek, J. Son, and J. Hong, “The adaptation of 3D stereoscopic video in MPEG-21 DIA,” Signal Processing: Image Communication, vol. 18, pp. 685-697, Sep. 2003.

[Ozbek 07] N. Ozbek, A.M. Tekalp, and E.T. Tunali, “A new scalable multi-view video coding configuration for robust selective streaming of free-viewpoint TV,” IEEE Intl. Conference on Multimedia and Expo, pp.1155-1158, July 2007.

[Schreer 10] Schreer, Oliver. Final Report of the 3D@SAT project. European Space Agency, Future Projects and Applications Division, September 2010.

[Visch 10] T. Visch, S. Tan, D. Molenaar, “The emotional and cognitive effect of immersion in film viewing,” Cognition & Emotion, 24: 8, pp.1439-1445, 2010.

Growing Documentary: Multi-User-Directed Interactive Storytelling Platform

Janak Bhimani*, Ali Almahr, Annisa Mahdia, Daisuke Shirai, Naohisa Ohta
Keio University, Graduate School of Media Design
Yokohama, Kanagawa, Japan
*janak; alia; annisa; dais; naohisa @kmd.keio.ac.jp

Abstract. The Growing Documentary is a platform for computer supported cooperative work (CSCW) using multi-user-directed content to produce digital video productions that can be remixed, reworked and built upon as the story, and storytellers, change and adapt. The interactive environment of the Growing Documentary supports and encourages more people to tell their stories together in new and innovative ways utilizing social media, digital video technology, and traditional production techniques.

Keywords: Growing Documentary, Interactive Storytelling, Non-linear Narrative, Digital Video Collaboration

1 Inspiration → First Prototype

The first iteration of the Growing Documentary was a short documentary about the devastation and aftermath of March 11th from the point of view of three amateur and professional photographers from different walks of life who were all affected by the earthquake and tsunami. The film, "lenses + landscapes" [2], was produced by crowd-sourcing the photographers, translators and audio via social networking services (SNS). The Growing Documentary is a social platform for individual and community cinematic expression. Through the use of a cloud-based file sharing system, the photographers' hi-resolution raw still images were combined with multi-framed HD video to render a very high resolution movie clip through the use of community resources. As a result of the interactivity and collaboration, "lenses + landscapes" was produced in less than two months and shown, in 4K (4096x2160) [1], at a special screening of the 2011 Tokyo International Film Festival.

2 Hands-on Non-linear Narrative Collaboration

The Growing Documentary is an important step in furthering the field of interactive storytelling [4]. The Growing Documentary platform allows collaborators to not only interact with one another in the production of a story, but also the elements of the story and, moreover, the story itself becomes an interactive experience. Ideas, images, videos and sounds can be shared by the storytellers as well as the audience. Content is not only generated by users; but multiple users become stake-holders in the creative

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

process through the interactivity in controlling and directing various aspects of the narrative.

By incorporating SAGE OptIPortables [3] into the creative process of the Growing Documentary, digital contents over gigabit pipelines around and throughout the world, people have the potential to collaborate, regardless of location, at any time in order to produce high-quality digital contents by and for the global community.

3 Demonstration at FutureTV 2012

For the demonstration at the FutureTV 2012 Workshop, we will illustrate how the Growing Documentary concept complements the vision of linked TV through the integration of various interfaces by users in different geographic locations collaborating in real-time to produce dynamic interactive contents. There are two parts to the demonstration: in Stage 1, assets, captured by users from different "directors" using DSLR & digital video cameras and smartphones will be acquired; in Stage 2, the "curator" in Berlin will be connected via a broadband connection to the "users" in Japan to collaborate in real-time through the implementation of a SAGE unit -- only in Japan. (Figure 1). The demo presenter in Berlin will be curating the demo from a laptop. A projector or monitor (screen) will be required to show the interactivity.

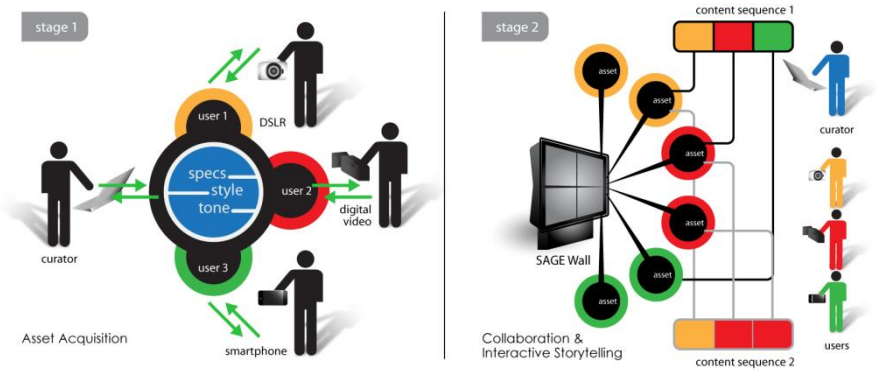


Fig. 1. Growing Documentary FutureTV 2012 demonstration schematic

4 References

1. 4K Resolution. http://www.pcmag.com/encyclopedia_term/0,2542,t=4K+resolution&i=57419,00.asp .
2. *"lenses + landscapes"*. http://vimeo.com/31093347
3. Renambot, L. et Al. 2004. SAGE: the Scalable Adaptive Graphics Environment. In *Proceedings of WACE 2004, (The 4th Workshop on Advanced Collaborative Environments)*
4. Ursu, M. F. et al. 2009. Interactive Documentaries: A Golden Age *COMPUT. ENTERTAIN.* 7, 3, ARTICLE 41 (SEPTEMBER 2009)

Semi-Automatic Video Analysis for Linking Television to the Web

Daniel Stein[1], Evlampios Apostolidis[2], Vasileios Mezaris[2], Nicolas de Abreu Pereira[3], and Jennifer Müller[3]

[1] Fraunhofer Institute IAIS, Schloss Birlinghoven,
53754 Sankt Augustin, Germany, daniel.stein@iais.fraunhofer.de
[2] Informatics and Telematics Institute, CERTH,
57001 Thermi-Thessaloniki, Greece, apostolid@iti.gr bmezaris@iti.gr
[3] rbb – Rundfunk Berlin-Brandenburg, 14482 Potsdam, Germany
nicolas.deabreupereira@rbb-online.de, jennifer.mueller@rbb-online.de

Abstract. Enriching linear videos by offering continuative and related information via audiostreams, webpages, or other videos is typically hampered by its demand for massive editorial work. Automatic analysis of audio/video content by various statistical means can greatly speed up this process or even work autonomously. In this paper, we present the current status of (semi-)automatic video analysis within the LinkedTV project, which will provide a rich source of data material to be used for automatic and semi-automatic interlinking purposes.

1 Introduction

Many recent surveys show an ever growing increase in average video consumption, but also a general trend to simultaneous usage of internet and TV: for example, cross-media usage of at least once per month has risen to more than 59% among Americans [Nielsen, 2009]. A newer study [Yahoo! and Nielsen, 2010] even reports that 86% of mobile internet users utilize their mobile device while watching TV. This trend results in considerable interest in interactive and enriched video experience, which is typically hampered by its demand for massive editorial work. This paper introduces several scenarios for video enrichment as envisioned in the EU-funded project "Television linked to the Web" (LinkedTV),[4] and presents its workflow for an automated video processing that facilitates editorial work. The analysed material then can form the basis for further semantic enrichment, and further provides a rich source of high-level data material to be used for automatic and semi-automatic interlinking purposes.

This paper is structured as follows: first, we present the envisioned scenarios within LinkedTV, describe the analysis techniques that are currently used, provide manual examination of first experimental results, and finally elaborate on future directions that will be pursued.

[4] www.linkedtv.eu

2 LinkedTV Scenarios

The audio quality and the visual presentation within videos found in the web, as well as their domains are very heterogeneous. To cover many possible aspects of automatic video analysis, we have identified several possible scenarios for interlinkable videos within the LinkedTV project, which are depicted in three different settings elaborated below. In Section 4, we will present first experiments for Scenario 1, which is why we keep the description for the other scenarios shorter. See Figure 1 for tentative screenshots of each scenario.

Scenario 1: News Broadcast The first scenario uses German news broadcast as seed videos, taken from the Public Service Broadcaster Rundfunk Berlin-Brandenburg (RBB).[5] The main news show is broadcast several times each day, with a focus on local news for Berlin and the Brandenburg area. The scenario is subject to many restrictions as it only allows for editorially controlled, high quality linking with little errors. For the same quality reason only links selected from a restricted white-list are allowed, for example only videos produced by the Consortium of public-law broadcasting institutions of the Federal Republic of Germany. The audio quality can generally be considered to be clean, with little use of jingles or background music. Local interviews of the population might have a minor to thick accent, while the eight different moderators have a very clear and trained pronunciation. The main challenge for visual analysis is the multitude of possible topics in news shows. Technically, the individual elements will be rather clear: contextual segments (shots or stories) are usually separated by visual inserts and there are only few quick camera movements.

Scenario 2: Cultural Heritage This scenario deals with Dutch cultural videos as seed, taken from the Dutch show "Tussen Kunst & Kitsch", which is similar to the "Antique Roadshow" as produced by the BBC. The show is centered around art objects of various types, which are described and evaluated in detail by an expert. The restrictions for this scenario are very low, allowing for heavy interlinking to, e.g., Wikipedia and open-source video databases. The main visual challenge will be to recognise individual artistic objects so as to find matching items in other sources.

Scenario 3: Visual Arts At this stage in the LinkedTV project, the third scenario is yet to be sketched out more precisely. It will deal with visual art videos taken mainly from a francophone domain. Compared to the other scenarios, the seed videos will thus be seemingly arbitrary and offer little pre-domain knowledge.

[5] `www.rbb-online.de`



(a) News Broadcast (b) Cultural Heritage (c) Visual Arts

Fig. 1. Screenshots from the tentative scenario material: (a) taken from the German news show "rbb Aktuell" (`www.rbb-online.de`), (b) taken from the Dutch Art show "Tussen Kunst & Kitsch" (`tussenkunstenkitsch.avro.nl`), (c) taken from the FireTraSe project [Todoroff et al., 2010]

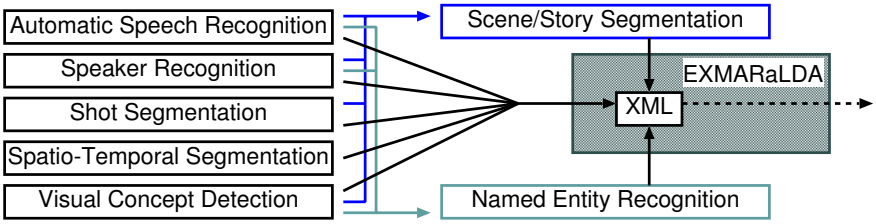


Fig. 2. Workflow of the data derived in LinkedTV

3 Technical Background

In this section, we describe the technologies with which we process the video material, and briefly describe the annotation tool as will be used for editorial correction of the automatically derived information.

In the current workflow, we derive stand-alone, automatic information from: automatic speech recognition (ASR), speaker identification (SID), shot segmentation, spatio-temporal video segmentation, and visual concept detection. Further, Named Entity Recognition (NER) and scene/story detection take the information from these information sources into account. All material is then joined in a single xml file for each video, and can be visualized and edited by the annotation tool EXMARaLDA [Schmidt and Wörner, 2009]. See Figure 2 for a graphical representation of the workflow.

Automatic Audio Analysis For training of the acoustic model, we employ 82,799 sentences from transcribed video files. They are taken from the domain of broadcast news and political talk shows. The audio is sampled at 16 kHz and can be considered to be of clean quality. Parts of the talk shows are omitted when, e.g., many speakers talk simultaneously or when music is played in the background. The language model consists of the transcriptions of these audio files, plus additional in-domain data taken from online newspapers and RSS feeds. In total, the material consists of 11,670,856 sentences and 187,042,225 running words. Of these, the individual subtopics were used to train trigrams with modified Kneser-Ney discounting, and then interpolated and optimized for

perplexity on a with-held 1% proportion of the corpus. The architecture of the system has been described in [Schneider et al., 2008].

Temporal Video Segmentation Video shot segmentation is based on an approach proposed in [Tsamoura et al., 2008]. The employed technique can detect both abrupt and gradual transitions; however, in certain use cases (depending on the content) it may be advantageous for minimizing both computational complexity and the rate of false positives to consider only the detected abrupt transitions. Specifically, this technique exploits image features such as color coherence, Macbeth color histogram and luminance center of gravity, in order to form an appropriate feature vector for each frame. Then, given a pair of selected successive or non-successive frames, the distances between their feature vectors are computed, forming distance vectors, which are then evaluated with the help of one or more SVM classifiers. In order to further improve the results, we augmented the above technique with a baseline approach to flash detection. Using the latter we minimize the number of incorrectly detected shot boundaries due to cameras flash effects.

Video scene segmentation draws input from shot segmentation and performs shot grouping into sets which correspond to individual scenes of the video. The employed method was proposed in [Sidiropoulos et al., 2011]. It introduces two extensions of the Scene Transition Graph (STG) algorithm; the first one aims at reducing the computational cost of shot grouping by considering shot linking transitivity and the fact that scenes are by definition convex sets of shots, while the second one builds on the former to construct a probabilistic framework towards multiple STG combination. The latter allows for combining STGs built by examining different forms of information extracted from the video (low level audio or visual features, visual concepts, audio events) while at the same time alleviating the need for manual STG parameter selection.

Spatiotemporal Segmentation Spatiotemporal segmentation of a video shot into differently moving objects is performed as in [Mezaris et al., 2004]. This unsupervised method uses motion and color information directly extracted from the MPEG-2 compressed stream. The bilinear motion model is used to model the motion of the camera (equivalently, the perceived motion of static background) and, wherever necessary, the motion of the identified moving objects. Then, an iterative rejection scheme and temporal consistency constraints are employed for detecting differently moving objects, accounting for the fact that motion vectors extracted from the compressed stream may not accurately represent the true object motion. Finally, both foreground and background spatiotemporal objects are identified.

Concept Detection A baseline concept detection approach is adopted from [Moumtzidou et al., 2011]. Initially, 64-dimension SURF descriptors are extracted from video keyframes by performing dense sampling. These descriptors are then

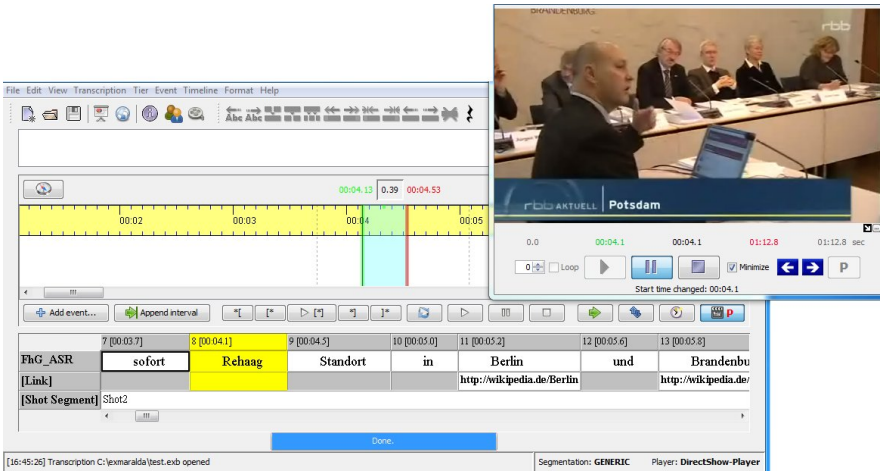


Fig. 3. Screenshot of the EXMARaLDA GUI, with automatically derived information extracted from an rbb video.

used by a Random Forest implementation in order to construct a Bag-of-Words representation (including 1024 elements) for each one of the extracted keyframes. Random Forests are preferred instead of Nearest Neighbor Search, in order to reduce the associated computational time without compromising the detection performance. Following the representation of keyframes by histograms of words, a set of linear SVMs is used for training and classification purposes, and the responses of the SVM classifiers for different keyframes of the same shot are appropriately combined. The final output of the classification for a shot is a value in the range [0, 1], which denotes the Degree of Confidence (DoC) with which the shot is related to the corresponding concept. Based on these classifier responses, a shot is described by a 346-element model vector, whose elements correspond to the detection results for the 346 concepts defined in the TRECVID 2011 SIN task.[6]

Annotation We make use of the "Extensible Markup Language for Discourse Annotation" (EXMARaLDA) toolkit [Schmidt and Wörner, 2009], which provides computer-supported transcription and annotation for spoken language corpora. EXMARaLDA supports many different OS and is written in Java. Currently, it is maintained by the Center of Speech Corpora, Hamburg, Germany. Here, the editor will have the possibility to modify automatic analysis results or add new annotations manually. This will be necessary especially with regard to contextualisation which still involves qualitative data analysis to some, if lesser, extent [de Abreu et al., 2006].

[6] `www-nlpir.nist.gov/projects/tv2011/`



Fig. 4. Spatiotemporal Segmentation on video samples from news show "RBB Aktuell"

4 Experiments

In this section, we present the result of the manual evaluation of a first analysis of videos from Scenario 1.

The ASR system produces reasonable results for the news anchorman and for reports with predefined text. In interview situations, the performance drops significantly. Further problems include named entities of local interest, and heavy distortion when locals speak with a thick dialect. We manually analysed 4 sessions of 10:24 minutes total (1162 words). The percentage of erroneous words were at 9% and 11% for the anchorman and voice-over parts, respectively. In the interview phase, the error score rose to 33%, and even worse to 66% for persons with a local dialect.

In preliminary experiments on shot segmentation, zooming and shaky cameras were misinterpreted as the beginnings of new shots. Also, reporters' flashlights confused the shot detection. In total, the algorithm detected 306 shots, which are mostly accurate on the video level but were found to be too fine-granular for our purposes. A human annotator only marked one third, i.e., 103 temporal segments, to be of immediate importance. In a second iteration with conservative segmentation, most of these issues could be addressed.

Indicative results of spatiotemporal segmentation on these videos, following their temporal decomposition to shots, are shown in Figure 4. In this figure, the red rectangles demarcate automatically detected moving objects, which are typically central to the meaning of the corresponding video shot and could potentially be used for linking to other video, or multimedia in general, resources. Currently, an unwanted effect of the automatic processing is the false recognition of name banners which slide in quite frequently during interviews, which indeed is a moving object but does not yield additional information.

Manually evaluating the top-10 most relevant concepts according to the classifiers' degrees of confidence, produced further room for improvement. As mentioned earlier, for the detection we have used 346 concepts from TRECVID, but evidently some of them should be eliminated for being either extremely specialized or extremely generic ("Eycariotic Organisms" is an example of an extremely generic concept), which renders them practically useless for the analysis of the given videos. See Table 1 for two examples, the upper one with good results and the second one with more problematic main concepts detected. Future work on concept detection includes exploiting the relations between concepts (relations such as "man" being a specialization of "person"), in order to improve concept detection accuracy.

Table 1. Top 10 TREC-Vid concepts detected for two example screenshots from scenario 1, and their human evaluation

.

	Concept	estimation
	body part	good
	graphic	comprehensible
	charts	wrong
	reporters	good
	person	good
	face	good
	primate	unclear
	text	good
	news	good
	eukaryotic organism	correct
	event	unclear
	furniture	unclear
	clearing	comprehensible
	standing	wrong
	talking	good
	apartment complex	unclear
	anchorperson	wrong
	sofa	wrong
	apartments	good
	body parts	good

5 Conclusion

While still at an early stage within the project, the manual evaluation indicates that a first waystage of usable automatic analysis material could be achieved. We have identified several challenges which seem to be of minor to medium complexity. In the next phase, we will extend the analysis on 350 videos of a similar domain, and use the annotation tool to provide ground-truth on these aspects for quantifyable results. In another step, we will focus on an inproved segmentation: since interlinking for a user is only relevant in a certain time intervall, we prepare to merge the shot boundaries and derive useful story segments based on simultaneous analysis of the audio and the video information, e.g., speaker segments correlated with background changes.

Acknowledgements This work has been partly funded by the European Community's Seventh Framework Programme (FP7-ICT) under grant agreement n° 287911 LinkedTV.

References

[de Abreu et al., 2006] de Abreu, N., Blanckenburg, C. v., Dienel, H., and Legewie, H. (2006). *Neue wissensbasierte Dienstleistungen im Wissenscoaching und in der Wissensstrukturierung*. TU-Verlag, Berlin, Germany.

[Mezaris et al., 2004] Mezaris, V., Kompatsiaris, I., Boulgouris, N., and Strintzis, M. (2004). Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(5):606 – 621.

[Moumtzidou et al., 2011] Moumtzidou, A., Sidiropoulos, P., Vrochidis, S., Gkalelis, N., Nikolopoulos, S., Mezaris, V., Kompatsiaris, I., and Patras, I. (2011). ITI-CERTH participation to TRECVID 2011. In *TRECVID 2011 Workshop*, Gaithersburg, MD, USA.

[Nielsen, 2009] Nielsen (2009). Three screen report. Technical report, Nielsen Company.

[Schmidt and Wörner, 2009] Schmidt, T. and Wörner, K. (2009). EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19:4:565–582.

[Schneider et al., 2008] Schneider, D., Schon, J., and Eickeler, S. (2008). Towards Large Scale Vocabulary Independent Spoken Term Detection: Advances in the Fraunhofer IAIS Audiomining System. In *Proc. SIGIR*, Singapore.

[Sidiropoulos et al., 2011] Sidiropoulos, P., Mezaris, V., Kompatsiaris, I., Meinedo, H., Bugalho, M., and Trancoso, I. (2011). Temporal video segmentation to scenes using high-level audiovisual features. *Circuits and Systems for Video Technology, IEEE Transactions on*, 21(8):1163 –1177.

[Todoroff et al., 2010] Todoroff, T., Madhkour, R. B., Binon, D., Bose, R., and Paesmans, V. (2010). FireTraSe: Stereoscopic camera tracking and wireless wearable sensors system for interactive dance performances - Application to "Fire Experiences and Projections". In Dutoit, T. and Macq, B., editors, *QPSR of the numediart research program*, volume 3, pages 9–24. numediart Research Program on Digital Art Technologies.

[Tsamoura et al., 2008] Tsamoura, E., Mezaris, V., and Kompatsiaris, I. (2008). Gradual transition detection using color coherence and other criteria in a video shot meta-segmentation framework. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 45 –48.

[Yahoo! and Nielsen, 2010] Yahoo! and Nielsen (2010). Mobile shopping framework – the role of mobile devices in the shopping process. Technical report, Yahoo! and The Nielsen Company.

Second-Screen Use in the Home: An Ethnographic Study

Jeroen Vanattenhoven and David Geerts

Centre for User Experience Research, KU Leuven – IBBT
Parkstraat 45 bus 3605, 3000 Leuven, Belgium

{jeroen.vanattenhoven,david.geerts}@soc.kuleuven.be

Abstract. This paper presents the results of an ethnographic study to investigate how people use so-called second-screens in front of the television in the home. Our goal is to inform the design of better and more unified TV experiences. In December 2011, 12 households participated in a diary study for a three-week period. They were asked to record media usage and communication behavior while watching TV or video. In the weeks following the diary study, semi-structured interviews were held at the participants' homes to discuss and clarify what participants had recorded in their diaries. These interviews were transcribed and coded in NVivo. We report the qualitative results concerning, second-screen uses, viewing modes, social media, devices, content (genre), discovery, and communication. Then, we formulate what these results imply for the design of future TV experiences.

Keywords: second-screen, interactive television, Social TV, ethnography, user experience, media use in the home, user studies

1 Introduction

In recent years, we have seen a convergence of broadcast television and Internet services such as video streaming and social networks. This more hybrid form of TV can move beyond today's second-screen applications and social TV experiences. An intelligent integration of TV and Web content can results in seamless new user experiences. The question is then: What does an intelligent integration entail? How should we design these new kinds of applications? To answer these questions, we first need to understand which kinds of applications are useful, and how these new TV services fit into the everyday life of the user. To that effect, we conducted a diary and interview study with 12 households on the uses of second-screens. The obtained results can inform the design of future hybrid TV applications and technologies.

This paper is structured as follows: section two provides an overview of related work in the area of media use in the home, section three describes how the diary and interview study was setup and how the results were analyzed, section four presents the main findings of the study, section five elaborates on the implications of the results on the design of future, hybrid TV applications, and section six concludes this paper and identifies topics for future research.

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

2 Related work

Before the era of interactive and digital TV, traditional broadcast TV already formed an important topic of research, in part because of its role as a mass medium. Lull [8] carried out an ethnographic study, using systematic participant observation to construct a typology of the social uses of television. The results indicate that mass media are valuable social resources that aid in the construction and maintenance of the relations between the members of a household. The typology describes structural uses on the one hand, i.e. how the television is an anchor around which family activities such as mealtimes are organized, and relational uses on the other hand, i.e. how the viewing of TV programs can diminish the need to talk to each other. Another extensive study by Lee and Lee [7] investigated how and why people watch TV in order to inform the design of interactive TV. Their findings show that people have different levels of attention when watching TV and that they sometimes engage in other activities simultaneously.

The advent of Digital TV and Digital Video Recorder (DVR) systems disconnected TV consumption from the broadcast schedules. Based on analysis of in-depth interviews with early-adopters, Barkhuus and Brown describe how the ability to watch video content on the users' initiative, resulted in a more active form of TV watching [1]. Darnell [2] observed how people use Digital TV and Digital Video Recorder (DVR) systems at home, and found that these devices are used to avoid commercials and to discover new programs to watch after the currently watched show ended. Rudström et al. performed a web survey to investigate attitudes and behavior related to on-demand TV [9]. Their findings show that people use time-shifting for rescheduling, catch-up and repeats. Furthermore, some people indicated that they did not necessarily want to depend on a broadcast schedule, except for news, sport events and other live broadcasts. Finally, a study on media use in the homes of 27 families [10] revealed that people would like support for media presentation and consumption through the TV, and look at photos and videos together with other household members. In addition, users expressed the need to access online services such as e-mail access and the use of social networks.

3 Methodology

For this study, 12 households were recruited in and around the cities of Antwerp, Ghent, Hasselt and Leuven in Belgium. The 12 households comprised three singles, two couples, two families with very young children, and five larger families with teenagers. By recruiting diverse households from different areas in and around cities, we aimed to gain a broad insight into the use of second-screens in the home. Participants were recruited via convenience sampling, social media, and the university newsletter. A 100€ gift certificate was provided for each household as incentive after the final part of the study.

To gain a thorough understanding of the complexity of media use in the home, a qualitative approach was chosen. Diary studies are a meaningful tool to investigate

private settings such as the home, and they provide a more accurate record of actual behavior. Therefore, they have been used before in similar studies, investigating media use in the home [6]. In the first phase a diary study was carried out, which formed the basis for the follow-up interviews. The diaries were specifically designed for this study and took the form of an A3-size, paper booklet (see **Fig. 1**). The first pages contained an introduction to the study, followed by a basic questionnaire about the participants including demographics, media use and preferences. The actual diary pages formed the main part of the diaries. Each day in the three-week diary period, participants could write on two A3-pages. On each line participants could fill in the starting and ending time as well as the title of each TV program or video they watched, with whom they were watching this program, and which second-screens they were using. In the families with teenagers, every family member over 12 years received an individual diary. In two cases, an individual diary was also handed to a nine-year-old boy and an 11-year-old girl who wanted to participate along with the rest of the family.

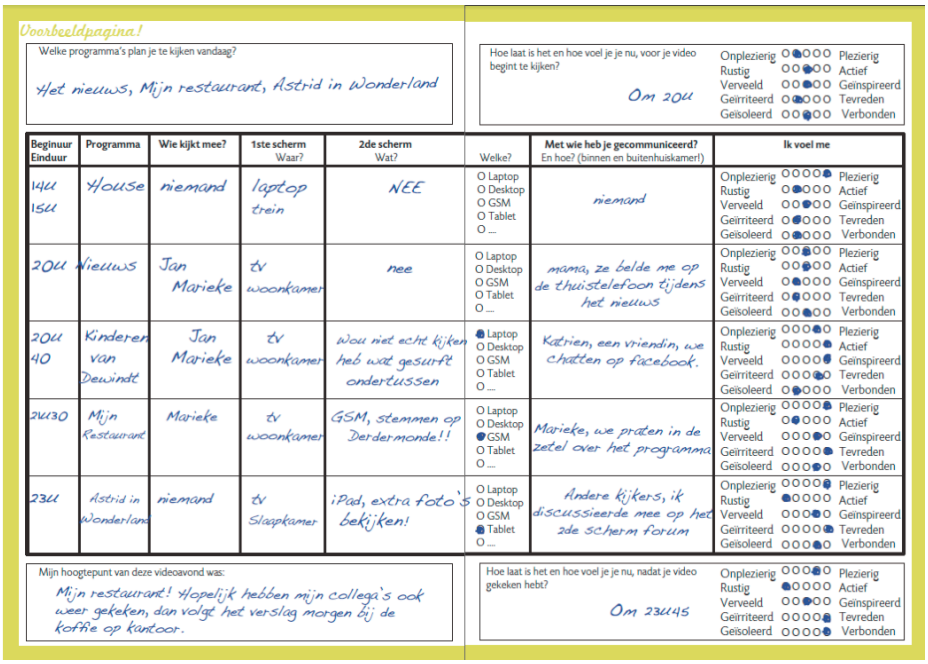


Fig. 1. Sample diary pages

In the weeks after the diary period, three researchers visited the households to conduct semi-structured interviews. In these interviews, we walked through the diary with the households to ensure the resulting data was firmly grounded into the participants' actual recorded behavior. Before the study, an interview guide was constructed to make sure all topics of interest would be covered, and to minimize possible differences between the interviews conducted by three researchers. The interviews approx-

imately lasted between 30 minutes, and one hour and 30 minutes. Then, the interview recordings were transcribed.

A Grounded Theory approach was used for analysis. This approach allows researchers to organize the data, and generate insights that are firmly grounded in the data [4]. To this extent, the interview transcripts were imported into NVivo, a qualitative analysis software package. One researcher carried out a first iteration of open coding. The resulting codes and coding structure were then discussed with another researcher, after which a second iteration of coding was applied to the data.

4 Results

By extensively coding the interview transcripts a rich data set was created, allowing us to explore a number of aspects relating media use in the home environment: second-screen uses, viewing modes, content (genre), and discovery. The results will be described per high-level category from Section 4.1 through Section 4.4. In each section, we will first explain the meaning of the categories, and describe the subcategories. Then, quotes from participants will be used to illustrate these subcategories.

Additional analysis of the data was carried out via "Matrix Coding". This operation in the qualitative software package NVivo allows us to retrieve overlap between different categories. This means, for example, that we can look at how second-screen use relates to participants' viewing modes. These results are presented from Section 4.5 through Section 4.9, and describe the relations between second-screen uses on the one hand, and devices, viewing mode, social media, content on the other hand, and the relations between communication and social media.

Table 1. Overview of high-level categories

High-Level Category	Households	References
Bad TV Habits	4	8
Content	12	158
Devices	12	187
Diary Use and Feedback	7	15
Digital TV	4	5
Discovery	7	9
Everyday Life	12	93
Locations	12	87
Mood & Emotion	11	24
Motivations	11	113
Quality	9	15
Second Screen Uses	11	104
Social Media	2	8
Togetherness	12	99
Viewing Mode	11	75
Watching Behavior	12	84

Table 1 provides an overview of the high-level categories that were derived from the data. The references column shows how many segments of the interview transcripts have been coded under the respective high-level category. The households column indicates the number of households that corresponds with each high-level category. There are, for example, 113 segments in the interview transcripts relating to "Motivations"; these references come from 11 households. The results described in this paper are a selection of all collected and analyzed data; this was done to present the most interesting findings related to the workshop.

4.1 Second-Screen Uses

In the data 104 references from 11 households for second-screen use were discovered. The first subcategory, "program-related", entails second-screen use with a specific relation to the TV program. The second subcategory, "not program-related", is second-screen use that is not related to what happens on TV. The references for "not-program related" (84 references from 11 households) outnumber those for "program-related" second-screen use (20 references from five households).

The **program-related** second-screen uses involve activities such as looking up information, talking to other viewers, and voting on candidates. A single woman (34y) explained: *"And then 'So You Think You Can Dance' was on TV. I remember that the winner of the previous edition also danced 'Pirates of the Caribbean', and that was really good. So I looked up that movie clip from last year"*. A woman (couple, 22y) was interested in talking to other viewers: *"I visit Twitter [using the hashtag of the TV program] to see if anybody else is watching, and if that's the case, I respond to them"* and *"I really like to see those opinions, such as, 'I find this important' or, 'that can be better'"*.

The **not-program related** subcategory consisted of arranging personal affairs, checking social networks for updates, and listening to music. A single woman (34y) illustrated how she made some personal arrangements: *"It's the evening you [the interviewer] visited. That evening, I actually didn't watch TV because I was going to a [friend's] party. But then I started watching TV after all, because I had to make some arrangements with my friends for after the party anyway; I also checked my email"*. Checking social networks also happened frequently, as a woman (27y) describes: *"Yes, usually I'm checking my emails, or I'm checking Facebook to see what everyone has posted"*. Finally, a part of the interview transcript shows how a single woman (60y) uses Skype to talk to her daughter during the news:

Interviewer:	*"And then we have a look at Monday, November 28th"*
Woman:	*"Using Skype for 30 minutes"*
Interviewer:	*"Yes, and again during the news"*
Woman:	*"Yes, [referring to her daughter who lives in Africa], her baby is in bed, she has time [to talk] then."*

4.2 Viewing Modes

A second category identified in the interview data, relates to the different viewing modes people experience when watching TV (75 references from 11 households). People's level of attention or immersion toward the TV can vary greatly. We have called these subcategories "dedicated" viewing in which people's full attention is direct toward the TV, "mixed mode" in which attention shifts back and forth between the TV and secondary activities, and "background" for situations in which the TV merely serves as background "noise" for other activities.

The first mode, **dedicated** viewing, has 21 references to the data in five households. A father (30y) illustrates: *"I find that going to the movie theatre offers a complete experience: the lights goe out, and there is only that movie, and for an hour and a half or two hours, you are really immersed. The feeling of being disconnected from the world for a brief moment, that's what I miss… Even if we start watching a movie here [at home] … and we want to see a good movie, … something might happen [with the kids] and you have to go upstairs, so you have to pause the movie"*.

In **mixed mode**, the attention shifts back and forth between the TV and the second-screen: *"Today, my boyfriend unplugged the Internet because I'm a total Facebook addict… That's why for me this is really 'work-a-light', when I'm working in the couch in front of the TV. It's a complete convergence of the entertainment of the TV, the social aspect from the social networks, and information"* (mother, 30y). This mode occurred the most in our data with 46 references in 10 households.

Finally, the TV is also used as **background** for other activities (eight references in five households) as a man (couple, 28y) describes: *"Sometimes the TV is switched on while we are doing the dishes, or during dinner"*.

4.3 Content (genre)

In the interview transcripts participants mentioned the kinds of content they were consuming while watching TV (158 references from 12 households). A whole range of genres, or different kinds of content, was reported: lighter TV genres (50 references), movies (31 genres), news (27 references), short movie clips (21 references), TV-series such as "Dexter" (12 references), sports (seven references), DVDs (six references), and programs for children (four references). It was not straightforward to derive mutually exclusive categories, as there are for example also lighter movies genres and TV-series for example.

Most references are related to **lighter TV genres**: *"We watched '[My Restaurant]'. We started watching this, so you also want to know how it ends"* (daughter, 18y). This subcategory included reality TV, talk shows, and dancing and singing contests.

Participants' statements about **movies** highlight a number of aspects related to media use in the home. A single woman (31y) explains how she records movies to watch them later on: *"Yes, it happens, especially with movies, that I want to record something to watch later on, but it can take months before I actually see it*". A mother (47y) says: *"For me it's about togetherness. I don't really know any programs that are important to me; for me it's more to have a seat and relax. I find that quite cozy.*

Only when we are watching a movie, then I really want to see it". The father of the same family clarifies the latter: *"Every week, we watch a movie with the whole family on Friday or Saturday evenings"*.

4.4 Discovery

This category describes how participants discovered new content, or used several resources to consult the broadcasting schedules. Subcategories include EPG and TV guide (17 references), recommendations (14 references), reviews (eight references), and shared (one references).

The **EPG and TV guide** subcategory contains statements about how participants used their EPG and paper TV guides. A single woman (34y) explains how reading TV guide from a free newspaper on the train, is actually part of her daily routine: *"I usually check the Metro newspaper in the morning, when I'm commuting by train. I have the habit of checking out the TV guide even though I sometimes already know that I won't be watching TV that same evening"*. Another woman (couple, 27y) talks about the application she uses on her iPhone, and the fact that the EPG offered by her digital TV provider, does not include the same channels as the iPhone application: *"It [the iPhone application 'TV Gids'] doesn't show all the channels, because we have 'TV Vlaanderen' [the digital TV provider]. This means we have a lot of English channels. The iPhone app doesn't provide information for these channels. Therefore, we often have to consult the EPG on our TV"*. A father of two (31y) illustrates how advertising of new TV programs and rigid broadcast schedules influence the need for a TV guide: *"We are not in the habit of checking the TV guides, because we never used to. We used to have only two channels, and the programming for those channels was so rigid that when you watched TV for one week, you immediately knew what would be on TV at any time. Nowadays, when a new program is launched, so much advertising is made that it's almost impossible not to notice"*. A father (45y) says how he chooses between using the EPG and the TV guide: *"I used the newspaper before, and the EPG while watching TV"*. Another father (47y) adds: *"The EPG provides a much better overview, and you're already sitting in front of the TV anyway"*.

The **recommendations** subcategory informs us how participants give and receive recommendations, and what kinds of recommendations are made. A single woman (34y) explains: "*Today, I informed a colleague who works out a lot, that he should watch 'Ook getest op mensen' [a consumer program], because they would talk about sports nutrition. I just happened to hear it on the radio*". The same participant shows what kind of recommendations she receives: "*Those are things they know I'm interested in*".

Next to social recommendations, participants also consulted various magazines that provide content **reviews**, as a man (couple, 36y) illustrates: "*There are a number of online forums and websites where you can find out how a certain drama series is rated. I regularly visit those sites to check out the new TV shows for the coming season, and then I try some of them out. Metacritic is one of those websites*".

4.5 Second-Screen Use and Devices

When we look at the overlap between second-screen use, and participants' devices in the encoded data, we notice that laptops (28 references) and smartphones (eight references) were used most for non-program related second-screen use. Other devices included an iPod and a tablet. A father of two (30y) explains: *"For me, the nice thing about smartphones is that you can install a lot of apps… It often happens that something is on TV, which is not interesting to me. So, I open my laptop to fool around, or I look on my Samsung to check if there are any new games to try out"*.

For program-related second-screen use, we found that laptops (three references), iPods (two references) and Smartphones (one references) were used. A man (couple, 28y) says: *"I use my iPod or laptop mainly to look up something or to use email while she [his girlfriend] is mainly occupied with Twitter and Facebook, more than I am anyway…. Most of the time it is not about the program, but sometimes it is"*.

4.6 Second-Screen Use and Viewing Mode

In this section, we describe the results for the relation between second-screen uses and participants' viewing mode (47 references from nine households). One thing we note here is that there are no "dedicated" viewing references for second-screen uses.

With regard to not-program related second-screen uses, there are 38 references for "mixed mode" and three for "background". A single woman (31y) illustrates the former: *"On Facebook I check who's online, what was posted, whether I have to wish someone a happy birthday"*. The same participant also clarifies the latter: *"'The Bachelor', actually I don't like it. But there is nothing else on TV so I started to chat with [a friend] on Facebook"*.

For program-related second-screen uses there are six references of "mixed mode". A mother of two explains (30y): *"If you really have to write a text, then you cannot watch TV at the same time. So either I disappeared into my text, and the TV merely became radio; then I was actually working without the TV even if it was switched on"*.

4.7 Second-Screen Use and Social Media

The search for overlap between second-screen use not related to the TV program, and social media shows 16 references to Facebook, and 11 references to Skype. A woman (couple, 22y) says: *"Sometimes I'm using Skype to talk with a friend when the TV is playing, but then I'm not really paying attention to the program"*. Another woman (couple, 27y) explains how she uses Facebook: *"Usually, I'm checking my emails, or I'm on Facebook to see what everyone has posted. This doesn't mean that I'm looking for what they've posted about the program, just what they've posted in general"*.

For second-screen use related to the TV program, Twitter (five references) is used the most. A woman (couple, 22y) says: *"Usually, I tell him: 'I tweeted this'. This is something I tell him [her boyfriend] first. It [Twitter] is not something I do on my own; it's part of our conversation"*.

4.8 Second-Screen Use and Content

The kind of content, or genre, is another aspect that relates to second-screen uses (22 references from five households). For the not-program related second-screen uses, there are only five references.

With 15 references, the lighter TV genres such as talk shows and song contests, are the most reported kinds of content for not-program related second screen uses. A single woman (34y) explains: *"I was searching for salsa music on YouTube. So it has nothing to do with 'American Topmodel'"*. A grandmother (60y) illustrates how she uses a tablet during the news (four references): *"I turn on the news almost every day. Then, I also use Facebook on my iPad to chat with my daughter"*.

For news, the second-screen use is illustrated by a single woman (34y): "*What I also did during the news, was surfing the web about toys. I was looking for something to buy for my godchild*".

4.9 Communication and Social Media

This final section reports on the overlap between communication (a subcategory of Togetherness), and social media. The communication subcategory (66 references) contains three subcategories: communication during TV watching (33 references), communication before and after watching TV (22 references), and no communication during TV watching (11 references).

Communication during TV watching, contains statements that show how participants communicate, with each other or via social media, while they are watching TV. A woman (couple, 27y) explains: *"When I'm watching 'The Voice' on TV with my sister, then we talk to each other, but we won't post anything on Facebook. We don't use external channels; that really bothers me. For 'The Voice' you can post messages on Twitter, and then they appear on TV."* She goes on to explain why this bothers her: *"You [her boyfriend] should check out 'The Voice'; the messages appear right in the middle of the screen. The contents of those messages are always very silly, useless. I don't see the added value in that"*. Her boyfriend (36y) adds: *"For 'Music for Life' these messages are displayed at the bottom of the screen, so you don't notice them that much. And I find those messages quite funny sometimes. It might not provide much added value but it fits into the concept of 'Music for Life'"*.

For communication before and after watching TV, there is only one reference to social media. Therefore, it might be more interesting to report how participants communicate in real life about TV programs: *"In the coffee room at work, we talk about nothing else. It is really terrible; every day it's about TV."* (mother, 31y). Her husband (30y) adds: *"My colleagues at work are very interested in sports, unlike me. Therefore, I sometimes pick up on some sports-related news to tease my colleagues"*.

The subcategory "no communication during TV" contains statements in which participants explicitly said they did not communicate in any way while watching TV. This mostly involved movies, but also includes others TV content that the participants really want to see.

5 Implications for the Design of Future TV Experiences

In this section, we elaborate on the results and try to formulate implications for the design of future TV experiences. Given the ethnographic nature of our research, it is not always straightforward to compile a clear list with design implications or user requirements [3]. Our goal then, is to communicate the meaning of our results, and where possible, to explore what these results could imply for the design of future TV experiences.

Before we present the discussion, there is one issue we want to note: laptops and smartphones were the most reported second-screen devices. The reason for the low amount of tablets probably stems from the fact that they are still quite new, and the uptake does not reach the level of smartphones yet. Another reason is that only one of the 12 households included a participant with a tablet. We aimed to include more people with tablets at the outset of the study because tablets are seen as one of the more promising devices that can be used as second-screen. The reasons for this are related to their very convenient form-factor. Tablets have a large screen, compared to a smartphone, enabling better video viewing, and a lower weight, making it more mobile than a laptop. Furthermore, tablets have short start-up times compared to laptops. One participant stated that, because of this, he really had to make a conscious decision for using his laptop in the couch, as it was quite heavy and took a very long time to start up.

5.1 Not program-related second-screen applications

An important distinction we found in the use of today's second-screen devices is that between uses related to the TV program, and not related to the TV program. The majority of the uses was not related to the TV program, and involved relaxing, socializing, and conducting professional activities. In these situations people are not fully immersed into the program. The attention shifts from time to time between the TV and the second-screen device [7], sometimes because people find a specific section of the program not interesting, or commercials are shown. Instead of waiting for interesting content to return on TV (lean-back), viewers actively search for something useful or enjoyable to do (lean-forward). In the meantime the TV is not switched off in most cases; the program continues to play so people can pick-up again when it becomes more interesting again. Here, a number of opportunities for the design of future TV experiences can be explored.

As the content is less compelling in these "idle" moments, the broadcast could be viewed on a smaller screen. The screen space that becomes available this way can then include a number of items: participants' social networks and email were mentioned in our study. One option is then to display these kinds of content on the TV. This way, viewers do not have to shift their viewing angles from a device on their lap to their TV set. A second option is to present this content on a second-screen.

Another opportunity is providing related content such as movies, and biographies on actors. As one participant stated, content from previous editions is something people might want to look up in the meantime. An opportunity for formats targeting

viewer interaction during the TV program lies in the engagement of the viewers during these idle moments. A number of reality TV shows and TV contests already make use of voting procedures; these kinds of interactions could be expanded, in order to keep viewers interested. Additionally, our results showed that people consult a number of magazines and TV guides for recommendations, reviews and program scheduling. People might find the presentation and availability of this kind of content valuable.

Our research has shown that interaction with the TV program depends on the kind of content, or genre, that is being watched, confirming existing research [4]. Reality TV is a very suitable genre since these are sometimes designed to provoke strong reactions. Talk shows could offer an overview of biographies on the guests, and videos of previous editions, to cover parts in a broadcast that viewers might not like.

Finally, people start to organize their viewing schedules independent from the broadcast schedules more often [2][9]. Some of this content is never touched, despite the fact that some participants complained that they did not do anything useful in front of the TV during the whole evening. Presenting an overview of the recorded content that has not been touched, might convince a viewer to really watch this content.

5.2 Program-related second-screen applications

Our study did not show a lot of program-related second-screen uses. There are a number of explanations for this. Not a lot of people have a tablet at home, and most applications for second-screen use related to TV programs has been developed for tablets in Flanders. Another reason might be that not a lot of people know that these applications are available. Nonetheless, we can explore some opportunities for the design of program-related second-screen applications.

A number of TV programs establish a Twitter channel, which they use to gather viewer reactions. Our results show that care has to be taken in the way this feedback is presented. A first recommendation here, is to moderate the content. Otherwise, useless, or even offensive remarks may be automatically displayed. A second recommendation is to display the tweets on a subtle place on the screen. Only when the program host wants to highlight a specific tweet, it could be useful to display it in a more prominent place for a brief moment.

With regard to program genres, program-related second-screen applications should avoid movies, drama and documentaries that require viewers full attention. Furthermore, notifications during these moments of "dedicated" viewing should be turned of on all devices.

6 Conclusion

By recording actual behavior using diaries, and clarifying this data via interviews, we developed an understanding of how people use second-screens in the home. Therefore, the results in this study can inform the design of new, hybrid TV applications.

The majority of the reported second-screen uses were not related to the broadcast because viewers lost interest in the program. Future TV applications should seize these moments to regain viewers' attention by providing information on their social networks, program-related information, recommendations and recorded content. For second-screen applications that are related to the broadcast, viewer input should be moderated, and shown on the TV in a position that is not too prominent.

Future research should investigate on which screen what information should be displayed. Furthermore, multi-user situations will have a profound effect on the use of social media on TV, certainly with regard to privacy. Therefore, in-situ evaluations of applications could provide more insight into this matter.

Acknowledgements. The research leading to these results was carried out in the EU HBB-NEXT project and has received funding from the EC Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n°287848, and in the SMIF project (Smarter Media in Flanders), co-funded by IWT (Agency for Innovation by Science and Technology), a Flemish government agency.

7 References

1. Barkhuus, L., Brown, B.: Unpacking the television. ACM Transactions on Computer-Human Interaction. 16, 1–22 (2009).
2. Darnell, M.J.: How do people really interact with TV?: Naturalistic Observations of Digital TV and Digital Video Recorder Users. Comput. Entertain. 5, (2007).
3. Dourish, P.: Implications for design. Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 541–550. ACM, New York, NY, USA (2006).
4. Geerts, D., Cesar, P., Bulterman, D.: The implications of program genres for the design of social television systems. Proceedings of the 1st international conference on Designing interactive user experiences for TV and video. pp. 71–80. ACM, New York, NY, USA (2008).
5. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Transaction Publishers (1967).
6. Hess, J., Wulf, V.: Explore social behavior around rich-media: a structured diary study. Proceedings of the seventh european conference on European interactive television. pp. 215–218. ACM, New York, NY, USA (2009).
7. Lee, B., Lee, R.S.: How and why people watch TV: Implications for the future of interactive television. Journal of Advertising Research. 35, 9–18 (1995).
8. Lull, J.: The Social Uses of Television. Human Communication Research. 6, 197–209 (1980).
9. Rudström, Å., Sjölinder, M., Nylander, S.: How to choose and how to watch: an on-demand perspective on current TV practices, http://soda.swedish-ict.se/3944/ (2008).
10. Tsekleves, E., Whitham, R., Kondo, K., Hill, A.: Investigating media use and the television user experience in the home. Entertainment Computing. 2, 151–161 (2011).

Challenges for Multi-Screen TV Apps

Jean-Claude Dufourd°, Max Tritschler*, Niklas Schmücker*, Stephan Steglich*

°) Telecom ParisTech/LTCI
37-39 rue Dareau
75013 Paris, France
jean-claude.dufourd@telecom-paristech.fr
*) Fraunhofer FOKUS
Kaiserin Augusta Allee, 31
10589 Berlin, Germany
{max.tritschler,niklas.schmuecker,stephan.steglich}@fokus.fraunhofer.de

ABSTRACT
HbbTV, an emerging standard for interactive television, is being deployed in various European countries, creating bandwidth demand for interactive services in existing saturated broadcast networks. It is a first step in the direction of TV apps, but we believe the winning technology in that area is multi-screen apps on and around the TV. In this paper, we describe the current situation with HbbTV and apps, what can be done with existing technologies, high-potential scenarios and our plans to make them possible by combining existing standards with limited extensions, creating a platform implementing a multi-application service model, and open source software tools spanning management and production of this new type of service, within the Coltram project.

Keywords HbbTV, interactive TV applications, resource management in broadcast, multi-screen and distributed applications, inter-widget communication.

1 INTRODUCTION

HbbTV (Hybrid Broadband Broadcast TV), a standard targeting interactive TV, is getting more and more traction since its inception in 2009 and standardization by ETSI in June 2010 [1]. HbbTV is actually a conformance, certification and promotional effort as well as an industry standard adding interactive functionalities to TV by aggregating and profiling existing standards such as CEA for CE-HTML, MPEG for video, audio and transport, DVB for signaling, OIPF (Open IPTV Forum) for APIs to manage the link between the interactive HTML and the various subsystems, etc. It builds on well-known and proven technologies, shared with most of the Web and the mobile environment, such as HTML, CSS and JavaScript to create interactive applications.

HbbTV was created from the premise that interactivity will be useful for the future of TV as we know it. Almost everybody infers that today's TV should become interactive, and this means that the TV screen should display interactive user interfaces and that the current TV command device, i.e. the remote control, should be the input device for the interaction.

There are many examples of attempts to improve on these assumptions, the most recent being the MOVL connect platform [7], which is a proprietary platform for developing multi-user and multi-device applications targeted at connected TVs and mobile devices, Apple's AirPlay [8], which allows controlling media playback on AirPlay-enabled devices remotely from e.g. a MacBook or an iPhone. There were also many versions of improved remote controls, including some with pointer technologies and others with touchscreens.

We do not believe in interactivity limited to the TV screen controllable only with the TV remote. And yet, we believe that HbbTV is a key enabler for an ecosystem of applications and services in the home around the TV. But even if the interactivity is TV-channel-related and received by the TV, it should not be confined to the usage of the TV screen and interaction through the remote control. More enabling technologies are needed.

There is currently intense activity around "multi-screen" scenarios [7][8][10][12][13], i.e. the possible cooperation between the TV and other connected devices in the home: ***smart phones***, representatives of a class of personal interactive mobile devices, ***tablets***, representatives of another class of family/shared mobile devices, ***desktop computers***, a class of non-mobile devices, etc. The activity has multiple forms:

- Developers try to use existing tools to provide services around TV, with varying degrees of synchronization; applications are either adding to the TV experience and focusing the viewer back on to the TV or trying to steal the viewer attention away from the TV towards a more personal, channel-independent experience.
- Larger industry actors are investigating which products or standards would help create an ecosystem around TV in the home, possibly on the basis of existing Web technologies.

With Coltram we aim to create a model for modular applications that are designed as constellations of smaller collaborating components. These components may be individually distributed across devices at runtime on request of the user, without losing their current state. Coltram will provide a location-transparent communication mechanism that lets the system handle the complex task of managing cross-device communication and synchronization between components, so that the developer may focus on his application. The scope of Coltram will go beyond using companion devices as remote controls for the TV [11] and not be confined to static distribution configurations that do not satisfy the highly dynamic availability of additional devices in the home. Further, the employment of Web technologies will enable cross-platform interoperability for Coltram-enabled applications and lower the barrier for developers and manufacturers to adopt this new approach to applications that are suited for multi-screen scenarios by design.

First we describe the multi-screen challenges from a service-oriented perspective (Section 2) and introduce the underlying concepts of the proposed Coltram architecture (Section 2.1). Then we will describe four scenarios for future multi-screen enabled TV apps (Section 3), from which we then derive requirements (Section 4) and challenges (Section 5) that need to be addressed to realize these scenarios. We conclude with a brief presentation of the Coltram components and concepts (Section 6).

2 A Service Point of View

Today, any service to the user is either implemented as a (native) software application, running on a particular device and OS, or if implemented as a Web application, then as one document (view/session) at a time and coming from the Internet (online).

The home environment is constituted of more and more devices with extremely varying characteristics. It is heterogeneous, whatever the point of view: devices have a screen or not, large or

small, have input capabilities or not, are personal and public, have computing power or not… Such an environment is, in a sense, multi-centralized. Specific features of a device implicitly make this device a server for these specific features. Any service requiring these features must connect to that device or to a service exported by that device. Parts of the service may be executed on a choice of devices, and as such, may need to be adapted. A service, on the other hand, should be easy to use. Detailed management of its mapping onto devices should be transparent to the user. More: if the availability of devices changes, then the service should be transparently reconfigured, without loss of execution context.

We believe in a new model of service, which is a collaboration of multiple applications running on multiple devices, including discovered devices, each application not tied to one device and able to move to another device without losing state, and coming from Internet or broadcast or the local network.

And yet, because native software applications will stay as an important component of the home architecture, the new service model should allow the seamless integration of native and Web components, and the seamless switching between native and Web components.

The challenge is to allow the convergence of Internet, the mobile world, the home media (TV, set top boxes, gaming consoles, etc.) world within the home network by providing this new service model, lowering the cost of designing and using these services for all actors.

In the WWW, services are implemented with Flash or Web standards, with one server for the application and possibly others for media. The situation is similar on mobiles, with less reliance on Flash but more dedicated apps and slightly different Web standards. Emerging standards such as HbbTV in the TV domain are pointing toward a similar model (without Flash or dedicated apps).

The home network however is quite different: there are many different devices, many of which are not always available; communications are not naturally centralized, but device-to-device; but Web standards are a common base for the design of services if peer-to-peer discovery and communication interfaces can be added.

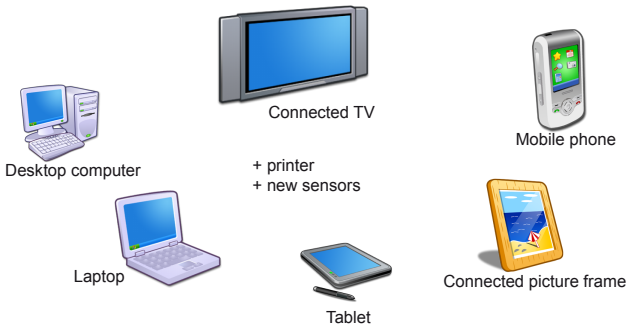


Fig. 1. Typical home network

Services on Internet or on mobiles are designed to run on one single, relatively powerful device, which has computing power and a screen/speakers and interaction capability as well as possible extra features such as a camera.

In the home network, capabilities required by a service are rarely concentrated on a single device: usually there are multiple choices for each requirement, and the best available choice may

change from moment to moment. Concepts from various origins need to be merged: ubiquitous/pervasive computing, where an application can run on any device; distributed computing, where a service may be implemented as several communicating applications running on different devices; and Web applications, where the service is implemented as a document presented in a Web browser (with scripting).

2.1 Enabling distributed TV apps

In this paper, we propose *Coltram*, a platform for distributed services in home networks and beyond. We see Coltram as an enabler for many innovative use cases in future TV scenarios. The Coltram infrastructure builds on top of the *webinos* [6] framework, a federated Web runtime that offers a common set of APIs to allow applications easy access to multi-user cross-service cross-device functionality in an open, yet controlled manner, to which Coltram will add a flexible model for distributed collaborative services and a service platform allowing for

- advanced, context-aided discovery of devices and services,
- seamless migration and ad-hoc distribution of atomic or composite applications between different devices,
- dynamic UI adaption for applications that were moved between devices with e.g. different display sizes.

Last but not least, Coltram will provide application authoring tools which help developers to easily make use of the mentioned Coltram platform features.

Pervasive computing addressed the various problems of mapping applications onto heterogeneous sets of devices. Service adaptation was studied in multiple contexts. However, there are no solutions yet addressing the Coltram combination of features: services implemented as constellations of communicating Web applications; Web applications can be split so that a device-dependent part can stay on that device; devices come and go, and applications may need to be moved during the life of the service while keeping the service state.

3 SCENARIOS

The following sections describe scenarios in the connected TV context which can be realized with the proposed Coltram architecture but not with the HbbTV architecture as it is. We further outline the corresponding challenges that need to be addressed in order to realize each scenario.

3.1 One on One

The general scenario is that of an application received by the TV, and requested for use by one viewer among many.

If the application does not require the use of TV APIs, then the application can be sent as is to the personal device of the viewer. Else, the application needs to be split into two communicating parts: the user interface should be sent to the personal device of the viewer, and the business logic including calls to the TV API should be executed on the TV itself. Communication between the two parts should be supported by the service platform.

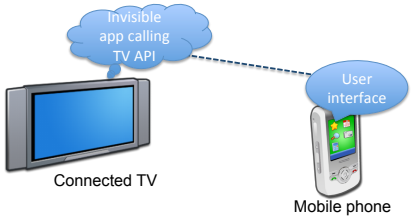


Fig. 2. One on One Scenario

A typical example is the electronic program guide (EPG) associated with the current channel. At this time, with HbbTV, the EPG could be part of the HbbTV application accessible through the red button (either directly, or as a sub-application). If there are multiple viewers, there is a good chance that only one viewer is actually interested in the EPG information. If the EPG information is displayed on the TV screen, only one viewer is satisfied, and the others are actually disturbed. If there is another screen available, it will be optimal to move the EPG display to that other screen.

To implement this scenario, there are many possibilities. HbbTV needs to be extended to be able to deal with this scenario, but how big the extension should be will be discussed below. Here is a list of features of this scenario that are not currently solved within HbbTV:

- the TV should be aware of the other device, i.e. discovery through UPnP or Bonjour;
- the application on the TV should be able to communicate with the application on the other device, to pass information that HbbTV application usually request by calling an JS API;
- the TV screen resolution and aspect ratio are not the same as that of the screen of the other device, so some layout adaptation needs to happen, or possibly a different layout needs to be provided;
- the interaction on the other device could be touchscreen (definitely not remote control), so user interaction should be adapted;
- the application on the TV and the application on the other device should be synchronized

3.2 Play Along TV Game

The general scenario is that of a TV game, with an application that allows viewers to play along. With HbbTV as it stands now, either there is just one local player, or all viewers have to collaborate as one player.

For such a game, the central application running on the TV would be designed to detect all local devices that would allow a viewer to play along, and to propose to all these devices to load the appropriate version of a play along application. Each version of the play along application would have the appropriate user input (touchscreen on a tablet and some phones, keyboard on some phones, etc.), and be synchronized with the TV program and the application running on the TV. Results for each player would be communicated by the play along application to the central application. The central application would possibly not need a user interface, just communication interfaces to all play along applications, and a connection to send results back to the server, e.g. to share high scores with all players in the country.

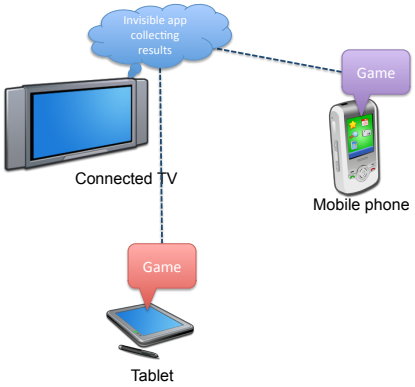


Fig. 3. Play Along TV Game

To implement this scenario, missing features are very close to those for the previous scenario:

- the TV should be able to discover many devices, not just one; UPnP and Bonjour provide that;
- the TV application should be able to communicate with multiple applications running on other devices, which makes the application more complex but is the same requirement;
- same requirement for screen adaptation, interaction adaptation and update propagation.

3.3 Device "Unavailable"

The viewer has engaged with the "One on one" scenario with the family tablet. During that usage, the tablet becomes unavailable, e.g. a kid "preempted" the tablet to play some new game. Then there are two possibilities for the viewer to go on with the service:

1. Before the tablet is preempted, the viewer requests a list of possible devices on which the service can be migrated, chooses one, and the service is restarted, keeping the service context, on e.g. the mobile phone of the viewer. We call this the application push migration scenario.



Fig. 4. Push Migration

2. From the viewer personal phone, a list of in-use services is requested, the viewer finds the one running now in background on the tablet, and requests its migration to the phone. We call this the application pull migration scenario.



Fig. 5. Pull Migration

For either push or pull migration of applications, the missing features are:

- the devices, or the browsers on the devices, need to recognize each other as "hosts of movable applications"; so each device needs to advertise itself as provider of that service and implement pull request, push request and the details of application migration;
- an application should be easy to move, so all of its resources should be packaged together and our suggestion is to adopt W3C Widgets-style packaging 5 which provided a simple organization of zip packaging for a web application;
- the execution state of an application should be serialized and sent with the application.

3.4 Service Mobility

In the fourth proposed scenario, a service is made available to a device residing on a different network: Alice and her husband Bob watch a movie on their HbbTV set. Bob decides to leave for work, but would prefer not to miss the rest of the movie. Hence, Alice authorizes Bob's tablet to receive the stream via their home network, using a modal menu on her TV, which automatically suggests nearby devices that are able to play video. Bob chooses to accept the offer by selecting the respective option in a dialog window shown on his tablet. Later in the bus on his way to work, Bob continues watching the stream, served directly from their home network.

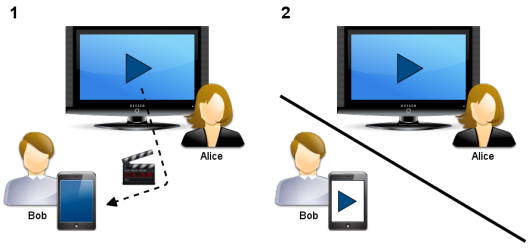


Fig. 6. Service Mobility

In order to enable this scenario, the Coltram platform must deal with several challenging aspects:

- Offering transparent access to local discovery mechanisms. The discovery results must be filterable by device capabilities and other constraints.
- Providing media content to remote devices. In particular, QoE and adaptive streaming mechanisms need to be put in place to ensure that home networks can handle the traffic load. Individual components might have to be placed in the Cloud.
- Political questions concerning content ownership and DRM need to be addressed in the implementation. For example, it should not be possible to stream copyright-protected content to arbitrary, alien devices without honoring restrictions imposed by copyright holders.

4 REQUIREMENTS

A generic requirement for all of the scenarios above is the automatic discovery of devices. It is not realistic to ask for the general public to declare all of their devices on their home network.

It should be possible to build text-based and graphics-based service interfaces. This should be non-controversial: use HTML+CSS for text apps, SVG for graphics apps.

Services should be available as widgets, compatible to W3C Widgets. The extra technology to make HbbTV compatible with W3C Widgets ("Packaging and Configuration" [5] and "Widget Interface" [9]) is comparatively very small.

A service should be available on any device, even if the service is about e.g. TV. A user should be able to vote on her TV (with a remote control), but also on her mobile, especially if there are more viewers. One way to achieve this is by solving another requirement: it should be possible to create a service as a constellation of cooperating elements (widgets) running of possibly multiple devices (distributed services). In the example of the TV vote, there should be one widget on each voting phone and one widget on the TV to collect and send the votes back. This implies communication between widgets.

It should be possible to build services with data from any origin, broadband or broadcast or local. The user should not be aware of these differences, and it should be possible to implement innovative business models on the complementarities between delivery modes. A widget should work the same way whether it is used during the broadcast, or when used with catch-up TV.

The user should also be able to switch devices at any time without losing the context of the current service, for example, for nomadic use, for a better and more organized user interface, and for extending the functionality of running applications through external device capabilities.

For example, the program guide may be easier to browse on a tablet, even if the data still comes from the TV.

To be able to deal with a dynamic set of available devices, the environment should provide discovery and service protocols. Discovery is necessary to deal with dynamic networks, to remove the need for the author of a service to know the address of a subservice in advance, as well as to deal with devices coming in or going out of the home network. You should be able to vote even if you are at a friend's home.

A service should behave the same, whether it is provided by hardware, by a native application, or by a widget. This is to be inclusive about the type of device that can be integrated in the home environment.

It should be possible to build services on top of other services, so as to easily create mash-ups, as well as multiple interfaces for the same service.

Even if we may have technology preferences, it should be possible to replace any chosen technology by an equivalent, without breaking the service model: replace a discovery protocol with another, a scripting language by another, a format by another, etc. The service model needs to be future-proof.

5 CHALLENGES

Although our every day's devices are getting more and more connected due to the latest technological achievements - almost every home has a wireless and wired network installation, TV sets and set-top boxes are connected to the Internet, mobile devices have WLAN and 3G data flat rates - the devices can talk to each other but there is no common application model yet. Some standards like UPnP addressing the interworking are partially used for interconnecting home devices for multimedia services, but often rely on communicating devices being on the same local network, which is typically not the case between, for example, the TV and the phone a user owns, or the tablet device of a visiting friend.

A real breakthrough is still missing! We need to address exactly this missing step in the ongoing convergence of home networks, TV, and Internet: developers of applications need to be supported to reflect the advantages of this convergence. They should not be forced to develop against all the different characteristics of devices (features, screen-size, I/O support, resources, etc.). Instead there is a clear need for a standardized, open cross-device platform, on which application run and can be automatically (or with the best possible support of the underlying platform) adapted to the respective execution context.

This goal is the innovating characteristic of our project. The pre-conditions are complex and requirements very high, as especially in the consumer electronics domain there is a very high claim for very good usability and user experience. Any solution not fully satisfying these claims will hardly be successful on the market.

On the other hand, the demand for this type of support for application developers is huge. With the introduction of Web 2.0 related technologies (JS, JSON, AJAX, etc.), which led the Web to become an interactive application platform rather than only a document exchange and presentation service. This led to an explosion of the Web and the content (including apps) in it. Today, it is the most dominant application platform used.

Coltram will support application developers in the consistent development of their applications across devices. This will also include the support for distributed execution / deployment; i.e. an "application" can be used on more than one device at the same time, where the different installations will synchronize and coordinate each. In this way the same application might represent a video playback widget on a TV device and a UI control on a tablet.

These different but adaptive representations of the same services need to be reflected by Coltram and the developer needs to be supported to deal with its complexity. Beside the technical aspects of the actual (distributed, orchestrated) execution of the application, the authoring aspect during the development of an application is a major objective and innovation of Coltram.

6 IMPLEMENTATION

In the previous sections we have described the requirements and challenges for innovative multi-screen scenarios in which consumers have the freedom to use any available device to serve as additional screen and even to control applications running on e.g. the TV. Within this section we present the underlying concepts of the Coltram service model and platform and how they meet the identified requirements. Coltram targets a wide range of device types including but not limited to smart phones, tablets, HbbTV connected TV sets and computers.

6.1 Coltram Service Model

In Coltram, services will be modeled as sets of communicating applications. Each application can be dynamically moved to the "best possible device" at any point in time, if its required features do not tie it to one device. The W3C widget specification [5] defines the basic model for a communicating application, extended with discovery, service capabilities and requirements on hard- and software, reaching forward in the direction of the work of the Home Network Task Force of the W3C Web and TV interest group [10]. Developers may use this model to combine multiple communicating applications into one complex service, which may be packaged and deployed as a single component. This composite service may in turn again be use within an even more complex service.

Inter-widget communication in Coltram will be implemented using event-based publish-subscribe messaging patterns [2][3][4], allowing for loose coupling of connected services and for full location transparency from developer perspective.

6.2 Coltram Application Platform

The Coltram service platform will be built on top of webinos [6], a federated Web runtime comprising a common set of APIs to allow applications easy access to multi-user cross-service cross-device functionality. Coltram will extend and augment the underlying technologies provided by webinos to support high-level functionalities, where webinos provides the lower-level technology enablers.

Coltram will provide a device and service discovery mechanism, which extends webinos' discovery with the additional features that are defined in the Coltram service model: service capabilities and requirements on hard- or software. Devices advertise their availability and will be registered in Coltram until they are not available anymore.

When services migrate from one device to another, Coltram has to ensure that the current state of the service does not get lost. When the user triggers a complete migration of any service, Coltram serializes the local state, pushes it to the new device und deserializes the data. For unexpected shutdowns of a device, e.g. when the battery died, this is not easily possible, although there are workarounds, ranging from storing periodical snapshots to continuous synchronization of the service's local state to a server. For services that are running distributed on different devices, Coltram will provide an efficient real-time synchronization mechanism to developers, which allows quickly propagation of changes between all of the distributed parts of the service.

6.3 Coltram Authoring Tools

The Coltram model induces extra complexity for the author, on top of the complexity of widget authoring. To compensate this, one part of the project is dedicated to making the creation of Coltram services easier. The following authoring aspects will be considered: authoring services consisting of a constellation of communicating widgets; authoring for hybrid delivery, i.e. using synchronized media coming from multiple delivery mechanisms, such as broadcast and broadband; authoring of applications that are mostly script.

The emphasis will be on extensible, collaborative tools, making it easy to reuse previous designs, and with facets for different authoring expertise levels, including one facet for a wide audience.

7 CONCLUSION

In this paper, we have presented several ideas for such use cases and outlined how Coltram will address the most important requirements that must be met. We strongly believe that Coltram will enable developing innovative and compelling scenarios for future TV apps that will increase user engagement and interaction.

Results from on-going research activities in Coltram regarding local and remote device discovery, service discovery, concepts for adaptive cross-device user interfaces and seamless synchronization of distributed services will be contributed to the standardization efforts within OIPF, HbbTV and W3C.

8 REFERENCES

1. HbbTV, ETSI Technical Specification 102 796, “Hybrid Broadcast Broadband TV”
2. C. Concolato, J. C. Dufourd, J. Le Feuvre, K. Park and J. Song, “Communicating and migratable interactive multimedia documents”. Multimedia Tools and Applications, May 2011 [DOI 10.1007/s11042-011-0805-2].
3. B. Hoisl, H. Drachsler and C. Waglechner, “User-tailored Inter-Widget Communication Extending the Shared Data Interface for the Apache Wookie Engine”. Proceedings of ICL 2010, Hasselt, Belgium
4. S. Sire, M. Paquier, A. Vagner, J. Bogaerts, “A Messaging API for Inter-Widget communication”. Proceedings of WWW 2009, Madrid, Spain
5. W3C, “Widget Packaging and XML Configuration”, September 2011, http://www.w3.org/TR/widgets [online; accessed 28/03/2012]
6. webinos, “webinos whitepaper”. June 2011
7. MOVL, http://movl.com [online; accessed 28/03/2012]
8. Apple AirPlay, http://www.apple.com/de/itunes/airplay/ [online; accessed 28/03/2012]
9. W3C, “The Widget Interface”, Candidate Recommendation, December 2011, http://www.w3.org/TR/widgets-apis/ [online; accessed 30/03/2012]
10. W3C, “Requirements for Home Networking Scenarios”, December 2011, http://www.w3.org/TR/hnreq/ [online; accessed 30/03/2012]

11. J.P. Barrett, M.E. Hammond and S.J.E. Jolly, “The Universal Control API v.0.6.0”. BBC Research & Development, Research White Paper WHP 194, June 2011.
12. S. Glaser, B. Heidkamp and J. Müller, „ Next-Generation Hybrid Broadcast Broadband: D2.1 Usage Scenarios and Use Cases”, HBB-NEXT Project Report, December 2011.
13. S. Slovetskiy and J. Lindquist, “Combining TV service, Web, and Home Networking for Enhanced Multi-screen User Experience”. Ericsson Position Paper for the 3rd W3C Web and TV workshop, September 2011.

Texmix: An automatically generated news navigation portal

Morgan Brehinier[1] and Guillaume Gravier[2]

[1] INRIA Rennes,
morgan.brehinier@inria.fr ,
[2] CNRS–IRISA,
guillaume.gravier@irisa.fr

Abstract. The Texmix demonstration presents an original interface to navigate a collection of broadcast news shows, exploiting speech transcription, natural language processing and image retrieval techniques. Navigation is performed through keywords search or through time or through maps, with links automatically created either between reports to follow the story or to the Web to know more about a story, a person or a fact. Image search technology is also integrated to find portions of the collection with similar images. We also present two original features to dynamically access videos, namely dynamic summary and geotagging ...

Keywords: HTML5, hypervideo, topic segmentation, speech to text, named entities, similar images, dynamic resume, geotagging

1 Introduction

The gradual migration of television from broadcast diffusion to Internet diffusion offers tremendous possibilities for the generation of rich navigable contents. However, it also raises numerous scientific issues regarding delinearization of TV streams and content enrichment. In this demonstration, we illustrate how speech in TV news shows can be exploited for delinearization of the TV stream. In this context, delinearization consists in automatically converting a collection of video files extracted from the TV stream into a navigable portal on the Internet where users can directly access specific stories or follow their evolution in an intuitive manner. The TEXMIX demonstration illustratess the result of the delinearization processand enables user to experience various navigation modes within a collection of news shows. An illustration of the interface is provided Figure 1. The demonstration can be accessed online at http://texmix.irisa.fr.

Structuring a collection of news shows requires some level of semantic understanding of the content in order to segment shows into their successive stories and to create links between stories in the collection, or between stories and related resources on the Web. Spoken material embedded in videos, accessible by means of automatic speech recognition, is a key feature to semantic description of video contents. At IRISA/INRIA Rennes, we have developped multimedia

Fig. 1. Overview of the navigation interface



content analysis technology combining automatic speech recognition, natural language processing and information retrieval to automatically create a fully navigable news portal from a collection of video files[1].

The purpose of this demonstration, aside from exploiting technologies developed in our research group, is to provide a new experience from watching videos on the Internet. Exploiting HMTL5 features, dynamic content is provided as the video plays, and allow the watcher to have additional information and to interact during his watching, being no longer passive. We also provide new search services based on the technologies developed.

In this summary, we describe the various features of the demonstration, providing brief insight on the underlying technology.

2 Timeline navigation

An important feature to access and navigate a collection of news bulletin is time. A timeline is used to locate in time all information that you can access. It is also used to locate in time information that are displayed by positioning the cursor accordingly.

3 Link creation

News shows are made of successive items which are usually introduced by the anchor speaker and developed in some story. The first step to structuring a collection therefore consists in segmenting each show into its constituting items. This segmentation step is done using statistical topic segmentation method which detect topic changes in the transcription, inspecting the distribution of words. Topic segmentation enables to automatically generate a table of contents of each show. Futhermore, keywords are automatically extracted from the transcript to characterize each story. Exploiting the keywords and the transcripts, links to the Web and to related stories in the collection are created and displayed, thus offering navigation capabilities to follow the evolution of an event or to know more about it.

4 Named entities

Named entities, that is proper names and location names, are also automatically detected using machine learning algorithms and used to locate news items and to dynamically provide additionnal information while playing the video. For instance, a Wikipedia pop-up is displayed whenever a person's name is uttered, with a link to the corresponding wikipedia page for more information. Similarly, locations are dynamically displayed on a Google Map as they occur, thus facilitating the localization of news stories.

5 Similar images

Spoken language provides crucial information but so do images! Indeed, the same image is often used to illustrate a story throughout time. Based on our image retrieval technology[2], we included a 'similar image' search feature in the news portal, as illustrated Figure 2. In this second illustration, we search for images similar to a graphic depicting the voting intentions for the 2007 French presidential election. This results in similar graphics at different dates, making it possible to follow the evolution of the poll in time.

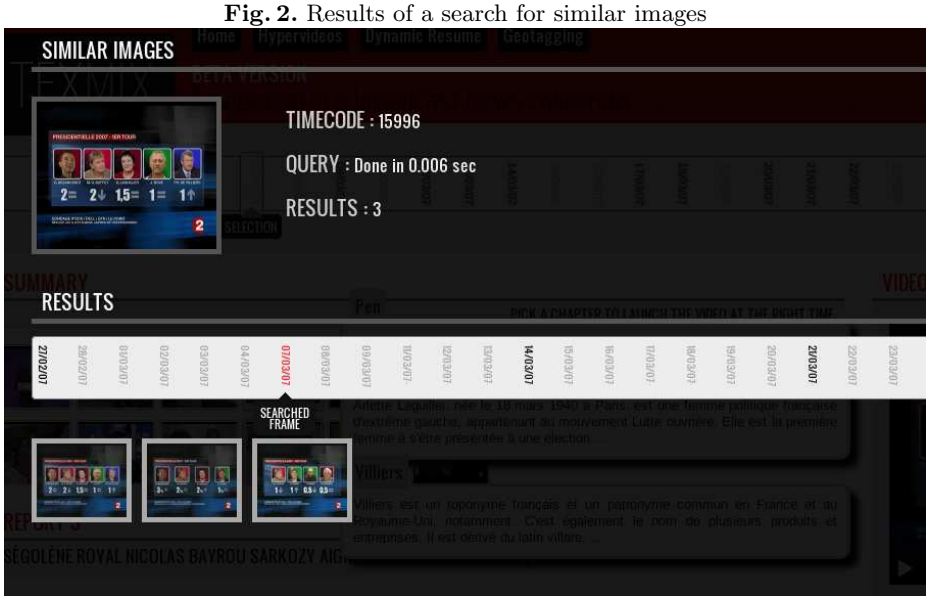
6 User services

Based on the previous technologies, we developed two services for the user.

6.1 Dynamic summary

Based on the keywords extraction technology, the dynamic summary offers the possibility to catch-up news on a specific topic in a very quick way. The user starts by selecting the period of time he is interested in, and then selects a keyword related to this period. Those two actions result in a dynamic resume of the topic, by sequencing the thirty first seconds of each story related to the keyword.

Fig. 2. Results of a search for similar images



6.2 Geotagging

In the same way 'dynamic resume' offered a navigation based on time and keywords, geotagging is a way to look for news stories related to their location. For this feature, the user starts by selecting the time period he is interested in, and markers are then displayed on a Google Map, showing where the news stories of this time period occured. The user can then zoom in or click on a marker in order to view the news bulletin related to the place he chose.

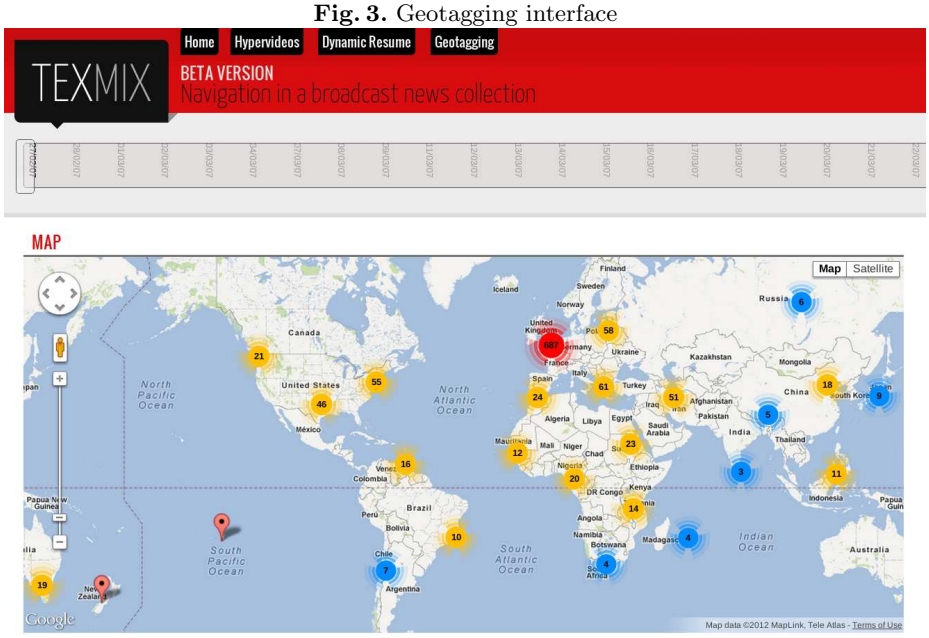
7 Technology

This interface has been developed in HTML5, CSS and Javascript. Those technologies allowed us to develop this multimedia portal without relying on proprietary software solutions. The synchronization of the contents and the videos is based on the popcorn.js library (www.popcornjs.org). Therefore, this demonstration is accessible on the Internet with any recent web-browser supporting HTML5 technologies.

8 Conclusion

This unique combination of speech and language technology and mulimedia retrieval offers countless opportunities to automatically generate Web content from TV streams, opening the door to a fully connected multimedia Web.

Fig. 3. Geotagging interface



9 Acknowledgements

This work was partially funded by the Quaero project and by EIT ICT Labs via the OpenSEM project.

References

1. Guillaume Gravier, Camille Guinaudeau, Gwenole Lecorve and Pascale Sebillot. Exploiting speech for automatic TV delinearization: From streams to cross-media semantic navigation. In Journal of Image and Video Processing, 2011, 2011:689780.
2. Herve Jegou, Mathias Douze and Cordelia Schmidt. Product quantization for nearest neighbor search. IEEE Trans. On Pattern Analysis and Machine Intelligence, 33(1):117128, 2011.

Towards the Construction of an Intelligent Platform for Interactive TV

Natasha C. Queiroz Lino , Clauirton de Albuquerque Siebra, Priscilla Kelly Machado Vieira and Omar Scher Ramalho

Centro de Informática
Federal University of Paraíba, Brazil
{natasha, clauirton}@ci.ufpb.br, {priscillakmv, omar}@lavid.ufpb.br

Abstract. *The Interactive TV (ITV) advances, such as the real time interactivity, are enabling the development of modern applications in this platform. However, several of such applications could take advantages from a convergence environment, via the use of information from other sources. The Web servers, in particular the knowledge servers, present interesting properties to the configuration of a convergence environment. This work explores such properties as a way to support convergence between the Web and ITV platforms. A prototype associated with semantic queries was implemented as a case study and also as a form to present the advantages and learned lessons of this research.*

1 Introduction

The *Knowledge as a Service* (KaaS) paradigm [1] was proposed in 2005 as a form to extend the potential of computational systems developed as a *Data as a Service* (DaaS). Two main advantages of this paradigm can be stressed. First, the models used by this paradigm are based on formal semantic representations. Second, the knowledge servers have the capacity of accessing data of different sources, instantiating their representations and generating knowledge to be delivered via intelligent process such as Data Mining algorithms [2].

This work discusses representational ways to KaaS, which enable an appropriate semantic description to data that comes from different platforms. Our main aim is to use the KaaS metaphor as a form to enable convergence among different computational platforms, such as the ITV and Web. We argue that the integration of data from different platforms can be very useful to extend the services that are running on a specific platform. A practical example of this affirmation is described in this work, where an ITV service uses information from Web via a KaaS that is semantically described in accordance with the Semantic Web concepts [3].

Knowledge as a Service (KaaS) is a service oriented computational paradigm where knowledge based services providers, or knowledge services, answer requests

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

sent by information consumers. These services are based on knowledge models that are in general expensive or impossible to be maintained by the own consumers. Figure 1 shows a conceptual view of services, according to the KaaS approach.

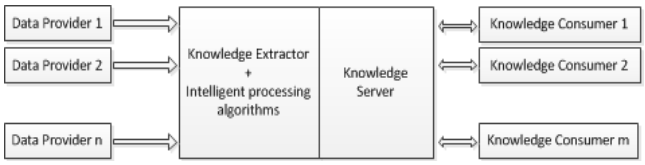


Figure 1. KaaS conceptual view [1]

It is important to highlight three essential features of KaaS, which make this computational paradigm appropriate for the configuration of convergence environments: (1) KaaSs are based on formal semantic models, for instance ontologies [4], which permit standardization of format and meaning. Thus, knowledge can be shared among participants that adopt a specific model; (2) KaaSs act like consolidation components, since they can seek information to instantiate their models, transforming that information in knowledge; (3) the intelligent processes used in Web servers, such as Data Mining algorithms, can emphasize knowledge originally implicit, or impossible to be derived from primary data sources due to the weakness and informality of their representations.

1.1 Towards Intelligent ITV and Web Convergence

One of the Knowledge TV [5] research project main goals is to configure a convergence environment between the ITV and Web platforms via a KaaS service. In Figure 2, the KaaS is represented by the Knowledge TV (KTV) Web Service. The KTV KaaS provides support for the following services: (1) a framework for data integration considering different data modeling paradigms used in the project, such as, operational, multidimensional or analytical, semantic and linked to the Linked Data [6]. This framework was build from the lessons learnt from the two prototypes developed (semantic queries and recommendation service) and aims to organize and integrate data in these different paradigms and also should be used as a guide to the new applications and services to be developed on top of the Knowledge TV platform; (2) a reference core ontology – CoreKTV [8] based on metadata standards for broadcast (Anytime TV) and multimedia content (MPEG-2, MPEG-21, etc.) in which the KTV services are based on; (3) a Knowledge Discovery in Databases (KDD) architecture to provide analytical data services, for instance, data mining processes and multidimensional queries (OLAP – On-Line Analytical Processing) that will permit computing and discovering of new and useful patterns and knowledge hidden in the data. These can be useful for other services such as the recommendation one; (4) a semantic multi-domain query service, integrated to the Linked Data methodology, that allows information retrieval and semantic enrichment via Linked Data; (5) a personalized semantic query service supported by KDD [2] that permits refined and tailored information retrieval based on users' patterns of ITV programs usage; (6) a semantic based recommendation service for recommending content on the broadcast program via a

personalized EPG (Electronic Program Guide) and also broadband web content; (7) a framework based on ontology alignment that will permit a multi-domain meta knowledge-base of domain ontologies to support the KTV KaaS services; and (8) an open access Application Program Interface (API) for interoperability with other systems, services and platforms based on the CoreKTV ontology.

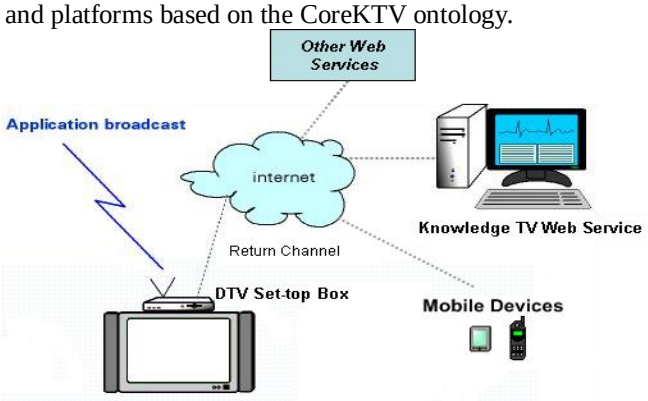


Figure 2. General Network Architecture

In Figure 3 computational modules are shown in detail. They are divided between the ITV (middleware - MD) and Web platform (KaaS server). However most of the components are located in the knowledge server. This is part of the KTV strategy to implement as less as possible resources in the ITV middleware. Thus, future middleware updates will not highly impact the system.

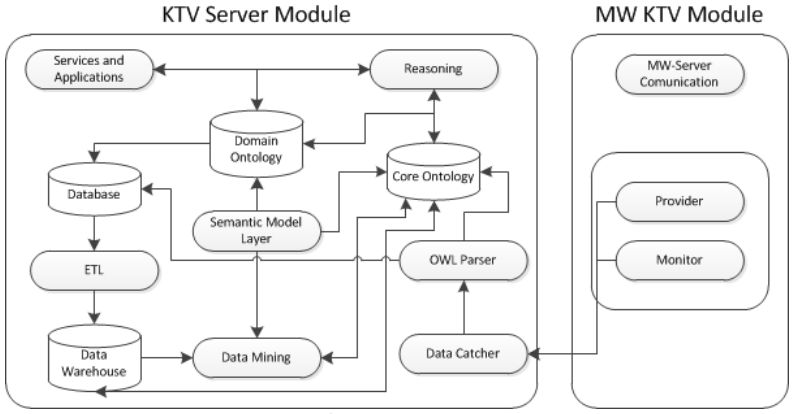


Figure 3. Modules of the convergence environment

In the middleware side, there are two components or agents: *Monitor*, whose task is to obtain TV user behavior by means of its interactions with TV (programs and applications) and, *Provider* that accounts for obtaining metadata using the middleware information content provider component.

In the server module, the following components are highlighted: *Data Catcher* that accounts for getting information obtained from the middleware and promoting a verification phase and cleaning checking before sending the information to the *OWL Parser*. The aim of this parser is to re-organize data under a single semantic format,

OWL, permitting sharing, re-use and grouping of concepts in the construction of *Domain Ontologies* and the *Core Ontology* (CoreKTV), whose formal specification is on the *Semantic Model Layer*. This layer accounts for maintaining a knowledge base of concepts and properties regarding multimedia content in ITV that can be used, shared and extended to any specific domain or service implemented on the KTV platform.

Information in the knowledge server is maintained according to the Domain Ontology models. For instance, the scenarios described in the next sections are dealing with the movies domain ontology. On this data is applied a knowledge discovery in databases process, including an *ETL* (Extraction, Transaction and Loading) pre-process so that the system is able to build a *Data Warehouse* [2]. *Data Mining* algorithms are applied on the Data Warehouse to extract important patterns and knowledge, in accordance with the desired knowledge that a specific service or application needs. For example, for the movies recommendation service, one data mining task is to derive association rules and/or user clusters that have similar preferences in movies.

2 The Semantic Queries Service

The Knowledge TV can be defined as an open platform to build intelligent services in convergence environments. To instantiate and validate that statement we developed a semantic query service on the KTV platform. The semantic queries service supports the extension of key words queries, so that the semantic of an initial query can be expanded and used during the retrieval of answers. The architecture of this service is based on Linked Data and developed in accordance with its principles [6]. The following steps are carried out to execute the service: (1) Clients send queries from applications running on the DTV or Web platforms. From the DTV platform, the query could be requested via a remote control button, so that users just need to press this button while they are watching a movie. The idea is to obtain more information related to this movie; (2) The keywords of the query are instantiated via metadata that is sent via broadcast by the content providers. This instantiation is carried out according to a pre-defined metadata. The provider agent collects this information and sends it to the KTV server in the form of a CoreKTV ontology instance; (3) A local search is carried out in the KTV server database. If the required data is locally stored, they are returned to users. Otherwise, a search is carried out in the Linked Data (DBPedia, LinkedMDB, etc.); and (4) Data from Linked Data are locally stored and, after that, sent to users.

The information from metadata (SI tables), in the ITV platform, is generally quite limited. For example, providers generally only send the name and gender of movies. Thus, the semantic extension, or semantic enrichment, of this information is relevant to obtain a result that attends the users' expectations.

3 Preliminary Experiments and Results

We are simulating an ITV video flow related to the "Time Out" movie. During this movie, users can press the remote control blue button, which accounts for performing

a semantic query about the current program. This is specified in a simple middleware embedded application. The idea of this application is to bring additional information about the current program, in this case, the "Time Out" movie.
When the button is pressed, the *Provider* module (see Figure 3) sends all the current content from the SI table to the KTV server. The *DataCatcher* receives this flow and calls the *OWL Parser*, which organizes the income data in a unique format (OWL), in accordance with the *Core Ontology*. Then, the *Application and Services* module identifies this scenario and requests, from a specific semantic server about movies, information to complete the *Domain Ontology* attributes. This ontology, specified to movies in this application, has the following attributes: director, main actor, actress, awards, synopsis, year of production and etc. All this information is formatted, according to the ontology, and returned to the middleware, where a specific application accounts for showing the information to users.

We are using two approaches to evaluate this application. The fist evaluation is a simple questionnaire about satisfaction and usability. We have presented three methods to obtain information about a current program to users: (1) SI table query: in this case only the SI information is presented. Users reported the minimal additional information (generally the norm compulsory information), which is dependent of providers. This means, providers can choose what they want to insert in the video flow; (2) Simple Web query: in this case the keyword "Time out" is used without a semantic treatment. Users reported the significant amount of garbage (low precision), which were returned by the research. This is easily verified if we carry out a simple google search using "Time out" as keyword. Note that the results are not related to a movie; (3) KTV Semantic query: using the KTV approach, users reported the good quality of the information and also its organization. This is easy to control, once the ontology forces this organization. However, if some information, which is expected by users, is not represented in the ontology, it will not be displayed. In future versions, we could think about an user-centered and customized extension of ontology, so that particular expectations of users can be filled.

The second evaluation approach considers a more formal framework, based on the performance and correctness measures of the Information Retrieval (IR) [8] area. In this context, two measures are being used: precision and recall. This approach is going to be described in more details in future publications.

4 Conclusion and Future Works

This paper discussed a computing paradigm, called *Knowledge as a Service*, as an appropriate way to provide convergence among different computational platforms. In particular, we focused on convergence between the ITV and Web platforms. We explored the high-level framework of this paradigm, fulfilling its modules with components of the KTV architecture. The main aim was to provide semantic processes and representations, such as ontologies, to represent ITV information and also to expand this representation via knowledge from a KaaS server (KTV server). A semantic queries service is discussed to exemplify the potential use of this technology. Our future

works are mainly directed to implement the remainder modules of our architecture, so that we have a more robust prototype to demonstrations. Currently, about half of the project is implemented, so that we have some limitations to create evaluation scenarios. The *Monitor* module, for example, is an important component in ongoing implementation. It will enable the capture of the historical behavior of users, so that we can provide real recommendations based on users profile.

5 References

1. Xu, S., and Zhang, W. 2005. Knowledge as a Service and Knowledge Breaching. *Proceedings of 2005 IEEE International Conference on Services Computing*, vol. 1, pp: 87-94, Orlando, Florida, USA.
2. Han, J., and Kamber, M. 2006. *Data Mining: Concepts and Techniques*, 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers.
3. Shadbolt, N., Berners-Lee, T., and Hall, W. 2006. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21 (3). pp. 96-101.
4. Gruber, T. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, vol 5, pp: 199-220.
5. Lino, N., Araújo, J., Anabuki, D., Patrício Junior, J., Batista, M., Nóbrega, R., Amaro, M., Siebra, C., and Souza Filho, G. 2011. Knowledge TV. *Proceedings of the 9th EuroITV Conference* (EuroITV 2011), Lisbon, Portugal.
6. Berners-Lee, T. 2009. Linked Data, Design Issues. Available in: http://www.w3.org/DesignIssues/LinkedData.html
7. Araujo, J. 2011. CoreKTV - Uma infraestrutura baseada em conhecimento para TV Digital Interativa: um estudo de caso para o middleware Ginga. *Master of Science Dissertation at Informatics Post-Graduation Program*, Informatics Department, Federal University of Paraiba (UFPB).
8. Singhal, A. 2001. Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24(4): 35-43.
9. N. Gkalelis, V. Mezaris, I. Kompatsiaris, "Automatic event-based indexing ofmultimedia content using a joint content-event model", Proc. ACM Multimedia 2010, Events in MultiMedia Workshop (EiMM10), Firenze, Italy, October 2010.
10. N. Gkalelis, V. Mezaris, I. Kompatsiaris, "A joint content-event model for event-centric multimedia indexing", Proc. Fourth IEEE International Conference on Semantic Computing (ICSC 2010), Pittsburgh, PA, USA, September 2010, pp. 79-84.
11. Soares, L., Rodrigues, R., and Moreno, M. 2007. Ginga-NCL: the Declarative Environment of the Brazilian Digital TV System. Journal of the Brazilian Computer Society. 12, 37-46.
12. Souza Filho, G., Leite, L., and Batista, C. 2007. Ginga-J: The Procedural Middleware for the Brazilian Digital TV System. Journal of the Brazilian Computer Society. 2007, 12, 47-56.

A study of interaction modalities for a TV based interactive multimedia system

Jože Guna
University of Ljubljana
Tržaška cesta 25
1000 Ljubljana
+386 1 47 68 116
joze.guna@ltfe.org

Emilija Stojmenova
Iskratel, Ltd., Kranj
Ljubljanska cesta 24a
4000 Kranj, Slovenia
+386 4 20 72 154
stojmenova@iskratel.si

David Geerts
3rd author's affiliation
1st line of address
3000 Leuven, Belgium
+32 16 32 31 95
david.geerts@soc.kuleuven.be

Andrej Kos
University of Ljubljana
1st line of address
1000 Ljubljana
+386 1 47 68 101
andrej.kos@fe.uni-lj.si

Matevž Pogačnik
University of Ljubljana
1st line of address
1000 Ljubljana
+386 1 47 68 101
matevz.pogacnik@fe.uni-lj.si

ABSTRACT
The problem of effective, user friendly and intuitive human-machine communication in home multimedia rich environments is becoming difficult and is challenging due to the increased complexity of different interfaces, great diversity of end-user terminal devices and especially rich content and media extensiveness which can potentially lead to the information overflow. With this in mind, we present the results of a user evaluation study comparing a standard unimodal television remote controller to the multimodal and simple to use WiiMote controller including tactile gesture and voice command inputs, in order to alleviate this problem. Both modalities were tested on a commercial IPTV system "INNBOX" based on the XBMC multimedia platform. Standard methodology including SUS, AttrakDiff and think-aloud methods were used. The results clearly show that users preferred the multimodal approach (82 average SUS score vs. 75 average SUS score), even though the system was not flawless. The AttrakDiff results and user comments are similarly positive as well. The study results show that the addition of voice and alternative navigation modalities significantly improve the overall user experience.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology, Graphical user interfaces (GUI), Haptic I/O, Input devices and strategies, Voice I/O;

General Terms
Design, Experimentation, Human Factors.

Keywords
Interaction Modalities, Tactile Gesture, Voice Control, Home Multimedia System, Evaluation, HCI.

1. INTRODUCTION
Recent advancements in the field of interactive television (iTV) have resulted in the enrichment of television experience in a number of ways. The technological development has changed TV sets into Internet enabled devices providing multimedia applications and services to the end users. Consequently, the TV sets are no longer used only for watching linear TV programs. Instead, they allow end users to connect to online social platforms and video sites, play games and access their home multimedia content (pictures, videos, music) [1]. In addition, the way users control their TV sets has changed drastically. The era of big remote controllers with dozens of buttons and controls seems to be in decline, which is a reasonable change due to the poor user experience they provide. Nowadays, some TV sets and set top boxes come with remote controllers, which have only a few buttons while some have programmable touch screens allowing end users to customize their remote (e.g. Philips Pronto, Logitech Harmony, Universal MX-900, etc.) Following the rise of smart phone and tablet sales, a number of solutions support remote control of TV sets via Android and iOS applications (e.g. Philips 6000+ series, Netgem set top box N8000 series, some Samsung models, etc.), which, in addition to mapping the remote controller buttons on the device's screen, give an added value to browsing through the Video on Demand (VoD) catalogues, Electronic Program Guides (EPGs) and during usage of the social applications. These solutions are better known as "second screen" applications. TV set manufacturers are integrating even more advanced interaction modalities using contact-less and controller-less modalities (e.g. Google TV, Samsung), originally known from the world of gaming consoles - the best known example is the Microsoft Kinect sensor device. These solutions allow for the control of the TV systems without any remote, using only hand gestures and voice commands. It is apparent that this field is advancing rapidly and those new solutions are gaining ground almost on a daily basis.

The idea behind our research was to evaluate the added value of new interaction modalities such as voice commands, applied to a commercial IPTV system "INNBOX" developed by the Iskratel Company [2]. The production version of the INNBOX system supports standard IPTV functionality such as linear TV and VoD, running of applications providing EPG and weather information and access to typical home multimedia content (pictures, videos and music). The set top box comes with a standard unimodal remote controller with directional buttons. Our hypothesis was that the inclusion of voice commands should improve the user experience for the most frequent navigational scenarios such as navigation between main menus (video, music, pictures, linear

TV, weather info, etc.). The hypothesis was tested on 14 end users, who followed the same usage scenario on two versions of the system, with and without the voice control.

The rest of the paper is organized as follows: related work is presented in section 2, interaction modalities, a description of the system setup and belonging architecture in section 3, experiment setup and evaluation procedures in section 4, evaluation study results in section 5, while discussion, key conclusions and future work references are drawn in sections 6 and 7 respectively.

2. RELATED WORK

In general, a modality in terms of HCI designates a method of communication between the human and the computer and is usually bidirectional in nature. Output modalities, such as vision, hearing or haptic modalities coincide with human senses and provide a sense through which humans receive the output of the computer. Input modalities allow the computer to receive the input from the human using various sensors and other devices, such as keyboard, mouse, accelerometer, camera, microphone, etc.

To provide a better user experience in human-computer interactions, multiple modalities are combined and a new class of interfaces emerges, called "multimodal interfaces". A definition of the multimodal interfaces is given in [3] by Sharon Oviatt, who states: "*Multimodal systems process two or more combined user input modes — such as speech, pen, touch, manual gestures, gaze, and head and body movements — in a coordinated manner with multimedia system output.*" Compared to traditional keyboard and mouse interface or a classical remote controller device, multimodal interfaces provide flexible use of input modes and allow the users a choice of which modality or their combinations to use and when [4]. Multimodal interfaces are thus perceived as easier to use and more accessible [5], provide better performance and robustness of interaction [3] and are suited for critical industrial [6] and clinical use [7] and applications. Results of the study in [8] show that speech input modality can be successfully used for enhancing accessibility in telemedicine services for the iTV. Voice input can be successfully utilized in a multimodal form-filling system, as shown in [9]. The study reports that with practice, users learn to develop interaction patterns that ensure reliable and efficient interaction, resulting in decreased dialogue duration and more user satisfaction. Furthermore, multimodal interfaces are also especially suited for entertainment and educational use, particularly in combination with virtual environments, such as Second Life as reported in study [10].

In the area of multimodal interfaces specifically applied to the interactive TV and multimedia systems we find a study [11], where the authors investigate the benefits of introducing speech and connected dialogue for iTV interaction by a speech-based natural language interface. The results show that study participants were positive toward spoken input modality and about the fact that the information load was carried by the system rather than by themselves. The presented system allowed for an ambitious goal of providing the natural language interface for accessing TV program related information. User evaluation was diverse, but included only five participants.

A study in [12] presents a concept of a speech enabled remote controller for television. The remote controller is simple in design and is equipped with a built-in microphone and press-talk switch. The authors focus mainly on the remote controller and speech recognition system and less on the TV multimedia interface itself. Chosen speech recognition trigger strategy is "press to talk" and not continuous.

A relevant study was conducted in [13] where the authors present results of a study comparing two remote controls (unimodal vs. multimodal) for a linear TV and VOD movie entertainment system. The results show that multimodality improves user experience despite clear differences in usability. Furthermore it was shown that multimodal interfaces are actually providing universal access. The different user groups showed slight differences when interacting with the multimodal remote control while older users seemed to benefit from multimodality.

In light of this fact, age and especially previous exposure to technology is relevant which is evident from a study [14] of this phenomenon. The authors focus on the technology and its acceptance taking into account the aspects of usability and accessibility by the elderly while also providing usability metrics and insights gained when confronting the elderly people with technology in general.

3. SYSTEM SETUP

3.1 Hypothesis

We state and evaluate the following hypothesis:

"*Multimodal user interface using voice commands and simple button remote control and tactile gesture based navigation modalities will be perceived as more intuitive and easier to use in comparison to a standard unimodal remote control interface.*"

3.2 Interaction modalities

In our experiment we made a comparison of two interaction modality setups:

- Interaction using a standard remote controller
- Interaction using a simplified remote controller, voice commands and simple tactile gestures

The standard remote controller (Figure 1-top) had directional buttons, a "HOME" button, shortcut buttons to video, music, picture and linear TV menus, playback related buttons (PLAY, PAUSE, FFWD, etc.), color buttons and ten buttons for text input.

The simplified remote controller was a programmable Nintendo WiiMote [15] remote (Figure 1-bottom) with directional buttons, an "OK" button, a "BACK" button and a "HOME" button. This controller was programmed in such a way that it allowed for gesture based commands (shake up and down), which resulted in "PAGE-UP" and "PAGE-DOWN" commands, allowing the user to scroll faster through the menus. Gestures swipe left and right were used for navigating between previous and next media items. In addition, we integrated the following voice commands: "HOME", "MUSIC", "VIDEO", "WEATHER", "PICTURES", "OK", "LEFT", "RIGHT", "UP" and "DOWN". Voice recognition was continuous and running without any special user interventions.

The selection of the Nintendo Wii remote in this setup was based on the fact that it is much simpler than the standard remote, which has a lot of buttons. Some of the missing buttons like "PLAY" and "PAUSE" for example, were replaced with contextual interpretation of the existing commands. For example, if the user wanted to pause/play the video playback he/she had to push the "OK" button on the remote. At the same time this button had the "OK/SELECT" functionality in other menus.



Figure 1. Traditional remote controller (top) and multimodal WiiMote controller (bottom).

3.3 System architecture

The TV multimedia system used in our experiment was a slightly modified prototype of a commercial IPTV system INNBOX, which is based on the XBMC multimedia platform [16]. The choice of XBMC was beneficial for inclusion of additional interaction modalities, since XBMC is an open platform, which allows for modifications and customizations. The INNBOX user interface skin was modified in such a way that it was the same for both setups. Thus we were able to compare the interaction modalities without the influence of the GUI specifics.

The experimental setups were installed on a standard laptop (Intel i5 CPU, 4GB RAM). The INNBOX setup using the unimodal remote controller was running on the Linux platform, while the voice enhanced setup was running on the Windows XP OS and was using a customized version of the XBMC multimedia system. The voice recognition software was written in C# using the Microsoft Kinect and Speech SDK [17] and Microsoft Kinect sensor as the cost-effective and hands-free microphone array input device. The Nintendo Wii remote was programmed using GlovePie environment [18]. The system architecture is presented in Figure 2.

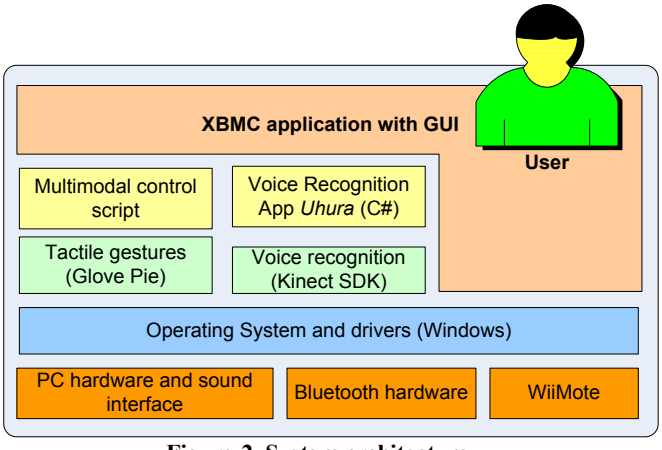


Figure 2. System architecture.

4. EXPERIMENT SETUP

Fourteen participants took part in the user evaluation study. Ten of them were male and four were female. The youngest participant was 22 years old, whilst the oldest participant was 42 years old. The average participant's age was 32 years (6.21 standard deviation). Four participants declared themselves as expert users of multimedia systems and eight of them declared as regular users. Two participants said they were occasional users and only had little experience with multimedia systems. Previous experience with multimedia systems was assessed on a five point Liker scale, where 1 indicated little or no experience and 5 an expert user. An expert user has a lot of experience with multimedia systems (e.g. IPTV, VoD, home multimedia repositories, etc.) on different devices (set-top-box, smart phone, tablet PC, etc.) on a daily basis. Among the participants one user was left handed, one ambidextrous and all the other users were right handed.

In the first part of the study (RC), the participants were interacting with the TV based multimedia system using only a standard remote controller. For the purpose of the evaluation study, the participants were asked to carry out seven different tasks, such as: play specific video, music, mute/unmute sound, open specific picture, show weather info. Tasks were the same for all participants. For every participant, for each task, the number of repetitions for a successful task completion was measured. During the study, participants were encouraged to think aloud i.e. to talk about the problems they encounter and to express their general impressions of navigating the system with the remote control. After some time of usage, the participants were invited to fill out the standard usability questionnaire System Usability Scale (SUS). Additionally, the participants were asked to complete the AttrakDiff questionnaire on-line. At the end of this part of the study, they were invited to suggest how the system should be improved to better fit with their expectations.

In the second part of the evaluation study (RC and voice), the whole procedure was repeated. However, in this part participants were interacting with the TV based multimedia system using a remote controller in combination with voice commands and tactile gestures.

The evaluation procedure was counterbalanced where half of the users (odd number, even number principle) performed the unimodal part first and the multimodal second and the other half users performed the multimodal part first and the unimodal second. The study was conducted in a home environment in order to provide as accurate and comfortable test conditions as possible, as shown in Figure 3.

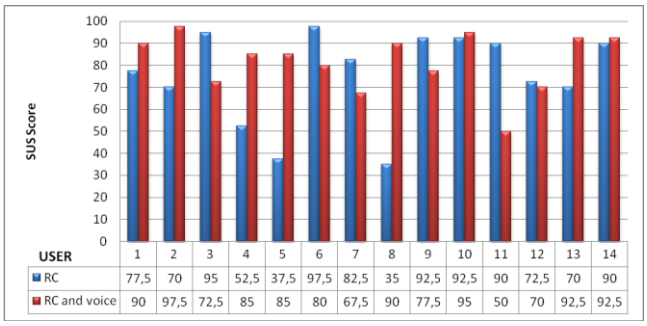


Figure 3. Evaluation setup in a home environment.

5. RESULTS

The average SUS score for the interaction with remote controller was 75.36. The lowest obtained SUS score in this part of the user evaluation study was 35, whilst the highest SUS score was 97.5. The average SUS score for the interaction with remote controller in combination with voice commands was 81.79. The lowest obtained individual SUS score in this part of the study was 50.

The highest individual SUS score was same as the one obtained in the first part of the study i.e. 97.5. Individual SUS scores for both parts of the study are presented in Figure 4.



USER	1	2	3	4	5	6	7	8	9	10	11	12	13	14
RC	77,5	70	95	52,5	37,5	97,5	82,5	35	92,5	92,5	90	72,5	70	90
RC and voice	90	97,5	72,5	85	85	80	67,5	90	77,5	95	50	70	92,5	92,5

Figure 4. Individual SUS scores for both parts of the user evaluation study.

The results obtained from the on-line questionnaire AttrakDiff for both types of interaction – unimodal remote controller (part A) and the combination of remote controller with voice commands and tactile gestures (part B) are presented in Figure 5.

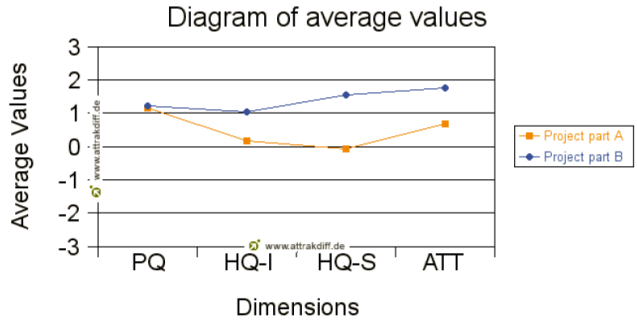


Figure 5. Mean values of the four AttrakDiff dimensions for both types of navigation.

6. DISCUSSION

To explain what the obtained SUS score means in describing the usability of described types of navigation we used Bangor et al. findings [19]. In terms of adjective rating, both types of interactions can be described as "Excellent". However, as presented above, the individual SUS score in the second part of the study (81.79) was higher than the individual SUS score in the first part of the study (75.36). According to these results, in terms of usability, study participants found the combined use of remote control and voice commands better than the solely use of the remote control for interacting with the TV based multimedia system. However since the difference in the individual SUS score values is statistically insignificant, further analysis about the usability are needed.

The AttrakDiff questionnaire was used in this study in order to evaluate the attractiveness of the presented types of navigation. As it can be seen from Figure 5, the interaction with the combination of the remote control and voice commands performs better than the interaction with the unimodal remote control. Even though, pragmatic quality is approximately the same, the hedonic quality for the combined interaction is higher.

In terms of pragmatic quality, the interaction with the combination of the remote control and voice commands perform better than the interaction with the unimodal remote control. The difference in value however is statistically insignificant. Both types of interactions are located in the above-average region and meet the ordinary standards. Even though the study participants find both interactions usable and they were successful in achieving their goals, further improvements are needed in both types of interactions.

The difference in the value of the hedonic quality of the interaction with only a remote control and the combined interaction is statistically significant. The identity parameter of the hedonic quality for the interaction with a remote control is located in the average region. On the other side, the same parameter for the combined interaction is located in the above-average region. The simulation parameter of the hedonic quality for the interaction with a remote control is located in the average region and it just about meets ordinary standards. On the opposite, in terms of aspects of stimulation, the combined interaction with remote control and voice commands is classified optimal.

According to the AttrakDiff results, overall, the study participants find the interaction with the TV based multimedia system with a combination of a remote control and voice commands as very attractive, while they find the interaction with only a remote control as moderately attractive.

While interacting with the TV based multimedia system, study participants reported some of the problems they have encountered:

- "Searching for the buttons on the remote control can be distractive."
- "The remote itself is unnecessary complex and has too many buttons."
- "The system is not reliable enough and sensitive to environment sounds." Note: user was interacting with the TV based multimedia system in a room with small children in background.
- "Sometimes voice commands are not recognized if one is talking in a lower voice."
- "Voice shortcuts are cool; however the overall reliability could be an issue."

Additionally, study participants made suggestions for some further improvements:

- "An indication for the user which (voice) command was understood (by the system)."
- "Additional second screen device with tactile screen interface and touch gestures."

7. CONCLUSION

In the paper, the results of a user evaluation study comparing a standard unimodal television remote controller to the multimodal WiiMote controller including tactile gesture and voice command inputs are presented. The study results showed that the addition of voice and alternative navigation modalities significantly improve the overall user experience. From the study results we discovered there is a lot of potential in multimodal interaction with the television and the other multimedia sets.

8. ACKNOWLEDGEMENTS

The operation that led to this paper is partially financed by the European Union, European Social Fund and the Slovenian research Agency, grant No. P2-0246. The authors also thank the participants who took part in the evaluation study.

9. REFERENCES

[1] Schatz, R., Baillie, L., Froehlich, P., Egger, S., Grechenig, T., "What Are You Viewing?" Exploring the Pervasive Social TV Experience. *Human-Computer Interaction Series, 2010, Mobile TV: Customizing Content and Experience*, Part 5, 255-290.

[2] Iskratel Innbox, accessed at http://www.innbox.net/en/ (last seen on 1.3.2012)

[3] Oviatt, S., Breaking the Robustness Barrier: Recent Progress on the Design of Robust Multimodal Systems. *Advances in computers*, vol. 56, 2002, 305-341.

[4] Barthelmess, P., Oviatt, S., Multimodal Interfaces: Combining Interfaces to Accomplish a Single Task. *HCI Beyond the GUI*, 2008, 391-444.

[5] Jaimes, A., Sebe, N., Multimodal Human Computer Interaction: A Survey. *Computer Vision and Image Understanding* 108, 2007, 116–134.

[6] Sodnik, J., Dicke, C., Tomazic, S., Billinghurst, M., A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, vol. 66(5), 2008, 318-332.

[7] Krapichler, C., Haubner, M., Lösch, A.,Schuhmann, D., Seemann, M., Englmeier, K., Physicians in virtual environments — multimodal human–computer interaction. *Interacting with Computers 11*, 1999, 427–452.

[8] Delgado, H., Rodriguez-Alsina, A., Gurgui, A., Marti, E., Serrano, J., Carrabina, J. Enhancing Accessibility through Speech Technologies on AAL Telemedicine Services for iTV. *D. Keyson et al. (Eds.): AmI 2011, LNCS 7040*, 300–308.

[9] Sturm, J., Cranen, B., Effects of prolonged use on the usability of a multimodal form-filling interface.), *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*, Springer 2005, 329-348.

[10] Guna, J., Kos, A., Pogačnik, M., Evaluation of the multimodal interaction concept in virtual worlds. *Electrotechnical Review, Journal of Electrical Engineering and Computer Science*, Vol 77(5), 2010.

[11] Berglund, A., Johansson, P., Using speech and dialogue for interactive TV navigation. *Univ Access Inf Soc* (2004) 3, 224–238.

[12] Fujita, K. ; Kuwano, H. ; Tsuzuki, T. ; Ono, Y. ; Ishihara, T., A new digital TV interface employing speech recognition. *Consumer Electronics, IEEE Transactions on*, 2003, 765-769.

[13] Wechsung, I. ; Naumann, A.B., Evaluating a multimodal remote control: The interplay between user experience and usability. *Quality of Multimedia Experience, International Workshop on*, 2009, 19-22.

[14] Holzinger, A., Searle, G., Wernbacher, M., The effect of previous exposure to technology on acceptance and its importance in usability and accessibility engineering. *Univ Access Inf Soc 10*, 2011, 245–260.

[15] Lee Johnny C.: Hacking the Nintendo Wii Remote, IEEE Pervasive Computing, Volume 7, Issue 3, 2008

[16] XBMC, accessed at http://xbmc.org/ (last seen on 1.3.2012)

[17] Microsoft Kinect and Speech SDK, accessed at http://www.microsoft.com/en-us/kinectforwindows/ (last seen on 1.3.2012)

[18] GlovePie Software, accessed at http://glovepie.org/ (last seen on 1.3.2012)

[19] Bangor, A., Kortum, P., Miller, J., Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies* (2009), Volume: 4, Issue: 3, 114-123.

Workshop 2: Third Euro ITV Workshop on Interactive Digital TV in Emergent Economies

Context Sensitive Adaptation of Television Interface

Aniruddha Sinha
Tata Consultancy Services Ltd.
Kolkata, India
aniruddha.s@tcs.com

Kingshuk Chakravarty
Tata Consultancy Services Ltd.
Kolkata, India
kingshuk.chakravarty@tcs.com

Debatri Chatterjee
Tata Consultancy Services Ltd.
Kolkata, India
debatri.chatterjee@tcs.com

Somnath Ghoshdastidar
Tata Consultancy Services Ltd.
Kolkata, India
som.ghoshdastidar@tcs.com

ABSTRACT
With the popularization of the smart television (TV), more and more new applications are emerging. In spite of the applications being interesting and useful for a large community, the presentation of the content and the user interaction plays a major role in determining its acceptability. Hence there is a need for personalization. In this paper, we propose a system for automating the adaptation of the TV display according to user profile information obtained from the second screen. Due to the high rate of penetration, mobile phones have been selected as the second screen. The user profile information is derived from the phone settings (text font, color, display intensity, audio volume, type of ring-tone etc), speed at which the user interacts with the phone, types of applications used and by processing of conversational speech. We have considered distance education application on TV as the use case to demonstrate the possible context sensitive adaptations based on the user profiles. We present the results using two user categories, namely, elderly people and housewives.

Categories and Subject Descriptors
H.5.2 **[User interfaces],** D.2.2 [**Design tools and techniques],** H.1.2 **[User/Machine systems]**

General Terms
Design, Human Factors.

Keywords
Adaptation, TV User Interface, User context, second screen

1. INTRODUCTION
A smarter world is to display essential and important information in a user friendly fashion. In the year 2009, household penetration of television in Asia Pacific region was significantly high at 75% and it is increasing day by day [1].

In spite of this high penetration, the potential of interactive television remains unfulfilled [2]. In a household, usually one single television is shared by all the members of the family. For each individual viewer, the Television settings might be different. These settings need to be changed manually. This process is time consuming and laborious. This effort can be minimized if we could automate the process of changing the settings of Television. or in other words, TV might be smart enough to adopt these preferences automatically.

Due to very high penetration rate of mobile phones (approximately 74% in India and in some countries as high as 120% [3]), new smart phone applications developed show that smart phones can be used for different purposes like mobile phone as input device for ubiquitous applications, as TV remote control, user context extraction device etc[12],[15],[16]. Some applications have been developed involving interactions between TV and mobile phone inter-audience interaction etc [13],[14]. In this paper we are proposing a novel system where the user interface of the Television which is the first screen, can be updated automatically by using the profile information extracted from the second screen.

User preferences differ based on languages, customs, cognitive levels, physical deficiencies and experience level. These differences are obtained from the usage of the mobile phones. If the user profile of the mobile phone is changed, then that change also gets reflected in the Television screen without user's intervention i.e ubiquitously. Here the mobile phone is used as the second screen. For implementing the proposed system, we have used an android phone (Google Nexus S) as the second screen. The user profile information stored in this mobile phone are extracted and sent to a Over The Top (OTT) box via Bluetooth. This box decides on the changes to be incorporated and finally those changes are automatically adopted on a Television screen.

As a demonstration of the system we have considered the Tele education application explained in [28]. This application can be used for individual as well as group teaching. Some of the possible target groups are Elderly, Kids, College students, housewives, physically challenged persons etc. Out of these groups, we have selected elderly people (age 60-75 yrs) and housewives as the target audience. The motivation for this selection is based on the statistics available in literatures. According to recent studies conducted in India, approximately 40% of the total enrolments in distance education are women [35] and elderly education is also gaining interest. The situation is similar for other developing countries also. In a cross sectional study conducted from University of Cape Coast, University of Education and foreign programs run by university of Ghana, it was found that 54% of the total enrollments are female [36].In fact a survey shows that the "class programme" broadcasted by UGC was popular amongst the housewives than students for whom it was meant.

Researches show that the number of elderly people is increasing and their characteristics are also changing. In the developed countries, in the year 2006, 20% of the total population was 60 years or older [4]. This population is projected to be 30% by the year 2050 [4]. The trend is similar for the Indian population also [10].Elderly population is expected to be approximately 10% of the total population by 2021.[10] This older population is very interested in new technologies, ideas and hence a new career. In a survey 71% of the Americans aged 25 to 70 said they want to continue working past their expected retirement age [8].15% of these people plan to start a new business, while 7% plan to work

fulltime in a new career and 30% would like to work part-time [8]. This brings out the importance of elderly education. Elderly people want to receive an additional education that quickly puts them on the path of a new career. In the year 2005, 27% of the total enrollments in different work-related courses were aged between 55 – 65yrs [8].

In current context of developing countries like India, the distance education is largely dependant on print media. Techniques like tele conferencing, interactive video are being experimental in a limited way. Hence the study materials are not at all interactive or cannot be adapted according to one's preferences. In the system proposed in this paper, the education application is interactive and the display can be automatically and ubiquitously modified in the smart TV according to student's preferences.

We initially present the mapping of different features from users' mobile phone to TV followed by a system solution that is used to perform the automatic adaptation of user interface (UI) on TV. Finally the adaptive TV interfaces are presented for the selected categories of the user, namely elderly people and housewives.

2. FEATURE MAPPING AND ADAPTATION

Based on the literature survey, a summary of the parameters and their usage guidelines for designing the user interface for elderly people is shown in Table 1. This depicts the features those are usually preferred by the elderly people. Usually these parameters on the phone are set by the user according to their own preferences. Apart from these, the brightness, contrast and languages are also set by the user. The inverse of the contrast detection threshold is defined as contrast sensitivity [25] which is more for the elderly and visually impaired people compared to the normal people [26]. The measure of response time to perform an action using the mobile phone and the speed of keypad interactions give important information about the cognitive level of the user [37]. Distance of the display from the user and the size of the display are the two major differences between a TV and a phone, which mainly demands for the adaptation.

Table 1: Mapping of features for TV and mobile phones

Parameters	Features for mobile phones	Features for TV
Display	it should be large	21' and above [22]
Color	Avoid high bright, fluorescent, vibrant colors that appear to blur [9]	Avoid specifying pure colors; that is, colors that have a non-zero value for only one of the RGB quantities. Avoid selecting any R, G, or B value that is within 20% of maximum; that is, 204 (255 - 20%). [9] Use TV safe colour palette [18]
Fonts	font size should be between 12pt – 14pt [3], [12]	font size should be between 18pt – 24pt [17]
Audio volume	70-85db [5]	70-85db [5]

2.1 Effect of Distance

Image size and viewing distance relationship is shown in Figure 1.

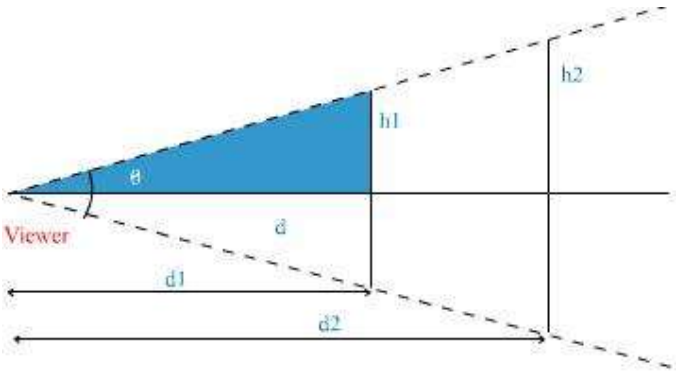


Figure 1 : Image size and viewing distance relationship

From Figure 1, we have

h1 = image size in mobile screen

d1 = distance of viewing for mobile phone

d2 = distance of TV viewing

h2 = image size in TV

Now, $\tan\theta = h1/d1 = h2/d2$, for $0 \text{ degree} < \theta < 90 \text{ degree}$

Thus for any given TV and mobile screen size, if h1= x inches, d1= y feet, d2 = 12y feet, then h2 needs to be 12x inches.

Mobile phone displays are physically smaller than television display. Most users today have a mobile that's 3.5″ or less in size, but 21″ and larger televisions are very common [22].



Figure 2: Comparative screen sizes

Consider a mobile phone having a diagonal display size of 3.5 inch with resolution of 640 x 960 pixels, having about 326 pixels per inch on the diagonal [20]. A 42″ TV monitor displaying 1080p video at resolution 1920 x 1080 pixels gives about 52 pixels per inch [20]. Figure 2 gives the comparative screen sizes. Generally people watch their televisions from 12 to 15 feet [19]. So an image of resolution 256 x 256 pixels is about an inch tall on the mobile screen, which is approximately 6 inch on 42" television but still appears relatively small to a viewer at 12 ft away as shown in Figure 3. Thus the perceived size mainly depends upon the distance between viewer and the screen. Hence an application icon which seems big on mobile screen seen from a distance of 1 foot appears to be small on TV screen seen at a distance of 12 feet; even it occupies the same number of pixels.



(a) **(b)**

Figure 3: Perceived size of same icon of resolution 256x256 pixels on (a) mobile (b) TV

Thus everything on TV user interface has to be approximately 1.5 to 2 times larger in pixel resolution as compared to the mobile phone user interface [19].

3. SYSTEM OVERVIEW AND IMPLEMENTATION

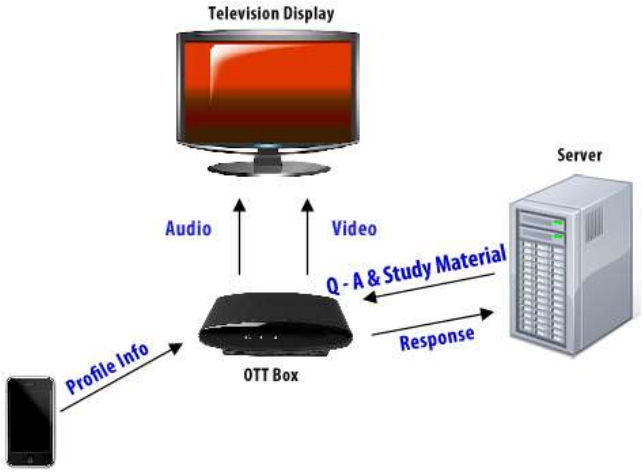


Figure 4: System diagram

We are recommending a system where we can customize the display on TV according to user cognitive ability and perception level. The system diagram is shown in Figure 4.

It is primarily motivated by the need to display education content i.e question-answer and study material, according to different users' widely varying cognitive ability and perception. For example, elderly people have totally different cognitive and perception level from others.

In the proposed system, the smart TV contains certain default settings for different user categories Xdi, for all the categories i. The features derived from the individual mobile phone settings is Y. Based on the feature set Y, the probability of the user category, P(Xdi/Y) is derived. The category i is selected which gives the maximum probability. The settings targeted for the TV interface is further refined based on the deviation of Y from the default feature set Ydi for the ith user category, such that the deviation is preserved in the final settings for Xi.

An easy option is to have the customized tutorials available in the education server and based on user category, one can access the same. However, this demands an increase in the storage requirement in the server. Hence we propose the dynamic adaptations in the OTT box.

We target two categories of user, namely elderly people and housewives. As Mobile phone is generally used as personal computing device for aged to young, it can be used to sense user cognitive ability and perception. Hence, these mobile-phone user profiles can be used to personalize display style of any TV set. As profile information we consider the following profile information vectors

1. Icon size 2. Font-Size 3. Foreground Color 4. Background-color 5. Brightness contrast 6. Microphone and speaker volume

Components

Implementation of the system requires following elements in its system [Fig.4]

1. Google Nexus S Android based mobile phone (android version 2.3.3) – capable of sensing user profile.
2. A traditional TV set
3. A remote Central Distance Learning (CDL) server which keeps all education contents like lecture video, image, text, questionnaire files.
4. Over The Top(OTT) box capable of
   - Communicating with interactive CDL server
   - rendering Question-Answer (Q-A) sessions on TV screen
   - Connecting mobile phone via bluetooth dongle. Here Linksys Bluetooth dongle is used to communicate with Google Nexus S phone.
5. Linksys Blue tooth dongle connected to OTT box.

Assumption

- OTT box is connected to CDL server via internet.
- OTT box is paired and connected via Bluetooth to Google Nexus S, so that it can send and receive files to/from the mobile phone.

System overview

When user group switches on the TV set to attend the lecture video session on some specific topic; OTT box connects with the remote CDL server for lecture video or study material, questionnaire and metadata. Metadata file has the necessary information about Q-A sessions. At run-time, OTT box constructs the questionnaire from the metadata file in html format. This questionnaire can be customized for different users, based on their preferences, as shown in Figure 5. Text to Speech [21] software is used to convert the questionnaire into Audio format. During this conversion, profile information gathered from mobile phone is also taken into account, e.g for elderly people, the question and instructions are read out slowly and clearly. After that, this audio format is blended with the html text in the OTT box, so that they could be played simultaneously. The Q-A sessions and the corresponding playback are also possible in native languages [21]. This language selection is done as explained in illustration 2 below.

The system workflow is shown above in Figure 5. A user profile sensing application (MOBUSR_PROF_APP) is designed on Android platform 2.3.3 so that it can run on Google Nexus S phone and correctly get the user profile information. Application will first check whether the profile information vector is set on

mobile-phone or not. If it is not set it will prompt user to set the profile manually. Otherwise it will send this profile information vector to OTT box. Then OTT box will extract feature vector set (mob_fv(i)) which will be used to categorize user profile. After getting feature vector set, OTT_App_Disp will match the feature vector set(mob_fv(i)) with its stored standard feature vector set(stndrd_fv(i)) to categorize user. This matching of mob_fv(i) with respect to stndrd_fv(i) is done in Decision Filter section of the application OTT_App_Disp. If the output of the Decision Filter is "Elderly", OTT_App_Disp will map the mob_fv(i) with elderly feature vector set (eldr_fv(i)). As an output, user can observe changes in TV display for education content.

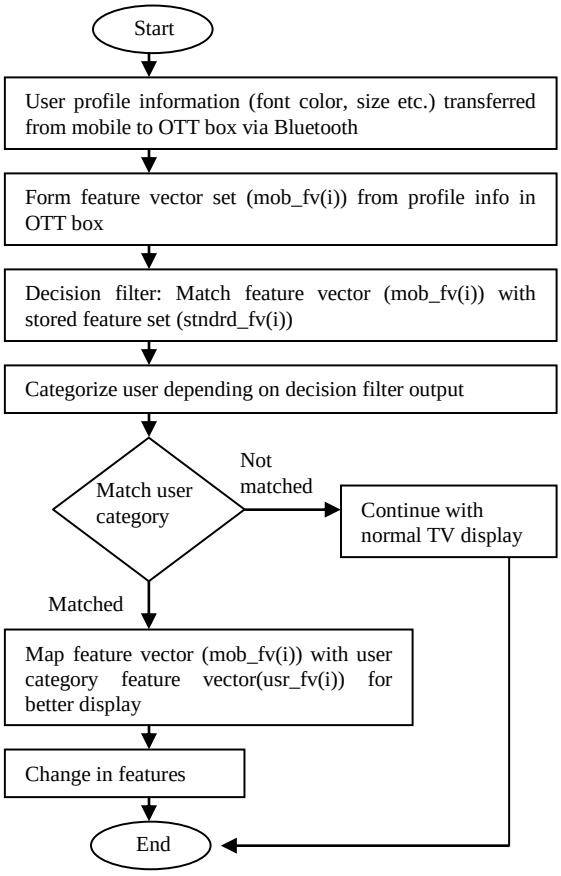


Figure 5: Implemented system workflows for interactivity

MOBUSR_PROF_APP application runs Parallel Phone Recognition followed by Language Modeling (parallel PRLM) [34] algorithm to identify the user conversational language from telephonic speech. This language information along with mobile phone default languages (set by the user) are stored in the mobile phone.

While communicating with the OTT box, language information is also transferred via Bluetooth. OTT box then checks for those language supports in

- Text Questionnaire
- Audio playback of the questionnaire
- Lecture video.

If the option(s) are available, then OTT box will ask for user permission to change the language. The entire process is shown in Figure 6.

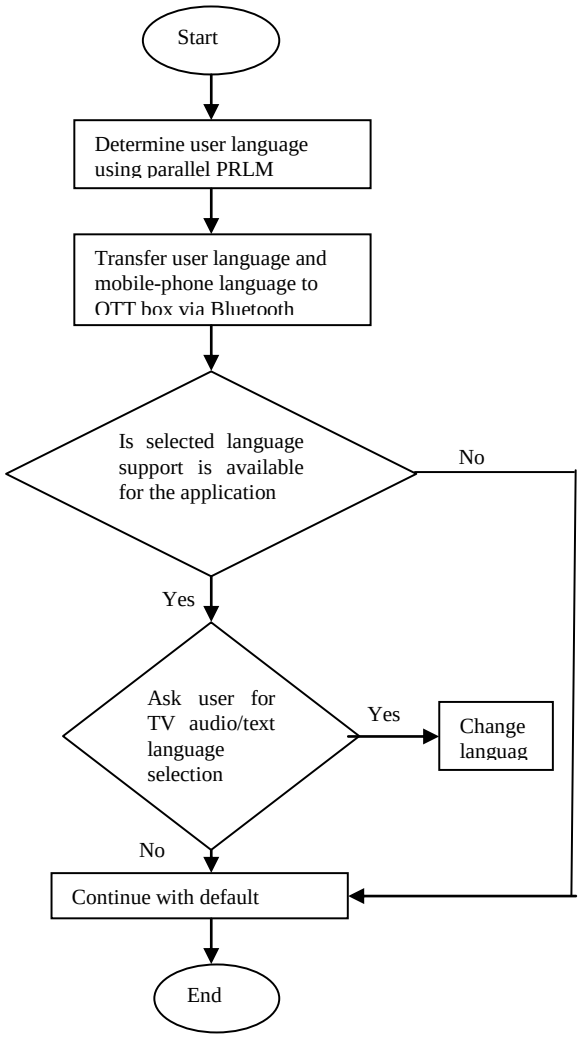


Figure 6: Language selection using parallel PRLM

4. INTERACTIVE TV INTERFACE

The creation of educational courses, study materials etc need consideration of the target age group. There are certain guidelines that are characteristics and preferences of elderly people [5], [6], [7]. The components of the displayed content for the interactive home-education system are given in Figure 7. Adaptation of each of these components is performed based on the user profile.

The user profile information stored in the mobile is transferred to the TV via OTT box which is used to automatically adopt the content locally at run-time. The screen shots for the adaptation of our interactive home-education application for different users are given for illustration. Figure 8 shows the menu screen for housewives or the normal display. Here the font size is 18pt, icons are of standard size and icons are colored ones. The same screen is modified if the user using it is an elderly person as shown in

Figure 9. Here the icon size is larger compared to the previous one, the font size is 24pts and also the icons are made black and white as this is the most preferred combination for elderly people. Similar adaptation is done for the display of the Electronic Program Guide (EPG).

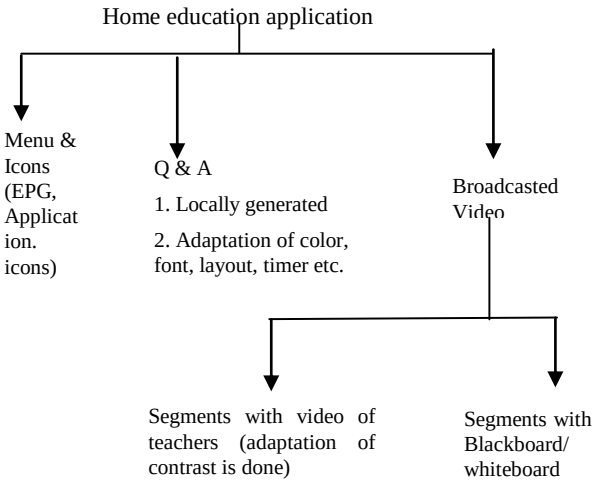


Figure 7: Components of home-education application

The locally generated HTML pages for displaying the Q&A [28], [29] undergo similar adaptation in terms of color selection, font size, brightness, and contrast with the guidelines as shown in Table 1. The sample Q&A for housewives is shown in Figure 10 where the answer options are given in the form of standard selection bullets. The modified version of the Q&A presentation for elderly people is presented in Figure 11. It can be seen that the presentation is done in the numbered form rather than selection of bullets which reduces the number of key interactions for the elderly people, whereas the answer selection using bullets is more natural for a normal housewife. The cognitive level is determined from the usage of the mobile phone [37] and accordingly the timer for the display of Q&A is adapted.



Figure 8: Menu screen for housewives



Figure 9: Adopted menu screen for elderly

The tutorial video is either downloaded over internet [28] or broadcasted over satellite TV network [30]. This contains majorly two types of segments, namely, the writings on the whiteboard or blackboard and the segments with teachers on the screen. These segments can be automatically detected in real time using scene change algorithms [31], [32], [33].



Figure 10: Sample screen of Q&A for housewives



Figure 11: Sample screen of Q&A for elderly people

The video for the segment with blackboard (Figure 12) is by default a multicolored one which is highly suitable for housewives. The same video is modified with the inputs from an elderly person's mobile phone as shown in Figure 13. Here the font color has been made white in black background and the contrast is enhanced as suggested in Table 1. This is done by eliminating the color components (U and V) from the video and enhancing the luminance (Y) components.

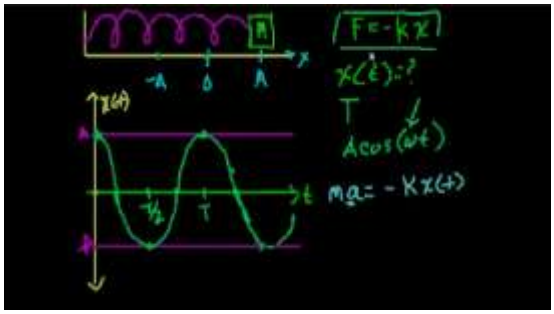


Figure 12: Lecture video with blackboard segments for housewives

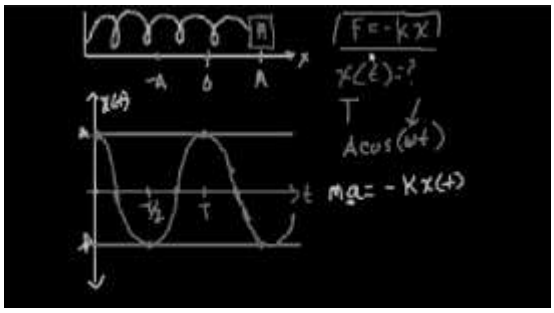


Figure 13: Lecture video with blackboard segments adapted for elderly people

The sample frame of video segments with teachers on the screen is shown in Figure 14 as the original content received by the OTT might be suitable for the housewives but may not be for the elderly people who in majority have higher contrast sensitivity [25]. Hence the video segments are adapted in real-time during display by modifying the pixel distribution [27] to enhance the important images locally and by using the edge enhancement algorithm based on Laplacian operator [27]. An image of such an adaptation is shown in Figure 15. This sample tutorial video is taken from the Lecture series on "Design of Reinforced Concrete Structures" by Prof. N.Dhang, Department of Civil Engineering, IIT Kharagpur, India. The video is generated under the National Program on Technology Enhanced Learning (NPTEL) and is available at the website http://www.nptelvideos.com/video.php?id=1647&c=11.



Figure 14: Lecture video with teachers on screen display for housewives

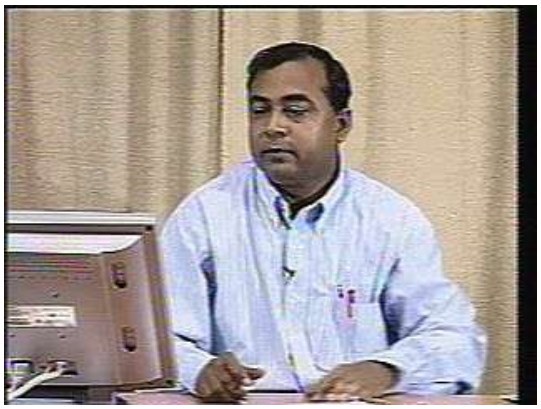


Figure 15: Lecture video with teachers on screen display adapted for elderly people

5. CONCLUSION

In this paper we have proposed a system solution that ubiquitously adapts the presentation of smart TV applications based on the user profile and context derived from the personal mobile phone. Distance education application is used to demonstrate the UI adaptation on TV. The features used to derive the user profile information include the settings on the phone related to color, font size, brightness, contrast, audio volume, language etc. The conversational language is also analyzed to derive additional profile information. Cognitive level of the user is derived from the interaction patterns with the mobile phone. These features are used to adapt the presentation of the content of the distance education such that it is more amicable to the said user category. The educational content is analyzed in real time to partition into different segments including QA, writing on blackboard/whiteboard, teacher's lecture etc. These segments are adapted using different types of algorithms. Further work needs to be done on the user study in order to measure the change in satisfaction index.

6. REFERENCES

[1] https: // www.itu.int/ITU-D/ict/publications/ wtdr_10 /material/WTDR2010_Target8_e.pdf

[2] William Cooper, "The interactive television user experience so far", Proceeding of 1st international conference on Designing interactive user experiences for TV and video, October 22-24, 2008.Sillicon Valley, California, USA.

[3] http://en.wikipedia.org/wiki/List_of_mobile_network_operators_of_the_Asia_Pacific_region

[4] Verstockt, S.; Decoo, D.; Van Nieuwenhuyse, D.; De Pauw, F.; Van de Walle, R.;, "Assistive smartphone for people with special needs : The Personal Social Assistant," Human System Interactions, 2009. HSI '09. 2nd Conference, vol., no., pp.331-337, 21-23 May 2009

[5] http://www.humanfactors.com/library/aug01uc.html

[6] Muhammad Mehrban, Muhammad Asif, "Challenges and Strategies in Mobile Phones Interface for elder people", Master Thesis, Computer Science, School of Computing, Blekinge Institute of Technology, Sweden, Thesis no: MCS-2010-10 January 2010

[7] http://www.tiresias.org/research/guidelines/telecoms/mobile.html

[8] http://www.acenet.edu/Content/NavigationMenu/ProgramsServices/CLLL/Reinvesting/Reinvestingfinal.pdf

[9] http://www.europosparkas.lt/phare/rekomendacijos.pdf

[10] http://www.cehat.org/humanrights/rajan.pdf

[11] http://mobisocial.stanford.edu/papers/mobicase11.pdf

[12] http://php.scripts.psu.edu/users/o/o/oob100/127TOCOMMJ.pdf

[13] http://dl.acm.org/citation.cfm?id=2071513

[14] http://www.tvtak.com/developers.html

[15] http://www.cs.cmu.edu/afs/cs.cmu.edu/Web/People/aura/docdir/sensay_iswc.pdf

[16] http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?reload=true&arnumber=1593574

[17] http://www.mhp.org/docs/itv-design_v1.pdf

[18] http://www.serlin.co.uk/colour/dtv_colour_chart.html

[19] http://www.softwarequalityconnection.com/2011/05/re-thinking-user-interface-design-for-the-tv-platform/

[20] http://en.wikipedia.org/wiki/List_of_displays_by_pixel_density

[21] Kurian S.A, Narayan B,Madasamy N, Bellur A,Krishnan R,G. Kasthuri,V.M Vinodh, Murthy A. H, "Indian Language Screen Readers and Syllable Based Festival Text-to-Speech Synthesis System" Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pages 63–72,Edinburgh, Scotland, UK, July 30, 2011

[22] http://www.adi-media.com/PDF/TVJ/annual_issue/002-Color-Televisions.pdf

[23] http://encyclopedia2.thefreedictionary.com/type+font

[24] http://www.brighthub.com/electronics/home-theater/articles/34539.aspx

[25] Congdon, N; O' Colmain, B; Klaver, C C; Klein, R; Munoz,B; Friedman, D S; Kempen, J; Taylor, H R; Mitchell P, "Causes and prevalence of visual impairment among adults in the United States," Archives of Ophthalmology, 122/4, 477-485(2004)

[26] E. Peli, "Recognition performance and perceived quality of video enhanced for the visually impaired," Ophthalmic and Physiological Optics, vol. 25, pp. 543-555, 2005

[27] Saha Biswanath, Bhowmick Brojeshwar, Sinha Aniruddha, "An Embedded Solution for Visually Impaired", IEEE International Symposium on Consumer Electronics (ISCE), 2009

[28] Sujal Wattamwar, Tavleen Oberoi, Anubha Jindal, Hiranmay Ghosh, Kingshuk Chakravorty, "A Novel Framework for Distance Education using Asynchronous Interaction", accepted in ACM MM MTDL 2011

[29] Sounak Dey, Kingshuk Chakravarty, Aniruddha Sinha, "An Interactive Low Cost Distance Education System Over Home Infotainment Platform Using Color Code", in IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS) 2011

[30] Anupam Basu, Aniruddha Sinha, Arindam Saha, Arpan Pal, Jan 2012, "Novel Approach in Distance Education using Satellite Broadcast", 3rd International Conference on e-Education, e-Business, e-Management and e-Learning, IPEDR vol.27 (2012) IACSIT Press, Singapore

[31] Bhowmick, B.; Goswami, K. "SVM Based Shot Boundary Detection Using Block Motion Feature Based on Statistical Moments", ICAPR 2009

[32] Bhowmick, B.; Chattopadhyay, D, "Shot Boundary Detection Using Texture Feature Based On Co-Occurrence Matrices", International Conference on Multimedia Signal Processing and Communication technology, 2009

[33] Tanushyam Chattopadhyay, Ayan Chaki, Utpal Garain, "A fast method for detection of video shot boundaries using compressed domain features of H.264 for PVR enabled Set Top Boxes: A novel approach", CICSyN 2009

[34] Marc A. Zissman, "Automatic Language Identification of Telephone Speech", The Lincoln Laboratory Journal Volume B. Number 2, 1995

[35] D. Janaki Dr., Mother Teresa Women's University, "Empowering Women through Distance Learning in India",The Fourth Pan-Commonwealth Forum on Open Learning, 2006

[36] "Ghana Distance Education Program" – case study of chapter 3 of the book Strategies for Sustainable Open and Distance Learning (www.col.org/SiteCollectionDocuments/Ch3_CS-Mallet.pdf)

[37] Holleis, P., Otto,Friederike., Hussmann,Heinrich.,Schmidt, Albrecht. Keystroke-level Model for Advanced Mobile Phone Interaction. In Proc. CHI'2007.

A Distance Education System for Internet Broadcast Using Tutorial Multiplexing

Sounak Dey, Arindam Halder, Arindam Saha, Abhishek Mishra, Aniruddha Sinha, Somnath G Dastidar
TCS Innovation Lab Kolkata 700091, India
+91336636 7433
{sounak.d, arindam.halder, ari.saha, abhishek11.m, aniruddha.s, som.ghoshdastidar}@tcs.com

ABSTRACT
The need for distance education is increasing in developing countries like India, especially in the rural areas. Though most of the solutions exist over internet protocol (IP), a very little focus is given on optimizing the use of the IP bandwidth. In this paper we present a system that uses IP multicast technique for sending video lectures to remote classrooms. The system is composed of broadcast station and an IP enabled client box, connected to a television (TV) where the broadcast station generates the stream in MPEG2 format containing the multiplexed tutorials. The tutorials generated in the teacher's workbench, using H.264 video encoder, are multiplexed before generating the MPEG2 stream. The broadcasted stream is recorded at the student's end based on the individual selection. The details of the implementation on the client box, at student's end, are presented using a message sequence diagram.

Categories and Subject Descriptors
J.6 [**Computer Aided Engineering**]

General Terms
Measurement, Performance, Design, Experimentation.

Keywords
Distance education, multiplexing, broadcasting, internet protocol.

1. INTRODUCTION
Recently solutions for distance education [5] have gained a revived focus specially in developing countries including India. TV based distance education systems popularly use different techniques to carry metadata within the lecture video streams like embedding water mark, color code [1], Quick reference (QR) Code [2], metadata [3] etc. Also, there are techniques to multiplexing different lecture videos and embedding rhetoric question answer sessions into single video stream [4]. This single video stream can also carry multiple audio streams within itself. This multiplexed single video stream is used to broadcast using television (TV) based distance education system [4]. Other solutions of distance education are proposed over internet protocol (IP) where the student downloads the required tutorial from the education server and the question and answer (Q-A) from the Q-A server [5]. During the peak hours, many students trying to download the contents from the server will put enormous load on the server. To overcome this problem, the solution with IP multicast [10] also exists for distance education. The average connection speed in developing countries like India is approximately 256 kbps which is majorly available in urban areas [11]. Though wireless broadband connection is improving the scenario, the state in rural areas remains same. In this paper, we present the multiplexing of multiple tutorial contents into one single IP broadcast channel to maximize the utilization of bandwidth. We use a method of broadcasting tutorials over IP, wherein the students can schedule a recording of the desired content and consume the same later. The selection and scheduling of the desired content is based on the electronic program guide (EPG) of the contents. Unlike the multiplexing of tutorials in analog video frame presented in [4], tutorials are digitally multiplexed in a single MPEG2 stream which is broadcasted over IP. In order to maximize the utilization of limited available bandwidth, we use a system, where multiple tutorials will be spatially and temporally multiplexed to reduce the broadcast cost for each tutorial with a penalty in the visual quality. In this paper we present a system solution of broadcasting tutorials over IP wherein the contents are multiplexed in a single IP channel.

2. SYSTEM DESCRIPTION
In our proposed system, the modules comprises of teachers workbench; lecture and Q-A repository; IP broadcast server and multiple student's station as shown in Figure 1. Brief descriptions of these modules are given below.

- Teachers Workbench – Teacher creates their video lectures and respective Q-A sessions. Using a special purpose editor, teacher denotes when exactly the Q-A sessions will appear in the lecture video timeline and using the same editor teacher publishes them. The lectures and Q-A files are kept separately so that modification in any one of them does not affect the other. All the Q-A files, video lecture files, related metadata files are kept in the repository. A detail of Teachers Workbench is described in [5].

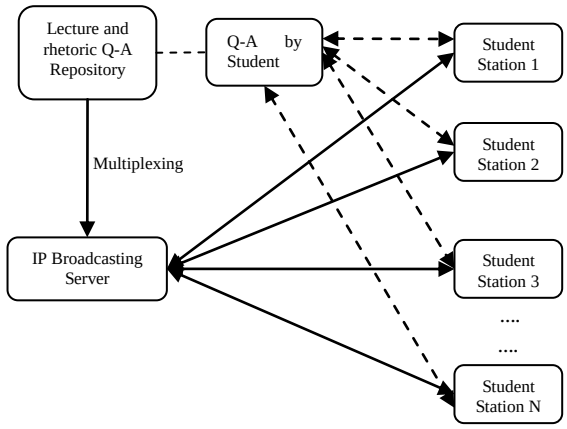


Figure 1. Proposed IP communication between Student Station and Broadcasting station.

- Lecture and Q-A repository – Published lectures videos along with Q-A sessions and their contents are stored here as mentioned above [5]. This repository can be maintained by broadcaster or content provider in a typical system.
- IP Broadcast Server – Before broadcasting, the server selects some tutorials (as per genre, class or subject) and scales down the resolution and frame rate according to the bandwidth requirements and digitally IP multiplexes these lecture videos and their related multiple language audios and Q-A contents into one single MP4 video stream.



Figure 2. Proposed UI for merging different lectures into a single stream.

```
Chemistry_11_146_Aldehydes_1327380635611-kolkata-Aldehydes_lecturevideo.mp4
Chemistry_11_139_MarkovnikovRule_1327382935978-kolkata-MarkovnikovRule_lecturevideo.mp4
Maths_11_173_QuadraticInEquality_1327450811992-kolkata-QuadraticInequality_lecturevideo.mp4
Maths_10_131_ComplexNumber_1327451671044-kolkata-ComplexNumber_lecturevideo.mp4
Physics_10_104_NewtonSecondLaw_1327453639152-kolkata-NewtonSecondLaw_lecturevideo.mp4
Physics_9_120_HarmonicMotion_1327454351837-kolkata-HarmonicMotion_lecturevideo.mp4
Maths_6_103_Average_1327887607786-kolkata-Average_lecturevideo.mp4
Chemistry_11_121_SimpleAlkanes_1327988515840-kolkata-SimpleAlkanes_lecturevideo.mp4
```

Figure 3. Proposed sample of EPG data file to be embedded into video frames.

Finally the EPG is inserted in this stream to make it ready for broadcast by creating the MPEG2 transport stream. A typical UI for such operation at IP broadcast server may look like Figure 2. One sample EPG data file that can be incorporated into video frames may typically look like one shown in Figure 3.

- Student Station – Student Station is the most important part of the proposed system as it interfaces the real class of student. It has following tasks to do.
    - Showing the EPG
    - Recording the desired lecture
    - Downloading the related Q-A files and metadata files
    - Playing such recorded lectures
    - Allowing students to record and upload their questions
    - Allowing students to download answers of their question in real time.

We will explain all these tasks one by one. When multiplexed lecture video MPEG2 stream is received at student station via IP multicasting then system detects and parses the EPG packets and one EPG viewer shows the details of each lecture that are coming in the multiplexed stream. Now it is the responsibility of the moderator of the class to select and record the required lecture from that bundle. One sample UI for such EPG viewer is shown in Figure 4.



Figure 4. Proposed sample UI for showing EPG data for multicast stream and recording one of the lectures.

The EPG screen can be extended for showing streams that are coming in multiple time slots throughout the day so that moderator can schedule his recording beforehand. After extraction, each lecture along with it's related audio streams and Q-A files are stored for student perusal. When a lecture is being created, the author may refer to some related audio/video/text files for further studies or reference. Not necessarily, such files would be a part of broadcasted video lecture. These files can be downloaded separately through an IP channel while recording the lecture video. This is a special scenario when it might not be possible to embed everything related to a lecture into IP stream. Student station can download these files (which may be big in size or seldom required or may be some setting files etc) as and when required. These files may be required while playing the recorded lectures by class moderator. Students now has option of asking question to teacher while viewing it and moderator can record such questions and upload them in real time to the IP broadcast server through the IP channel. The last part of operation is downloading and viewing the answers for these questions. Again one sample UI for such purpose is shown in Figure 5. The red signifies the questions that are unanswered and green signifies answered ones. Presently the system supports audio and text mode of question and answer. Later video mode can be incorporated.

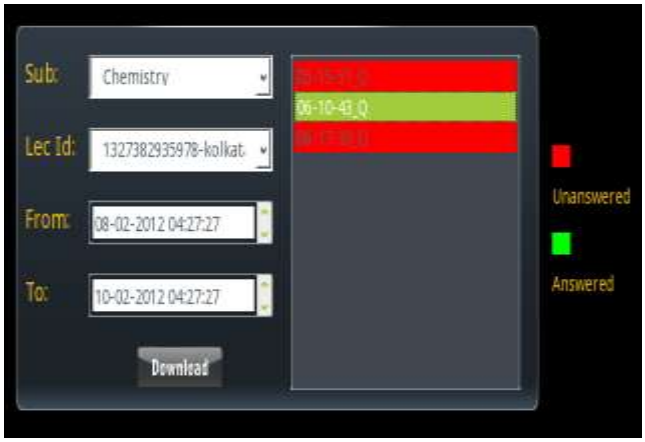


Figure 5. Proposed sample UI for showing questions asked by students and answers given.

3. MESSAGE SEQUENCE BETWEEN APPLICATION AND ALGORITHM

There are 5 modules in student station, namely EPG App, Recorder App, Storage, Packet Parser and Display Manager. The message flow between all these modules is shown in a sequence diagram of Figure 6 and explained in short below.

EPG App, Recorder App and Packet Parser are the three major modules in Student Station. The main responsibility of EPG App is to construct the EPG from the IP packets extracted by Packet Parser and display the complete EPG through Display Manager. Recorder App is the application responsible for recording a lecture using Packet Parser and constructs the final lecture video with rhetoric Q-A. It can also be used to schedule a future recording of lecture. The Packet Parser is responsible for extracting the IP packets from the network upon getting a request from any other application like EPG App, Recorder App. The Storage App will store any lecture to any storage medium (internal or external) connected with. User can check the state of the system as reflected in the UI as and when modified by these modules.

As Student Station is switched on and starts receiving the IP packets from broadcasting end, the EPG App module in it sense that and asks Packet parser module to parse the IP packets to build a proper EPG structure. Once the EPG structure is fully built, EPG App module shows it through Display Manger (on any TV or similar display device). Next, user selects one (or more) lecture video to be recorded using a simple UI (as shown in Figure 4). Consequently, Recorder App module asks Packet Parser module to extract the corresponding video lecture packets. As and when these packets are extracted, Recorder Module gets notified and it stores those packet data into a video file for later usage. Once the parsing and storing is over for a selected lecture (or set of lecture), the UI (Figure 4) gets updated automatically.

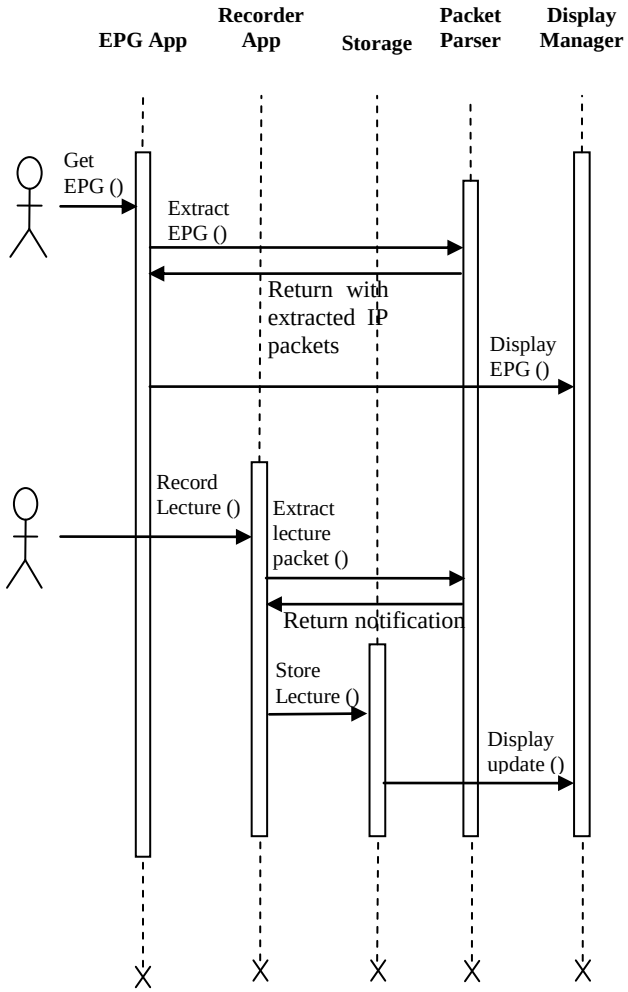


Figure 6. Message sequence diagram.

4. COMPARISON WITH PRESENT BROADCAST SYSTEMS

There are some existing solutions [5] which use these modules given above, wherein the student's station connects the tutorial storage for required lectures and creates a separate session with that server and downloads the required lecture. During peak hours, this kind of architecture creates a slow deadlock type scenario when every student station tries to create separate session with IP broadcast server thus dividing the available bandwidth into small pieces and as a result downloading process slows down a lot, especially when available bandwidth is not much.

There are other techniques where the multiplexed video can be broadcast using existing TV broadcast networks as given in [4]. In that case the lecture video can be extracted at student station but the student's query against a lecture cannot be recorded and sent to Broadcast Station unless there is an IP network attached to the student station which does the job. So Satellite broadcast scenario needs a parallel live IP network to complete the circuit hence making it a dependent architecture. Table 1 describes all these scenarios based on availability of internet bandwidth.

Table 1. Different Types of Content Distribution

Quality of internet connection	Data via Internet connection	Data via TV broadcast
Very high bandwidth internet connection [5]	Tutorial audio, video, QA from teachers end. Students can send queries to the teachers.	Not required
Low bandwidth internet connection [1]	Tutorial audio and QA from teachers end. Students can send queries to the teachers	Tutorial Video
No internet connection or very low bandwidth internet connection	Students can send queries to the teachers	Tutorial audio, video, QA from teachers end
No internet connection		Tutorial audio, video, QA from teachers end

Thus the only other viable option is multiplexing the whole lecture video along with audio streams and QA contents into MP4 stream and broadcasting through IP network (like one does in the case of IPTV) using MPEG2 transport stream, thus enabling the student station to record those video/audio/text contents and upload the student query through IP. In this scenario, all student station gets the stream of multiplexed stream as IP packets and lecture schedule as EPG packets. As per EPG packets, student station makes the schedule to record the required video and on getting the corresponding lecture IP packet it starts recording.
Since the IPTV architecture is both way i.e. student station can upload queries to QA server through same IP network (Figure 1).

4.1 CONTENT GENERATION

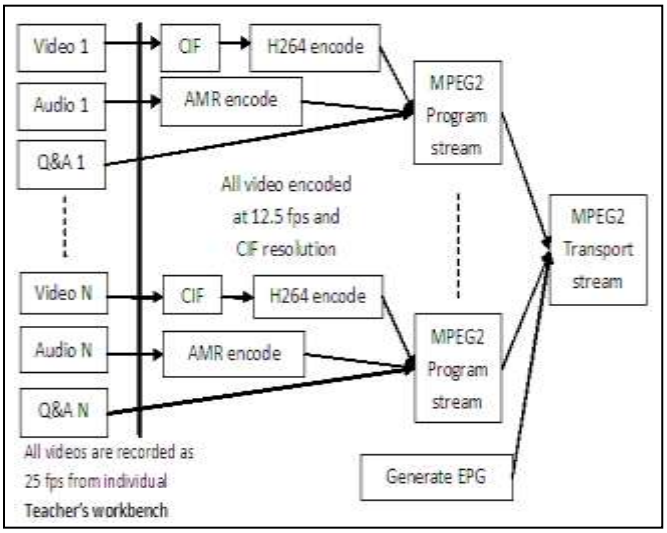


Figure 7. Content Generation

The method for broadcast content generation is shown in Figure 7. The video, audio and Q-A are received from the teacher's workbench. The video resolution and framerate of the teacher's workbench can be CIF/VGA/SD at 25 frames per second (fps). Each such video is converted to CIF (352x288) resolution and 12.5 fps, which is encoded using H.264 encoder. The audio is encoder with Adaptive Multirate (AMR) speech encoder. The audio, video and Q&A are converted to separate program stream. Finally all the program streams are multiplexed to generate a single MPEG2 transport stream, which is ready for broadcast. The EPG contains the relative time information for all the tutorials and is embedded as part of the transport stream.

4.2 CONTENT MANAGEMENT

As one video lecture and its rhetoric questions comes in a bundled way along with other lecture video, audio and their corresponding rhetoric questions, so if any one of the component needs modification then that individual component needs to be edited separately and then have to rebundled with whole stream again. This effort may be handled by some automated script like module which can re-bundle the edited component once again with other components. This has been depicted in Figure 8.
Now upon recording the lecture at student station, the quality of lecture video does not degrade much. A user study for such videos reveals the fact that such videos are well accepted for student perusal [8].

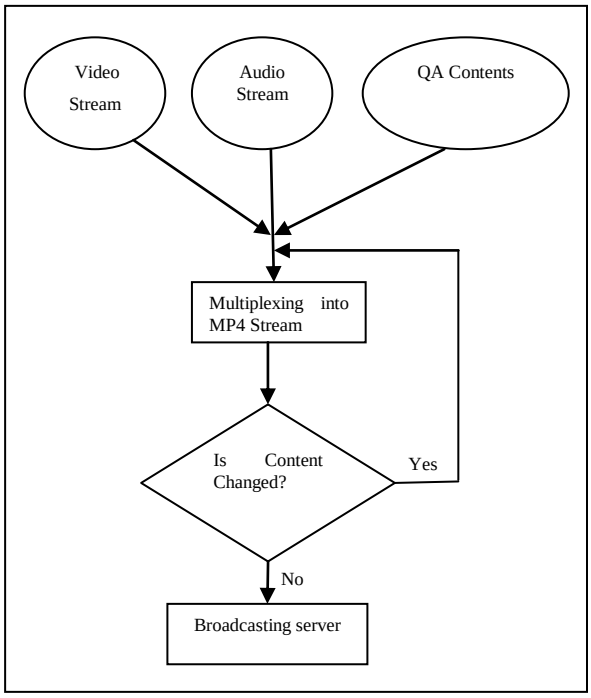


Figure 8. Content Management

5. CONCLUSION

We have discussed how IP broadcasting can be used for distance education systems with a multiplexed content management system. The degree of multiplexing will depend on the cost of each tutorial to broadcast. The feasibility and acceptability of multiplexing multiple tutorials in spatial and temporal domain needs to be studied. This type of system can work both ways so that student's station can also communicate with broadcaster through same channel. Future scope of work is personalization of the presentation of lecture content and rhetoric questions, wherein the same broadcasted content is adapted dynamically in the student station.

6. ACKNOWLEDGMENTS

Our sincere thanks to our colleagues Ranjan Dasgupta, Hiranmay Ghosh, Tavleen Oberoi, Romil Bansal for their valuable contributions.

7. REFERENCES

[1] Sounak, D., Kingshuk, C., Aniruddha, S. 2011. An Interactive Low Cost Distance Education System Over Home Infotainment Platform Using Color Code. In *IEEE International Conference on Intelligent Computing and Integrated Systems* (ICISS)

[2] Diptesh, D., Priyanka, S., Avik, G., Chirabrata, B. 2011. An interactive system using digital broadcasting and Quick Response code. In *IEEE International Symposium on Consumer Electronics* (ISCE), Singapore, 2011. DOI= 10.1109/ISCE.2011.5973857

[3] Arindam, S., Aniruddha, S., Arpan, P., Anupam, B., 2012. Embedding Metadata in Analog Video Frame for Distance Education. In *Proceedings of the IJEEEE 2012 Vol.2(1): 011-018 ISSN: 2010-3654*. DOI= http://www.ijeeee.org/abstract/073-Z00056F00017.htm.

[4] Arindam, S., Aniruddha, S., Arpan, P., Anupam, B., 2012. Multiplexing of Tutorials in Distance Education Using TV Broadcast Network. Accepted in *4th International Conference on Computer supported Education 2012*. DOI= http://www.csedu.org/.

[5] Sujal, W., Tavleen, O., Anubha, J., Hiranmay, G., Kingshuk, C. 2011. A Novel Framework for Distance Education using Asynchronous Interaction. *In Proceeding of MTDL '11 Proceedings of the third international ACM workshop on Multimedia technologies for distance learning*

ACM New York, NY, USA ©2011 ISBN: 978-1-4503-0994-3. DOI= http://dl.acm.org/citation.cfm?id=2072601

[6] Basu, P., Thamrin, A. H., Mikawa, S., Okawa, K. and Murai, J. 2007. Internet technologies and infrastructure for asia-wide distance education. In *Proceedings of the 2007 International Symposium on Applications and the Internet* (SAINT'07)

[7] Akamai Technologies, Q3 2010, Report on "State of the Internet", http://www.akamai.com/stateoftheinternet/

[8] Arindam, S., Aniruddha, S., Arpan, P., Anupam, B., 2012. User experience in multiplexing of Tutorials in Distance Education. Accepted in EdMedia 2012 - World Conference on Educational Media and Technology *2012*. DOI= http://www.aace.org/conf/edmedia/.

Analyzing Digital TV through the Concepts of Innovation

Paloma Maria Santos[1], Airton Zancanaro[1], Gregório Varvakis Rados[1], Klaus North[2], Fernando Spanhol[1]

[1]Post-Graduate Program in Knowledge Management and Engineering – Universidade Federal de Santa Catarina (UFSC)
Campus Universitário, Trindade, Florianópolis/SC, Brazil.
[2] Wiesbaden University of Applied Sciences

paloma@egc.ufsc.br, airtonz@egc.ufsc.br, grego@deps.ufsc.br, k.north@gmx.de, spanhol@led.ufsc.br

ABSTRACT
Innovation has been discussed intensively in the academy, government and in the market. The capability for innovation in a country or territory flourishes in the relationship between economic, political and social actors. To foster the dissemination of the innovation represented by Digital TV is a challenge in Brazil, and demands understanding of innovation and the characteristics involved in innovation processes. This article aims to analyze Digital TV based on the concepts of innovation, and provides a bibliographical review. The conclusions present important technical, economic and institutional factors, both to stimulate the adoption of Digital TV technology and to restrict its use. By looking at the impacts of Digital TV in the Brazilian economy it is possible to realize that major changes are to come in the future especially regarding economic and social issues.

Categories and Subject Descriptors
A.1 [Introductory and Survey].

General Terms
Management.

Keywords
Innovation, Digital Television, Brazilian Scene.

1. INTRODUCTION
The process of digitization of TV signals is due to the natural course of technological evolution. Along with forecasting a better image and sound quality, Digital TV provides the viewer with the possibility to have a more participatory position with the TV, through the adoption of interactive features. In addition, future viewers will have more channels available in their programming schedule, credited to the implementation of multiprogramming.

In this new scenario, Digital TV can be considered as an innovation as it decisively influences the direction of the industry in which it is inserted. It is an innovation that offers changes to the way products and services will be distributed to users.

Such technological innovation will almost certainly produce innovations in business processes. This means more activities and added value on services and products received by customers. Regarding Digital TV, these changes require a more dynamic view of business processes mainly due to the inclusion of new actors and stakeholders to the new business model.

This article aims to understand Digital TV as an innovation and discuss the implications and impacts of using such a concept related to this technology.

To this end, Section 2 addresses the path from idea to innovation. Section 3 describes the types of innovation according to the Oslo Manual. Section 4 explains the degree of novelty inherent to innovation. Section 5 presents the diffusion process and the factors that induce technological change. Section 6 deals with the analysis of Digital TV from the concepts of innovation and, finally, in section 7, the conclusions of the work are shown.

2. FROM IDEAS TO INNOVATION
Ideas move the world. Simple or complex, the ideas begin with the recognition and understanding that, at some point, there is a gap to be explored [1]. The ideas come from the creative capacity of individuals.

Creativity is understood by Silva [2] as the potential to generate ideas that appeal to preferences, establish smart strategies, modify products, seek solutions to problems, avoid the conventional and, therefore, to differentiate themselves. Creativity is the element that adds value and brings competitive advantage to organizations.

According to Prahalad and Hamel [3] the real sources of competitive advantage resides in the core competencies of organizations, as reflected in "the ability to consolidate corporate technology and production skills into competencies that can enable individual businesses to quickly adapt to constantly changing opportunities".

In this sense, we can say that idea is the product of creativity; it is the engine of innovation [1], which does not mean that every idea necessarily results in innovation. As highlighted by Silva [2], countless brilliant ideas are forgotten at the bottom of the drawer. "The idea is only transformed into innovation as it is executed, achieved." The issue is to know how to choose the right ideas and actually implement them [1].

Silva [2] states that the path from idea to innovation passes through four sequential phases, namely: identifying their problems, challenges, opportunities and dreams; generate ideas; evaluate ideas; and make it happen. Therefore, to explain this process, the author works with the representation of concepts linked to each letter of the word "ideia" (idea in Portuguese): I (initiate, anticipate the future), D (define what you want), E (explore your ability to think differently), I (imagine), A (act).

But what is innovation? Innovation is the ability to use knowledge to improve or create new products, processes or services bringing a competitive advantage to an organization. It is creating new opportunities through a combination of knowledge put into practice [4] [5].

According to Tidd Bessant and Pavitt [5], innovation occurs from the ability to detect opportunities, establish relationships and

make the best use of them. This can be considered the fundamentals of the dynamic capabilities of innovative companies, such as the ability of perception, capture and transformation.

Tidd, Bessant and Pavitt [5] point out that one of the problems of innovation management is that innovation is often mistaken for invention. While invention relates to the creation of a process, technique or a new product, innovation occurs with the effective practical application of an invention. Thus, there is no innovation without invention [6].

And why innovate? Silva [2] argues that as well as obtaining competitive advantage, companies who dedicate to innovation can increase their margin of profit, improve performance, take advantage of opportunities, obtain better returns, and help the development of the economy and job creation.

The limits to innovation are nothing more than the barriers we find to express our ideas. The main obstacle is that many companies look at innovation as something complex. Among the many barriers created for innovation, fear, lack of knowledge and negative associations are considered the three most frequent [2].

3. TYPES OF INNOVATION

The Oslo Manual [7], which is the conceptual and methodological reference most applied to analyze the innovation process, monitors four types of innovation:

- **Product innovation**: refers to significant changes in the potentialities of products/services that a company offers and they may be completely new or improved from existing products.
- **Process innovation**: refers to significant changes in how the products/services are produced and distributed.
- **Organizational innovation**: refers to changes in the company's management structure and organizational methods.
- **Marketing innovation**: refers to changes in marketing methods, including changes in product design and packaging, in product promotion and placement, and methods of pricing goods and services.

3.1 Technological Innovation vs. Business Model Innovation

Frequently when we talk about innovation, people usually think of technological innovations. According to Reis [8], the technological product and process innovation focuses on activities of Research and Development (R&D) and can be defined as the implementation of new technological knowledge, which results in products, processes or services which are technologically new, or significant improvement in some of their attributes.

Nevertheless, business model innovation is almost as important as technological innovation. Business model innovation focuses on strategic activities presenting reflections on commercial ventures and also in industry. According to Davila, Epstein and Shelton [1], business models define the way an organization creates, sells and transfers value to its customers. For the authors, the successful innovator is one who knows how to change, the models of business and technology, individually and in a group. The innovation matrix, illustrated in Figure 1, displays the interaction between technological innovation and business model.



Figure 1 – Innovation Matrix [1].

According to Davila, Epstein and Shelton [1], the success of an innovation depends on the integration of technological change and the business model. The concentration in only one of the two aspects will not produce a successful innovation or have a perspective of continuity.

The authors work with six levers, areas in which change can guide the innovation of technology and business model, shown in Figure 2, and highlight that innovation requires improvements in at least one of them.



Figure 2 – Levers for change [1].

The definition of each of the levers involves [1]:

- **Value proposition**: what is sold and released in the market;
- **Supply chain**: how it is created and brought to the market;
- **Target customer**: who receives the transfer of value;
- **Products and service offerings**: the product/service itself;
- **Process Technology**: production and delivery of better, faster and less costly services;
- **Enabling Technology**: faster strategy execution and better leverage time.

4. DEGREE OF NOVELTY INVOLVED

Not all innovations are treated equally. By definition, every innovation must contain a degree of novelty, which may be incremental (for the company), semi-radical (for the market) and radical (for the world), as discussed below.

4.1 Incremental innovation

According to Davila, Epstein and Shelton [1], incremental innovation seeks to make minor improvements through changes in both, technology and model of small business, in products and services that already exist in order to sustain market share and profitability of the product for a longer period of time. This type of innovation is also characterized by the OECD [7] as an innovation for the company, because it can utilize an organizational method of production, processing or marketing that may have been used by other companies, being new to the company that is using it now.

4.2 Semi-radical innovation

The semi-radical innovation involves substantial changes in one of the two fronts, technology or business model of an organization. This type of innovation depends on a set of core competences and leverages crucial changes in the competitive environment, which are not allowed by incremental innovation [1]. Semi-radical innovation is characterized by OECD [7] as innovation for the market, because it emphasizes the fact that the company is the pioneer introducing such innovation in its market, which may include a territory or a line of product.

It may happen that the change in one dimension depends on the change in another, although this concomitant change may not be as revolutionary as the first [1].

4.3 Radical innovation

Radical innovation often results from architectural changes and is characterized by presenting significant changes that affect both the technology and business model of a company. Radical innovation usually causes significant impact on the competitive landscape of an industrial sector. It has greater impact on and inherent risk degree to the organization than the incremental and semi-radical innovation [1].

This type of innovation is characterized by the OECD [7] as an 'innovation for the world', as the company is the first to introduce the innovation in all markets and industries, nationally and internationally. When successful, radical innovation has the potential to rewrite the rules of competition in the industry [1].

5. DIFFUSION OF INNOVATION

As well as the degree of novelty involved, there is the way by which an innovation is disseminated in the economy, from its first introduction to the different public (consumers, countries, regions, sectors, markets and companies). Without diffusion, an innovation has no economic impact [7].

For Tigre [6], when evaluating the process of diffusion of technological innovation, it can be understood as the course of the adoption of this technology in the market, focusing on their technological characteristics and other elements that condition pace and direction.

Thus, this diffusion process can be examined from four factors to induce technological change [6]:

Direction or course of the technology: refers to the technical choices made during the evolution of the technology. In the case of process innovation, this course may present:

Incremental technological changes: the most basic level of technological change, including improvements in design or the quality of products and processes, new organizational and logistical arrangements.

Radical technological change: usually results from R&D, being discontinuous in time and sectors. Such change disrupts the existing paths and commences a new technological route.

Technological system change: involves the transformation of one or more sectors due to the emergence of a new technological field. The innovation is accompanied by organizational change both within the company as well as in its relationship with the market.

Changes of Techno-economic paradigm: embraces technology innovation and innovation in the social and economic fabric in which they are inserted. Although they do not occur often, their influence is pervasive and long-lasting.

Diffusion rate or diffusion speed: refers to the speed in which the society adopts the technology, measured by the number of adoptions over time within the potential universe of users.

Conditioning factors: refers to factors that can both stimulate the adoption of technology or restrict its use. They may be of a technical (regarding the degree in which the innovation is considered hard to be understood or used), economic (related to the costs of the acquisition or implementation of the new technology) or of an institutional nature (regarding the social stratification, culture, religion, regulation, laws of the sector or country, including the availability of human capital and support institutions).

Economic, social and environmental impacts: refers to changes in the industrial structure, the quantity and quality of work and the relationship between technology and environment.

According to Tigre [6], the diffusion process leads to greater economic impacts than innovation itself, in that it "represents the effective adoption of a new technology by broader segments of society". In fact, the exact pattern of adoption of an innovation is directly related to the interaction of factors associated with supply and demand [5].

6. ANALYZING DIGITAL TV FROM THE CONCEPTS OF INNOVATION

Digital TV can be considered an innovation in that it definitely influences the direction of the industry in which it operates. Its diffusion brings positive and negative consequences to different sectors of the economy and society.

From the moment we interpret Digital TV as a technological innovation process – which offers significant changes in how the products/services are produced and distributed to customers (in this case, the viewer) – we understand that it is important to discuss the course of adoption of this technology in the market, as well as other factors that influence its pace and direction.

As said by Davila, Epstein and Shelton [1] it is very rare for a technological innovation not to produce innovations in business processes. The opposite is true. These innovations go together and should be considered as well as implemented as a whole.

Santos [9] points out that the model of Interactive Digital TV, in general, brings significant changes in traditional business processes, created for the analogue environment. When applied to the digital environment, these changes require a more dynamic view of the business processes mainly due to inclusion of new actors and stakeholders to the new model.

In this context, we characterize the degree of novelty brought by Digital TV as radical, according to the Innovation Matrix proposed by Davila, Epstein and Shelton [1]. This classification is due to the fact that both the technology and business model have different characteristics from that of the one currently used, being therefore new.

Specifically when it comes to technological innovations, Reis [8] argues that the inclusion of such technology will not necessarily lead to the complete abandonment of earlier technology. This is exactly what is happening with Digital TV.

Tidd, Bessant and Pavitt [5] say that the velocity of the diffusion of a technological innovation is usually drawn in an "S" curve,

with phases of introduction, growth, maturity and decline. First, the adoption rate is relatively low, and the "innovation is restricted to those so called 'innovators'. The next to adopt are the 'early adopters', followed by the 'late majority' and finally the curve falls, and only the 'laggards' remain".

The direction or technological course of Digital TV is embedded in changes in the technological system. The natural evolution from the analogical to digital television system brings a new emerging field of activity.

Analyzing the economic, social and environmental impacts, the diffusion of Digital TV in the Brazilian economy presents great changes, especially regarding the economic and social impact. From an economic viewpoint, there are several possible impacts on the devolution of the communication industry. The digitalization of TV facilitates the entry of new companies in the market as it removes the exclusivity of the broadcasting companies (open or closed) in providing television content.

Moreover, when the Internet and other telecommunication companies enter this scenario, multiplying the program schedule and eliminating the limitations presented by the current allocation of the frequency band for analogue transmissions, new business models will be established and a new course of innovation will come into the market [6].

Given the importance that the conditioning factors have, since they are both stimulating adoption of technology and restriction of its use, we will explore in more detail each one of their natures: technical, economic and institutional.

6.1 Technical Condition

Regarding the technical conditioning, on risk analysis, certainly the adoption of a new technology is the factor that impacts the most on its success, especially when it comes to a technological transition such as the one that is currently occurring with Digital TV.

One of these risks is usability. Television is considered as a well-known device by the population; it is very familiar and is easy to use. With Digital TV, the remote control, which was previously only used to change channels, turn on and off and control the volume, is appended with new features, such as accessing the interactivity available in a given program.

In this case, one runs the risk of users being unable to fully use the technology. Such users may not grasp the features of the convergence between TV, internet and phone in one device.

The main challenge of Digital TV will be of offering this accessibility to the handicapped, elderly and children who spend much of their time watching television. For this reason it is important to identify and offer content ensuring that all users are able to use the interactivity with autonomy.

When it comes to interactivity, this feature will offer interactive content directly on the television. But to make this possible, it is necessary for converters with Ginga middleware to be available on the market at an affordable cost for the population.

When thinking of interactivity, we find that since the television would be for common use, there is a complicating factor when one of the group members wants to interact with the program because it will disturb others who want to passively watch the program. To resolve this problem, Costa, Moreno and Soares [10] propose the usage of multiple devices.

Software available on multiple devices such as tablets and smartphones offer interactive content related to the program the user is watching, via set-top-box or the connected TV. For instance, users can watch their favorite program and, using a tablet, they can purchase some of the products announced during the show.

Thus, the user will be modifying the way to watch TV, but not the experience of watching TV. The secondary screen to allow individual interaction promises to be a new way to use the TV, as a vehicle for integration and communication, and this is already happening.

However, despite the presented scenario, interactive applications for Digital TV are still in testing and are still restricted to small niches, such as sports. Withal, to expedite this process and provide the consumer with an innovative product, the TV manufacturers were able to envision a business model for interactivity before the broadcasters. It is Connected TVs also known as Broadband TV, which combine the Web and television.

6.2 Economic Condition

The implementation process of Brazilian Digital TV requires investment by both television stations and users. The television stations will invest substantially in the digitalization of broadcasting with the purchase of cameras and digital editing stations, as well as transmitters and amplifiers. While the users, to receive digital signals and use the new services, should purchase a set-top-box with embedded Ginga middleware or a TV with this equipment already internally installed.

Currently in Brazil, there is a lack of public policies that favor the industry that produce converters, prepared for interactivity, available for sale in the market at an affordable cost for classes C and D of the population. There are still no incentives for the development of an industry of local and regional content, stimulating researches that have as their objective the production of content and services for education, electronic government, entertainment and news, using, in this case, the structure of transmission of channels of Public, College and Community TV.

Moreover, the population does not know exactly what Digital TV is, how to install it, what equipment is needed and what can be done with such technology. There is a lack of intense and permanent clarification campaigns in the general means of communication, public transport, airports, among others, giving citizens the possibility to assist the government in determining the fate of Digital TV in Brazil.

6.3 Condition of an institutional character

The society needs to benefit from the implantation of Digital TV. Otherwise it does not make sense to purchase a set-top-box to have access to the digital signal. This benefit can occur in three ways: improving the quality of audio and video, increasing the number of channels, and the possibility of interactivity. The first item is intrinsic to the digitalization process, but the other two lack public policies to materialize.

Public policies should be thought of as incentives for content production in disadvantaged communities, which promotes identity preservation, increased self-esteem, as well as employment opportunities.

In order to implement these policies, there is need for graduate and postgraduate courses, encouraging dialogue between different

areas considering the demand for generation and dissemination of knowledge and innovation in television.

In addition, professors and students need to be trained beyond the area of engineering and computer science, getting involved in research and production of content and interactive services. In this context, the benefit of Digital TV for the people includes the understanding of what interactivity is and how to use it in an attractive and easy way.

7. CONCLUSION

The aim of this paper was to understand Digital TV as an innovation and from that, consider all the conditioning related to this concept. Considered as technology process innovation, Digital TV features significant changes in its business processes, especially considering the presence of new actors in the supply chain. In this regard, there are several factors that work both to stimulate technology adoption and restrict its use.

The importance of discussing the process of the adoption of Digital TV in Brazil is due to the feeling that there are many factors that are cluttering up the process.

The diffusion of Digital TV is affected by the political and economic factors addressed in this article, but also by the perceived level of complexity, the ability for experimentation, the compatibility with the values, experiences and needs of users and by the degree to which the innovation is perceived as better than the product it replaces.

It is observed that to have the full diffusion of innovation provided by Digital TV and minimize the difficulties involving the implementation of new technology there is the need for public policies that address the issue more objectively.

The wait for the advancing of the technology and the perceived lack of government commitment is delaying research and the provision and adoption of the benefits offered by this new technology for users.

8. REFERENCES

[1] DAVILA, Toy; EPSTEIN, Marc J.; SHELTON, Robert. As regras da inovação. Porto Alegre: Bookman, 2007.

[2] SILVA, Antônio Carlos Teixeira da. Inovação: como criar idéias que geram resultados. Rio de Janeiro: Qualitymark, 2003.

[3] PRAHALAD, C.; HAMEL, G. The core competencies of the corporation. Harvard Business Review, mai-jun, 79-91, 1990.

[4] NORTH, Klaus. Gestão da inovação, 21-24 de mar. de 2011. Notas de aula.

[5] TIDD, Joe; BESSANT, John; PAVITT, Keith. Gestão da inovação. 3rd. ed São Paulo: Bookman, 2008.

[6] TIGRE, Paulo Bastos. Gestão da inovação: a economia da tecnologia do Brasil. Rio de Janeiro: Elsevier, 2006.

[7] OCDE. Oslo Manual: guidelines for collecting and interpreting innovation data. 3rd ed. Paris: OECD Publishing, 1997. 184 p. Available on: <http://www.finep.gov.br/dcom/brasil_inovador/arquivos/manual_de_oslo/prefacio.html>. Acessed on: 22 jul. 2011.

[8] REIS, Dálcio Roberto dos. Gestão da inovação tecnológica. 2nd. ed. Baueri: Manole, 2008.

[9] SANTOS, P. M. Modelagem de processos para disseminação de conhecimento em governo eletrônico via TV Digital. Dissertação de Mestrado. Departamento de Engenharia e Gestão do Conhecimento. Universidade Federal de Santa Catarina, Florianópolis, Brasil, 2011.

[10] COSTA, R. M. D. R.; MORENO, M. F.; SOARES, L. F. G. Ginga-NCL: Suporte a Múltiplos Dispositivos. Simpósio Brasileiro de Sistemas Multimídia e Hipermídia. Anais do XV Simpósio Brasileiro de Sistemas Multimídia e Hipermídia. Fortaleza. 2009.

The Evolution of Television in Brazil: a Literature Review

Paloma Maria Santos, Marcus de Melo Braga, Marcus Vinícius Anátocles da Silva Ferreira, Fernando José Spanhol
Post-Graduate Program in Knowledge Management and Engineering – Universidade Federal de Santa Catarina (UFSC)
Campus Universitário, Trindade, Florianópolis/SC, Brazil
paloma@egc.ufsc.br, marcus@egc.ufsc.br, marcus.ferreira@hotmail.com, spanhol@led.ufsc.br

ABSTRACT
Since its emergence in the 1950s, television has had the power to congregate the family around it, generate discussions and above all offer moments of leisure, information and entertainment to its viewers. With the advent of Digital TV, the technical capacities of the televising system are extended, as a wide array of interactive resources can be offered to the population. Considering this perspective, this paper aims to present an overview of the evolution of television in Brazil based on a literature review. As a result of this analysis, one may notice that Digital TV has all the necessary potentialities to turn the viewer into an active participant, but a great deal still needs to be done to transform this technology into a tool of social inclusion that is established in the institution decree of the Brazilian system.

Categories and Subject Descriptors
A.1 [Introductory and Survey].

General Terms
Documentation.

Keywords
Digital Television (iDTV), TV Evolution, Literature review, Brazil.

1. INTRODUCTION
Brazil is a country with continental dimensions[1] that still preserves extensive areas of forests and Atlantic vegetation. This great land mass, combined with a high concentration of population in urban centers, must be taken in consideration in any initiatives related to Information and Communication Technology (ICT). The most recent population census, in 2010, registered 190 million inhabitants in Brazil [1].

In spite of these dimensions, open television reaches about 97% of the urban population, thus becoming the means of communication with the largest penetration in Brazilian homes [1]. As such, television is an instrument of national integration. For the consumer, it is the only free network of telecommunications that has national coverage almost 24 hours a day [2].

These figures contrast with those of cable TV in Brazil. According to [3], cable TV represents only 4% of penetration in the Brazilian context. With regard to the Internet, penetration in the Brazilian homes is low, if compared to open TV - only 27% [4].

These figures show that, among the modalities of access to television, open TV is of bigger relevance to the population as a whole. This information can be seen as a good justification for investments in the Brazilian System of Terrestrial Digital TV, since they will be able to bring benefits to a larger slice of the population.

2. TV HISTORY IN BRAZIL
The history of television in Brazil is associated with the enterprising spirit of Assis Chateaubriand, a journalist, entrepreneur, lawyer, politician and diplomat. Chateaubriand was one of the more influential public personalities in Brazil in the 1940s and 1950s. He served a term as a senator between 1952 and 1957, and was also a law professor and writer [5].

Television in Brazil had its origins in two experimental broadcasts: a closed-circuit broadcast in 1939 during the Feira Internacional de Amostras[2] in Rio de Janeiro, and another broadcast in 1948 during the celebration of the centenary of Juiz de Fora city in Minas Gerais state. People from the city of Rio de Janeiro then got a taste of what was to come; on July 29, 1950, TV finally arrived in Brazil. Some stores and advertisers in Rio de Janeiro sold imported television sets from the United States, bought and distributed by Assis Chateaubriand himself [7].

Almost two months later (September 18, 1950), television was officially inaugurated in Brazil by Chateaubriand, but this time in the city of São Paulo [7]. Latin American pioneer[3] PRF-3 (Tupi Difusora, Channel 3) went on the air unscripted, with transmitters bought from the Radio Corporation of America (RCA), in the United States.

In the following year (January 20, 1951), without enough technical structure, TV Tupi (Channel 6) emerged in Rio de Janeiro. (Later this TV station would become responsible for training new technicians who would constitute the body of employees in Chateaubriand's group of TV stations.) The studios did not have soundproofing and scenes were often invaded by ship whistles and the noise of cars, since the

[1] 8,5 million km^2

[2] An International Fair

[3] According to [6], Tupi was not the first TV station in Latin America. Tupi aired its first TV program 18 days after Mexican TV had done so.

windows had to remain open to prevent overheating equipment [7]. "Tupi, both in Rio de Janeiro or São Paulo, was the university of Brazilian television", as everybody there would learn together in scenes full of improvisation. "Everything was a question mark" [7].

In 1955, the second TV station in Rio de Janeiro–TV Rio, Channel 13–went on air. In the following years, Chateaubriand added new TV stations to his empire, introducing them all over Brazil [7]. Thanks to advertising agencies in Rio de Janeiro and São Paulo, and also to so-called ad brokers, television started to gain customers and signal its presence in the market.

Still in the 1950s, several other networks start to emerge [8]: São Paulo TV (1952), TV Record (1953), TV Itacolomi from Belo Horizonte (1955) and TV Excelsior (1959). In 1956, TV stations in Porto Alegre, Curitiba, Salvador, Recife, Campina Grande, Fortaleza, São Luís, Belém and Goiânia were authorized and inaugurated. By the end of the 1950s, the city of São Paulo had approximately two thousand television sets, belonging to wealthy families who brought these TV sets from abroad. In 1951, São Paulo and Rio de Janeiro inhabitants together possessed a total of seven thousand television sets, and by 1952 the number had gone up to eleven thousand. According to [8], the value of a television set was equivalent to three times the value of the most sophisticated radio, and a little cheaper than a car.

In 1963, the first color TV sets imported from the United States arrived in Brazil, but it was only on February 19, 1972 that the first official color transmission in the country occurred, with the coverage of Feast of the Grape in Caxias city in Rio Grande do Sul state[8].

Different from other countries, the standard system chosen for analog transmission in Brazil was PAL-M. This system is a variation of the Phase Alternate Line (PAL) system for color television, developed in Germany and used for the first time in 1967. PAL-M uses a frequency of 60 Hz, and has 60 fields, 525 lines and 30 pictures per second. It differs from NTSC (National Television Standard Committee), the system for color television developed in the United States, in that NTSC has a chrominance frequency of 3.579545 MHz and PAL-M of 3.575611 MHz [9].

On April 26, 1965, TV Globo was inaugurated in Rio de Janeiro and, on the same day at 11:00 a.m., TV Globo São Paulo went on air through Channel Five. On July 17, 1980, after months of strikes, various dismissals and having its concession annulled by the government, TV Tupi closed down its broadcasting. From its inauguration to this fateful day, there was no single month in the history of TV Tupi when the employees had been paid on time. "Tupi was a pioneer in everything, including in the delay of wages, (...) which never arrived complete in our hands" [7].

The federal government then announced, on July 23, 1980, the opening of a competition for two new TV networks that emerged out of the seven concessions belonging to TV Tupi. Apart from these stations, there were also other two TV stations that had gone extinct in the beginning of the 70s. The new concessionaires were the executives Silvio Santos and Adolpho Bloch [8].

The 1990s were characterized by the first cable TV concessions and the beginning of high-definition broadcasts in the United States (1991) and Japan. In Brazil, the first experimental high-definition broadcast was carried out in 1998 by Rede Globo.

Sixty years after the beginning of analog broadcasting in the Brazilian television system, today we are crossing a decisive phase of renewal, with the gradual implantation of the Digital TV system. As in other countries, the digitizing process of the Brazilian television system is gaining force and visibility. The system of analog TV broadcasting in Brazil will last until 2016, pursuant to art.10 of decree n º 5,820 of June 29, 2006 [10].

December 2, 2007 marked the beginning of Digital TV broadcasting in the country. Apart from improving the quality of sounds and images, now without "statics", "ghosting" or motion blurs, Digital TV opens up the possibility of the viewer assuming a more active position in front of a television set, choosing among various additional options in the program being broadcast. Additionally, the viewer can exchange information with the TV station, a reality that is aligning TV sets with other devices and information systems.

This evolution from an analog model to a digital one involves the digitizing process and the adoption of interactivity resources. Apart from configuring the technological evolution of a system, Digital TV arrives with a perspective on diminishing social exclusion, offering the viewer interactive resources that can meet the needs and expectations of different audiences, propitiating new forms of expression, supporting the intercultural dialogue and promoting social mobility.

Today, given the importance that this medium has in people's lives, television is seen as a basic piece of equipment in Brazilian homes, with 97% penetration, lagging behind the stove with 98.5% [1].

3. THE BRAZILIAN TERRESTRIAL DIGITAL TV SYSTEM (SBTVD-T)

SBTVD-T, also known as the Nippo-Brazilian system[4], is a derivation of the Japanese Digital TV model. Apart from the inherent characteristics of the digital system itself, which are better quality of sound and image and better exploitation of the frequency spectrum, the Brazilian model, through technological advances, extends the Japanese model and adds new functionalities to it such as the use of an exclusive middleware and more updated compression standards, thus making the Brazilian model the most modern system among existing ones.

Owing to these technological advances, the Brazilian model offers as a main competitive advantage the possibility of

[4] Countries that have adopted the Nippo-Brazilian standard, apart from Brazil, are: Argentina, Bolivia, Chile, Costa Rica, Ecuador, Paraguay, Peru, the Philippines, Uruguay and Venezuela.

interactivity, mobility and portability. According to [2], while portability enables the reception of the Digital TV signal in cellular and other mobile devices, mobility is associated with the reception of the signal during the movement of the receiver, enabling access to Digital TV in any place and at any time [11].

Interactivity, meanwhile, is undoubtedly the greatest advantage of Digital TV. It is the key for the receiver's access to the world of production and the sharing of content and knowledge by means of television [12].

SBTVD-T was created to guarantee digital inclusion by exploring interactivity resources that enable future access to the Internet and the democratization of access to information. Although these features are still under construction, SBTVD-T's technical features enable the design of various applications in diverse areas of knowledge.

By exploring the Brazilian model of interactivity resources, a lot of applications and services can be created. It is possible to foresee that soon the same bidirectional interactivity resources that exist today on the Internet will also be available for Digital TV.

4. TRAINING OF HUMAN RESOURCES – THE RHTVD PROJECT

The UFSC (Universidade Federal de Santa Catarina) participation in the Brazilian program of Digital TV had already been registered when the NTDI (Núcleo deTelevisão Digital Interativa) participated in a project entitled "Requisito Formal de Proposta" (RFP) 6 in 2004, in which the aim was to generate multimedia content on the theme of "health" that could be accessed by common viewers as well as health professionals in Santa Catarina state.

This work was marked by the interdisciplinary nature of specialists coming from different areas such as telejournalism, software development, medicine, engineering, healthcare and design [13]. Various programs were produced for SBTVD such as: Live More, Depression Test, Spider Bite, Obesity and Obesity Test. All these programs aimed to explore the possibilities of interactivity for Brazilian Digital TV.

The project got financial resources from the Brazilian agency FINEP (Financiadora de Estudos e Projetos), in accordance with the invitation letter of MC/MCT/FINEP and FUNTEL [14]. The aim was to finance and "implement networks of academic cooperation in the country in the area of Digital TV, enabling and stimulating the production of scientific and technological research and the education of post-graduated human resources on the subject" [15].

In order to make the Brazilian Digital Television System viable, in 2006 the Brazilian government invested in the education of human resources by creating the project "Formação de Recursos Humanos para TV Digital (RHTVD) com Foco em Conteúdo e Serviços" controlled by the Brazilian research funding agency CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). The Postgraduate Program in Engineering and Knowledge Management (EGC) of UFSC was one of the institutions contemplated by CAPES.

As a result of this project, the Sambaqui[5] Digital TV research group of EGC/UFSC was created in 2008. This research group aims "to produce and to spread knowledge by means of contents and services for the Digital TV with a focus on the inclusion of Brazilian society in the economy of knowledge" [16], acting in an interdisciplinary way.

In this sense, each participant of the research group seeks to focus their academic research in order to foster the development of contents and services for Brazilian Digital TV. The Sambaqui group, in a period of only four years, has produced eight master theses and two PhD dissertations, contributing to the education of researchers on the subject of Digital TV in Brazil.

5. FINAL CONSIDERATIONS

Despite the fact that Digital TV migration in Brazil has already been underway for two and a half years, its incorporation is still very minimal; this is a challenge assumed by the federal government.

The majority of users still do not know what advantages Digital TV will make available, apart from a better image without ghosting or statics, as the resources offered until now have been limited to these types of improvement. The availability of a friendly and intelligent tool that accumulates attractive, useful and easy-to-access resources is indispensable for the success of digital TV. Moreover, the interaction process does not have to incur costs for the user, which would risk preventing the user's participation in the process.

It is a government priority to facilitate citizens' access to information and governmental bodies in a quick, free and democratic way, strengthening the relations between them. This process of approximation and inclusion permeates digital literacy.

Citizens, armed with the tools needed to condition them to this access, need to move on to the next stage; that is, to be able to deal with such tools. This process is related to the qualification of the citizen within the process of inclusion and digital education (digital literacy).

As citizens feel more confident and qualified, they start to be more active; they are not only consumers of information anymore. They start to generate knowledge, changing their status from mere users to become partners.

Much more than electronic development that promotes access to the government, it is necessary to educate and prepare people for the communication society, contributing to the speed of economic development and at the same time creating new channels of discussion and exchange of opinions.

The government needs to continue investing in public policies that condition the viewer to be part of this digital network, not only in terms of making these tools available, but also in terms

[5] http://tvdi.egc.ufsc.br/

of qualification. It is only by encouraging and fostering the reduction of barriers that users will be part of the collaborative network of knowledge construction.

6. REFERENCES

[1] IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Síntese de Indicadores Sociais: uma análise das condições de vida da população brasileira 2010. Available at: <http://bit.ly/9HUTvI>. Accessed in May, 2011.

[2] ZUFFO, M. K. TV Digital Aberta no Brasil: Políticas Estruturais para um Modelo Nacional, 2006. Available at: <http://bit.ly/oqRUjh>. Accessed in August, 2008.

[3] ANATEL – Agência Nacional de Telecomunicações. Panorama dos Serviços de TV por Assinatura. Brasilia, 40 ed, 2010.

[4] CGI – Comitê Gestor da Internet no Brasil. TIC Domicílios e Empresas 2010 – Pesquisa sobre o uso das Tecnologias da Informação e Comunicação no Brasil, 2010. Available at: <http://bit.ly/vdo1Ic>. Accessed in August, 2011.

[5] ENCICLOPÉDIA ITAÚ CULTURAL. Assis Chateaubriand: biografia, 2007. Available at: < http://bit.ly/wcQRZS>. Accessed in April, 2010.

[6] XAVIER, R.; SACCHI, R. Almanaque da TV: 50 anos de memória e informação. Rio de Janeiro: Objetiva, 2000.

[7] LORÊDO, J. Era uma vez a televisão. São Paulo: Allegro, 2000.

[8] VALIM, M.; COSTA, S. História da TV. Televisão: Tudo sobre TV, 2010. Available at: <http://bit.ly/59jVg8>. Accessed in April, 2010.

[9] PIZZOTTI, R. Enciclopédia básica da mídia eletrônica. São Paulo: Senac, 2003.

[10] BRASIL. Decreto n° 5.820, de 29 de junho de 2006., 2006. Available at: <http://bit.ly/FPqhn1>. Accessed in November, 2008.

[11] BITTENCOURT, F. TV aberta brasileira: o impacto da digitalização. TV digital: qualidade e interatividade. IEL/NC. Brasília. 2007.

[12] ZANCANARO, A.; SANTOS, P. M.; TODESCO, J. L. Ginga-J ou Ginga-NCL: características das linguagens de desenvolvimento de recursos interativos para a TV Digital. In: I Simpósio Internacional de Televisão Digital. 2009. p. 1084-1108.

[13] WANGENHEIM, A. V. Sistema Brasileiro de Televisão Digital: Especificação de testes de integração RFP 6 - Serviços de Saúde. Universidade Federal de Santa Catarina. Florianópolis. 2005.

[14] NTDI. Sistema Brasileiro de Televisão Digital Interativa. Núcleo de Televisão Digital Interativa, 2011. Available at: <http://bit.ly/FOYrv3>. Accessed in March, 2011.

[15] CAPES. Programa de Formação de Recursos Humanos em TV Digital. CAPES (Coordenadora de Aperfeiçoamento de Pessoal de Nível Superior), 2011. Available at: <http://bit.ly/Asraav>. Accessed in March, 2011.

[16] SAMBAQUI. Planejamento Estratégico. TVDI SAMBAQUI - Grupo de pesquisa TV Digital EGC/UFSC, 2011. Available at: <http://bit.ly/hDFH0J>. Accessed in March, 2011.

Using a Voting Classification Algorithm with Neural Networks and Collaborative Filtering in Movie Recommendations for Interactive Digital TV

Ricardo Erikson V. de S. Rosa[1]
Federal University of Minas Gerais
Belo Horizonte - MG, Brazil
ricardoerikson@ufmg.br

Vicente Ferreira de Lucena Junior[2]
Federal University of Amazonas
Manaus - AM, Brazil
vicente@ufam.edu.br

ABSTRACT
Many times users are faced with selecting and discovering items of interest from an extensive selection of items. These activities seem to be quite difficult in Interactive Digital TV (IDTV) given the large amount of available alternatives among TV programs, movies, series and soup operas. In this scenario Recommender Systems arise as a very useful mechanism to provide personalized and relevant recommendations according to user's tastes and preferences. In this paper, we describe a method for providing movie recommendations to IDTV users by using neural networks and collaborative filtering. We use a voting classification algorithm to improve recommendations. The approach described in this paper is used to predict user ratings. The method achieved above 80\% of good recommendations for users with representative data samples used during training.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Search and Retrieval – *Information filtering*. H.4 **[Information Systems]**: Information Systems Applications

General Terms
Algorithms, Design, Experimentation

Keywords
Collaborative Filtering, Interactive Digital Television, Neural Networks, Recommender Systems

---

[1]Graduate Program in Electrical Engineering - Federal University of Minas Gerais - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil.

[2]Professor at UFAM-PPGEE. Electronics and Information Technology R&D Center (CETELI).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
EuroITV'12, July 4–6, 2012, Berlin, Germany.
Copyright 2010 ACM 1-58113-000-0/00/0010…$10.00.

1. INTRODUCTION
Advances in Interactive Digital Television (IDTV) technologies have enabled a enhanced delivery of multimedia and interactive content through TV. As a result, users are supplied with an increasing amount of content offered by multiple content providers and are constantly confronted with situations in which they have too many options to choose from. This brings new and exciting challenges to the information processing research field in IDTV.

In recent times, recommender systems emerged as a solution to alleviate the problem of information overload. Recommender systems are aimed at building user models based on user preferences modeling, content filtering, social patterns, etc. These models are used to recommend the most relevant items to users while items considered as not relevant are removed or moved to a lower-ranked position [7]. Many companies such as Amazon.com and Google are aimed at using recommender systems for several reasons including keeping customer loyalty, boosting sales, attracting customers and increasing user satisfaction [6].

In IDTV scenario one of fields of application for recommender systems is to provide recommendations on audio visual content such as movies, programs, series, etc. The IDTV model has enabled a data transmission flow where the user can also transmit data to content providers by using IDTV applications. Now imagine the following scenario:

Scenario. *While watching some audio visual content such as TV programs, movies, series and soap operas the user x can access and see information (genre, age restriction, synopsis, characters, actors, etc) about this content. After the end of the program the user opens an application and gives a rate to that program. This rating is sent to the content provider in order to create a profile based on the preferences of the user x. After a certain number of evaluations it is possible to model the user profile and this profile is used to recommend new programs to the user. Once the content provider has a set of recommendations based on the user profile, these recommendations are sent to the user x's IDTV where he/she can choose and watch one of the recommendations.*

Knowing preferences and tastes of the users is essential to give good recommendations. The scenario described above presents a way where users can provide information about their preferences and tastes to content providers. In this way, content providers can gather data from many users. Once the user profile is known, the content provider can recommend programs, movies and other audio visual content that are liked by other users with similar tastes and preferences.

In this paper we propose an approach to recommend movies for IDTV users. The approach consists of: (1) predicting ratings of movies that the user has not yet seen or considered; and (2) using these ratings to recommend new movies. The approach uses Neural Networks and the bootstrap aggregating (Bagging)[1] algorithm in association with collaborative filtering methods to predict movie ratings. We ran the experiments through a dataset of movie ratings called MovieLens[4].

The remaining sections are organized as follows. In Section 2 we describe our approach and how we employed collaborative filtering and neural networks to provide recommendations. In Section 3 we present the dataset and the methodology of the experiments. We also present some experimental results that were achieved using the approach described in this work. The related works are presented in Section 4. We conclude the paper in the Section 5.

2. OUR APPROACH

In our approach we combine two techniques to provide good recommendations: (1) collaborative filtering based on Pearson's Correlation Coefficient (PCC); and (2) content filtering based on Neural Networks. These techniques were applied in a dataset containing user ratings (ranging from 1 to 5) on movies. In the following sections we describe how these techniques were employed and how they were combined to perform recommendations.

2.1 Collaborative Filtering-Based Recommendation

Collaborative filtering assumes that a good way of suggesting items that the user might be interested is finding other users with similar tastes [10]. Thus, collaborative filtering is based on items observed by other users aiming to predict user interest on items that were not observed yet.

According to [11], the most important step for collaborative filtering is to determine similarity level between two users such that it is possible to find a neighborhood to obtain relevant recommendations. One of the techniques used to determine the similarity level is the Pearson's Correlation Coefficient (PCC), which has proven performance for recommendations [9].

The PCC is defined by the Equation 1:

$$\rho(x,y) = \frac{\sum_{m \in I_{xy}} (R_{x,m} - \bar{R}_x)(R_{y,m} - \bar{R}_y)}{\sqrt{\sum_{m \in I_{xy}} (R_{x,m} - \bar{R}_x)^2 \sum_{m \in I_{xy}} (R_{y,m} - \bar{R}_y)^2}} \quad (1)$$

where I_{xy} represents the set of all movies that were rated by the users x and y. $R_{x,m}$ is the rating of the user x on the movie m and $\bar{R}_x$ is the average of user x's ratings. Equation 2 mathematically demonstrates $\bar{R}_x$, where I_x is the set of all movies that were rated by user x.

$$\bar{R}_x = \frac{1}{n} \times \sum_{m \in I_x}^{n} R_{x,m} \quad (2)$$

The correlation coefficient $\rho(x,y)$ is a decimal value ranging from -1 to 1, where -1 means strong negative correlation, 1 means strong positive correlation and 0 means the users x and y are independent from each other.

After obtaining the similarity levels among users, it is possible to determine the k nearest neighbors to predict the ratings of items that were not considered yet. The ratings are predicted by using the algorithm proposed in [8] and can be calculated according to the Equation 3.

$$\hat{R}_{x,m} = \bar{R}_x + \frac{\sum_{y \in NBS_x} (R_{y,m} - \bar{R}_y)\rho(x,y)}{\sum_{y \in NBS_x} |\rho(x,y)|} \quad (3)$$

where NBS_x represents the set of the k nearest neighbors (users with highest similarity levels) of user x.

A representative neighborhood can be obtained by using only the similarity criterion. However, in order to predict the rating of a specific movie (according to Equation 3), we are interested in users that also have explicitly rated the movie to be predicted. Thus, the neighborhood is restricted to users with highest similarity that also have rated the movie of interest.

To demonstrate the differences between the neighborhood selection approaches we present an example using MovieLens dataset. A neighborhood for user 1 was selected aiming to predict the rating for movie 1. Table 1 presents the 15 nearest neighbors that were selected by similarity criterion. From this neighborhood it is possible to observe that only the user 691 (similarity 0.823 and ranking 15) has evaluated the movie 1. Table 2 presents the 15 nearest neighbors that also have evaluated the movie 1. Although a user has a highly relevant neighborhood according to similarity criterion, most of the neighbors are removed because they do not have evaluated the movie that must be predicted.

Table 1 : Neighbors with highest similarity.

Ranking	$y \in NBS_x$	$\rho(x,y)$	$R_{y,m}$
1	685	1.000	*n.d.*
2	155	1.000	*n.d.*
3	341	1.000	*n.d.*
4	418	1.000	*n.d.*
5	812	1.000	*n.d.*
6	351	0.996	*n.d.*
7	811	0.988	*n.d.*
8	166	0.981	*n.d.*
9	810	0.901	*n.d.*
10	309	0.894	*n.d.*
11	34	0.887	*n.d.*
12	105	0.884	*n.d.*
13	111	0.868	*n.d.*
14	687	0.834	*n.d.*
15	691	0.823	5

2.2 Recommendations Based on Neural Networks

Artificial Neural Networks were used to perform content filtering based on movie genres. The main reasons to use neural networks in this approach were the learning capabilities and capacity of modeling complex relationships between input and output. The approach consists of building up a neural network for each user in order to predict which rating this user would give to a movie taking into account only genre.

The network training consists of using the sample of the dataset that contains only the movies that were rated by the user. Part of

this sample is used in training and the other part is used in testing. Each network has 19 inputs (19 genres), one hidden layer with 2 neurons and 5 outputs (ratings ranging from 1 to 5). Figure 1 illustrates the structure of the neural network used to predict user ratings.

Table 2 : Neighbors with highest similarity that also have rated the movie 1.

Ranking	$y \in NBS_x$	$\rho(x,y)$	$R_{y,m}$
15	691	0.823	5
25	134	0.764	5
31	895	0.745	4
33	550	0.744	3
38	549	0.727	5
43	769	0.684	4
48	199	0.651	1
50	274	0.645	4
53	730	0.627	4
54	697	0.616	5
57	800	0.610	4
60	66	0.603	3
62	526	0.601	5
68	923	0.589	3
71	120	0.584	4

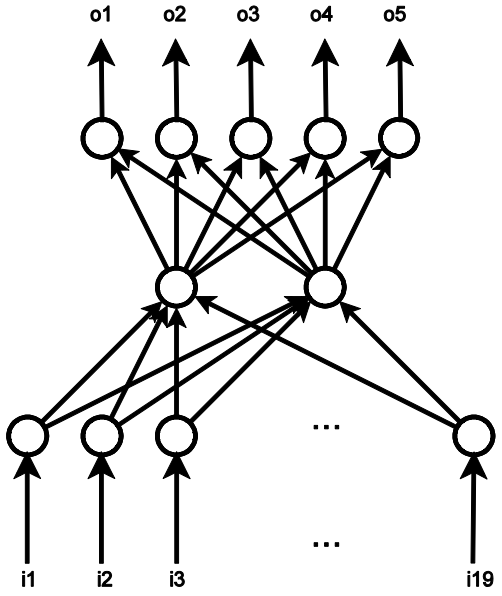


Figure 1 : Structure of the neural network used to predict user ratings on movies.

To ameliorate the prediction, we used a technique called Bagging[1], which is a voting classification algorithm. 70% of the dataset is used for training and the remaining 30% is used for testing. In the Bagging approach 20 networks are trained using random samples of 60% of the training set. The remaining 40\% of the training set was used as validation set in Bagging technique. After the training, the testing set is used to obtain the output of each network. The final output is the voting of the outputs of the 20 networks.

2.3 Movie Recommendations

According to [2], the main goal is not to recommend all the possible movies (since it would lead to the initial of too many choices), but to recommend a number that may be interesting for the user. Thus, the recommendations are limited to 5 movies per user.

The rating system was interpreted as "good" and "bad" recommendations. Movies with ratings 4 or 5 were classified as good. Movies with ratings 1, 2 or 3 were classified as bad. In this way, if the user has given a rating 5 to a movie and the system recommends the same movie with rating 4, even so the movie is classified as good.

To choose the 5 movies, we follow 3 rules that are applied in sequence until the system reaches the number of desired recommendations. The rules are the following:

1. At first, combine the recommendations provided by the two methods (collaborative filtering and neural networks). That is, the movies that are recommended by the two methods at the same time. If the total number of movies that were recommended at this rule is greater than 5, the system randomly selects the 5 movies to avoid other movies always being excluded from the list;
2. If the 5 movies are not reached at rule 1, it is necessary to select movies that were recommended by only one method. If the total of movies is greater than 5, then movies of this rule need to be randomly selected until the total of 5 recommendations is reached;
3. If the 5 movies were not reached yet, it is necessary to randomly select the most popular movies until the total of 5 recommendations is reached.

By following these rules, 5 movies will always be recommended. However, in real cases it is recommended the use of the two first rules. The accuracy of the recommendation depends on the representativeness and completeness of the data samples rated by the user. If the samples that were used in the training set are not sufficient to create a representative profile[1] for a specific user, the accuracy tends to decrease for this user.

3. EXPERIMENTS

The MovieLens dataset [4] was used in the experiments. In the following sections we describe this dataset and how it was used to perform the experiments on the approach described in this work.

3.1 Dataset

In the experiments we used a dataset called MovieLens 100k [4]. The data was collected through the MovieLens web site (*movielens.umn.edu*) by the GroupLens Research Group [5] from University of Minnesota during the period September 19th, 1997 through April 22nd, 1998.

The MovieLens web site has two main goals: and providing personalized recommendations for users. Users provide data to the web site by rating movies they have already watched in a 5 point scale, where 1 is the lowest rating and 5 is the highest.

[1] Users that have a representative profile are commonly the ones that have rated more movies and have strong positive correlation with their *k* nearest neighbors.

The dataset consists of:

- 100.000 rating from 943 users on 1682 movies;
- Each user has rated at least 20 movies (users who had less than 20 ratings were removed from the dataset);
- Each user has simple demographic information such as age, gender, occupation and zip code (users who did not have complete demographic information were removed from the dataset).

3.2 Dataset Interpretation

The ratings system used in MovieLens were analyzed as follows:

- Good. Movies rated with 4 or 5 were classified as being good to recommend to other users;
- Bad. Movies rated with 1, 2 or 3 were classified as being bad to recommend to other users.

Demographic information like users' gender, occupation, and age are available in MovieLens dataset, but they were not used in the experiments. According to [2], demographic information does not contribute enough to justify the increase in system complexity.

With respect to movies content, only information related to movie genres was used in the experiments. Movies from MovieLens dataset were classified according to 19 genres (action, adventure, comedy, etc) and each movie has at least 1 associated genre.

Recommendations based only on genre may frustrate the users. Many times a user may prefer movies from a specific genre, viewing a number of movies related to this genre. However, this does not necessarily mean that all movies from this genre are rated positively by the user. As a result, recommendations based only in genre may not be of total interest for the users.

In the experiments we combine recommendations based in genre with recommendations based in user ratings on movies. Thus, the experiments consist of two main data matrices: genres (1682 movies × 199 genres) and ratings (943 users × 1682 movies).

3.3 Experimental results

The approach was implemented using the MATLAB Neural Network Toolbox. The dataset was divided into sets: training set (70%) and testing set (30%). The training set was used to predict user ratings in both neural networks and collaborative filtering. The final result was achieved by using the testing set for evaluating the proposed approaches.

In the MovieLens dataset, the recommendations were achieved by using only the rules 1 and 2 described in Section 2.3. The main results of the approach can be summarized as follows:

- For some users where it was possible to create a representative profile, the approach described in this work achieved above 80% of good recommendations;
- For other users where it was not possible to create a representative profile, the rate of good recommendations was below to the contrary case achieving an average of 50%;
- The total percentage of success for recommendations based on movie genres using neural networks was 63.8%;
- The total percentage of success for recommendations based on user ratings using only collaborative filtering was 53%.

In the MovieLens dataset the minimum number of movies evaluated by a user is 20. Thus, while some users rates only 20 movies, there are users that evaluated more than 500 movies. This interferes in the training of the network because some users have a significative number of samples to train while others have only a few samples. We believe the approach was successful in the first case. However, in the later case (where users have few training samples) the high sparsity of the MovieLens dataset had a negative influence on the results. Other authors also have reported the sparsity issue as a negative factor when providing recommendations in MovieLens dataset [9].

4. RELATED WORK

There are two main concepts used in recommender systems: (1) collaborative filtering, where items are recommended based on the past preferences of people with a similar profile; and (2) content filtering, where new items are identified by analyzing user past preferences for similar items. A number of techniques are built upon these concepts by employing and combining different machine learning approaches in order to improve recommendations [2,3,11,12].

In [11] the authors describe an approach aimed at reducing the sparsity level of the dataset. The approach described by the authors employs back-propagation neural networks to predict the unknown ratings of items until the sparsity be reduced to a specified level. Thenceforth, the authors employ techniques based on correlation coefficient to predict the ratings of the remaining movies. The method used in [3] is similar to the one used in [11], however it only uses neural networks to predict user ratings. The correlation coefficient is used to provide recommendations.

The method described in [2] is based on content filtering and collaborative filtering. For recommendations based on content filtering, the authors used 3 neural networks for each user. Each one of the 3 network was trained to perform recommendations based on genre, actors and synopsis, respectively. For recommendations based on collaborative filtering, the authors used the PCC to obtain the similarity among users. The recommendations were obtained by combining the 3 neural networks and collaborative filtering.

In [12] the authors propose a recursive algorithm to predict the ratings and generate recommendations. This algorithm is based on the PCC to obtaining the similarity among users. The main purpose of obtaining the similarity is to select a representative neighborhood for each user. The prediction of the ratings is based on the method described in the Equation 3 [8] with addition of recursivity.

All the related works are based on neural networks and collaborative filtering and are aimed at predicting ratings for movies that were not evaluated yet. All the approaches described in this section were performed in the MovieLens dataset, which is the same used in this work.

5. CONCLUSIONS AND FUTURE WORK

Recommender systems help people to find interesting items. Several techniques are employed to learn user preferences and provide recommendations. These techniques can be applied separately or can be combined to achieve better results.

In this work we presented an approach that is based on movie genres and user ratings on movies. The approach combines neural networks and collaborative filtering based on similarity (PCC) to predict user ratings on movies that were not observed yet. In addition we used a technique called Bagging to boost neural network predictions. By combining these techniques we achieve above 80% of good recommendations for users with a representative profile. However, the total average of good recommendations was 50%. The high sparsity in MovieLens dataset may has negatively affected the results.

As future work, we pretend to run the experiments after reduce the sparsity of the dataset. Through this procedure, we expect to improve the recommendations even with users that do not have a representative sample of data.

6. ACKNOWLEDGMENTS

The authors would like to thank CETELI, UFAM, UFMG, FAPEAM, CAPES and CNPq for providing the supportduring the development of this work. The work of Ricardo Erikson is funded by the Programa de Formação de Doutores e Pós-Doutores para o Estado do Amazonas (PRO-DPD/AM-BOLSAS).

7. REFERENCES

[1] Breiman, L. Bagging Predictors. *Mach. Learn.*, 24:123-140, August 1996.

[2] Christakou, C. and Stafylopatis, A., A hybrid movie recommender system based on neural networks. In *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, ISDA'05, pages 500-505, Washington, DC, USA, 2005. IEEE Computer Society.

[3] Gong, S. and Ye, H., An item based collaborative filtering using bp neural networks prediction. In *Proceedings of the 2009 International Conference on Industrial and Information Systems*, IIS'09, pages 146-148, Washington, DC, USA, 2009. IEEE Computer Society.

[4] GroupLens Research. MovieLens Data Sets, Accessed on march 2012. Available at: *http://grouplens.org/node/73*

[5] GroupLens Research, Accessed on march 2012. Available at: *http://grouplens.org/*.

[6] Liu, J.-G., Chen, M. Z. Q., Chen, J., Deng, F., Zhang, H.-T., Zhang, Z.-K. and Zhou, T., Recent advances in personal recommender systems. *Journal of Information and Systems Sciences*, 5(2):230-247, 2009.

[7] Montaner, M., López, B. and Rosa, J. L. D. L., A taxonomy of recommender agents on the Internet. *Artif. Intell. Rev.*, 19:285-330, June 2003.

[8] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J., Grouplens: and open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on computer supported cooperative work*, CSCW'94, pages 175-186, New York, NY, USA, 1994. ACM.

[9] Sarwar, B., Karypis, G., Konstan, J. and Reidl, J., Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, WWW'01, pages 285-295, New York, NY, USA, 2001. ACM.

[10] Shih, U.-Y., and Liu D.-R., Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands. *Expert Systems with Applications*, 35(1-2):350-360, 2008.

[11] Zhang, F. and Chang., H.-Y., A collaborative filtering algorithm embedded bp network to ameliorate sparsity issue. In *Machine Learning and Cybernetics, 2005. Proceedings of the 2005 International Conference on*, volume 3, pages 1839-1844, aug. 2005.

[12] Zhang, J. and Pu, P., A recursive prediction algorithm for collaborative filtering recommender systems. In *Proceedings of the 2007 ACM conference on Recommender Systems*, RecSys'07, pages 57-64, New York, NY, USA, 2007. ACM.

THE USES AND EXPECTATIONS FOR INTERACTIVE TV IN INDIA

Arulchelvan Sriram
Anna University
600025 - Chennai
India
achelvan@annauniv.edu

Pedro Almeida
Aveiro University
3810 – 193 Aveiro
Portugal
almeida@ua.pt

Jorge Ferraz de Abreu
Aveiro University
3810 – 193 Aveiro
Portugal
jfa@ua.pt

ABSTRACT
This research aims to study the uses of TV along with the expectations for interactive TV among the Indian population. A survey was conducted as part of the research in the end of 2011. The results show that TV, computer and mobile phone usage is increasing although many Indian houses have only one TV and one computer. Concerning TV viewing habits, the majority see TV with family members. While watching TV viewers are performing other activities including talking over the phone; working with the computer; reading; or chatting online. Two thirds of the participants regularly recommend TV programs to their friends and others. Concerning the iTV domain, the Indian market is still a greenhorn. Regarding the expectations for upcoming iTV services, the majority of respondents are highly interested in having more interactive services for communication; recommendation and making comments; participating in shows/contests; social network features and a small percentage in shopping through the TV. These results, that expose a clear scope and market potential for those iTV services and programs, are being used in the next phase of this research, which consists of evaluating possible iTV applications for the Indian market.

Categories and Subject Descriptors
Media, Social and Economic Studies

General Terms
Media, Social, Communication, Interactivity

Keywords
iTV, Social iTV, TV Consumption, India

1. INTRODUCTION
TV has been a place of innovation and progress from its inception, while interactive Television (iTV) is a central issue in the TV industry nowadays. It offers the immediacy of interaction with content allowing instant feedback and a wide variety of appealing applications [4]**.** Social Interactive Television (SiTV) is a rather recent development in iTV. It focuses on new opportunities to the viewers to socially engage through the TV screen, sharing presence awareness and viewing information along with freeform communication. 2BeOn from the University of Aveiro, AmigoTV from Alcatel, Media Center Buddies from Microsoft Lab, ConnectTV from TNO are some of the pioneer experiments in this domain [7]. SiTV serves many social purposes, such as providing topics for conversations, easing interaction and promoting feelings of togetherness around TV [10]. The experiments and field trials of the aforementioned Social iTV projects showed that participants were considerably motivated for chatting and performing recommendations while watching TV [1, 8, 3, 2].

This research focuses on identifying possible opportunities for further exploration concerning the development of proposals to support different levels of penetration of iTV in India. India is a very strong technology adopter, and the iTV offer is still very limited. Therefore it can be expected that people are eagerly waiting for the next level.

2. GOALS
The specific goals of this study include: i) to understand the uses and behaviours related with TV consumption; ii) to collect the expectations towards iTV and Social iTV and; iii) to pave the way to the identification of guidelines on how to develop iTV and specially SiTV applications for the Indian market.

3. TV MARKET
To better understand the potential for iTV in India, it is important to provide some data concerning the actual uses of TV. India is the world's third largest TV market with 138 million TV Households, 600 million viewers and more than 550 channels [6]. Cable and Satellite penetration has reached 80%. The Indian market has attracted many foreign TV channels and a high number is also requesting permission to join in. The pay TV offer includes Cable, Direct To Home (DTH) and Internet protocol TV (IPTV). Pay TV households increased in the last years, but IPTV has still a very low penetration rate [13]. Active small operators have mainly run Indian TV distribution industry. Nevertheless, the emergence of large operators is now having a great contribution to changes in the market, mainly in the

major cities where they are concentrated [11]. Cable TV is the most relevant type of offer with 74 million subscribers, though the signal transmission is mainly analogue [12]. Nevertheless, digital cable TV is gaining popularity and its deployment is also expected to pick up soon [5], as digital cable subscribers reached 5 million in 2010.

On the other hand, DTH subscriber base has expanded 30 million in 2010 and it is expected that it will cross the 50 million mark by 2014. Considering IPTV, in a country with many constraints in infrastructures in terms of last mile connectivity the near future growth is not straightforward. The technology is promising due to its superior flexibility and interactive possibilities but the reach is limited [6]**.** The actual numbers point to over 1 million IPTV users, a negligible value for Indian population**.** Considering perspectives of service convergence, triple-play offers are now starting to be available by different operators.

4. MOBILE PHONE & INTERNET

When discussing perspectives for iTV including the communication and interaction possibilities, one needs to consider the use of complementary equipment such as mobile devices and networking. Therefore and to get a better characterization of the technology adoption, we need to consider some data about mobile phone usage. India has the fastest growing mobile telecom network in the world. It is even expected that it will surpass China by 2013 reaching 1.2 billion mobile subscriptions [13]. It is also important to refer that the Internet penetration in India is about 7-8% and out of that 4-5% is through mobile phones [9]. Smartphones accounted for 48% of Internet traffic in India.

TV market potential and scope, audience strength and technology adoption behaviors are highly favorable for new interactive TV services in India.

5. METHODOLOGY

To study the uses and expectations for iTV among people in India the main method to gather data was survey based research. It was conducted through a structured questionnaire released online and promoted by e-mail and through social networks in November and December of 2011 in India. It was not intended to be representative of the population and therefore it was targeted to a limited number of people, mainly at media students, academics and professionals. These groups of people are among the early adopters of the new technologies. The questionnaire gathered a total of 153 responses and it was structured in four sections including personal characterization of the respondents; ownership of media gadgets; TV and other communication media consumption; and usage and expectations for iTV. The personal characterization of the respondents included: their gender; education level; occupation; age; and city. The ownership of media gadgets was targeted at TV; computers and Internet; fixed telephone and mobile phones. The consumption and usage included: the type of TV consumption; sharing behaviours along with other activities while watching TV; and experience with interactive services. Finally, the section about expectations for iTV was targeted at getting the respondents opinion and level of interest regarding several types of interactive applications and features in TV like: communicating with friends; sharing program recommendations and comments; shopping; accessing social networks; reading news and information; and getting additional information about TV shows.

6. RESULTS AND DISCUSSION

6.1. Personal Characterization

As referred, the research gathered a total of 153 respondents with 97 (63.4%) male and 56 (36.6%) female from the cities of Chennai, New Delhi, Hyderabad, Bangalore and Madurai. The study had the participation of various age groups of respondents ranging from 16 to 70 years old. Among the respondents the 21-25 age group was the largest segment with 73 respondents (47.7%) followed by 26-30 age group with 29 (19.0%), 16-20 age group 16 (10.5%), 41-45 age group 10 (6.5%) and other age groups. Considering the level of education of the sample 77 respondents referred Masters (50.3%) followed by Bachelors 33 (21.6%), M.Phil. holders 26 (17.0%) and Ph.D holders 12 (7.8%). Among the respondents students are the highest participants, 41 (26.8%), followed by media professionals (designers, web developers and IT technicians), 34 (22.2%), Professor/Teacher, 22 (14.4%), Management professionals, 25 (16.3%), Journalists, 18 (11.8%), and engineers, 10 (6.5%).

6.2. Ownership of Media Gadgets

Regarding the ownership of gadgets, most of the respondents own mobile phones (91%), Internet access (89%) and fixed telephone (64%). Considering the way services are provided, the sample is dependent from many providers for different media and communication services. The majority of the respondents (57.5%) have cable connection followed by DTH (42.5%).

Although there is an high penetration of TV and computers, these devices are mainly shared in the household, as we can extract from the number of devices in Table 1.

Table 1 - Ownership of TVs and computers

Number of devices owned	% of houses	
	TV Sets	Computers
1	66,0	62,1
2	28,7	24,1
3	3,9	4,6
<3	0,7	2,0
Others	0,7	7,2

Considering the equipment for watching TV, viewers mostly use the TV set (97%), about 15% use their computer and a minimum 3% are using mobile phones.

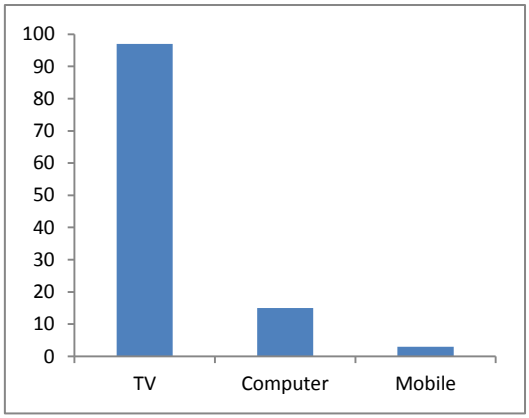


Figure 1- Devices used for watching TV

6.3. Television Viewing Habits

To better understand the TV viewing habits, some dimensions were observed. Considering the time spent watching TV, a major group (42%) watches 1 to 2 hours a day, 26% watch less than one hour, 24% watch 3 to 4 hours and 9% watch more than 5 hours a day. The majority of the respondents (73.2%) are not watching TV alone, something that could be expected by the number of devices available in each house. From those who watch TV with others, a majority is doing it with their family members (62%), followed by friends, colleagues and others. The viewers (79.7%) are doing many other activities while watching TV. It includes talking over the phone (44%), working at the computer (38%), reading (27%), chatting online (23%) and doing some other household activities (47%). In what concerns with social behaviors, about 2/3 of respondents usually recommend programs/shows to others. These recommendations are mostly performed during the program or in the following days (Figure 2).

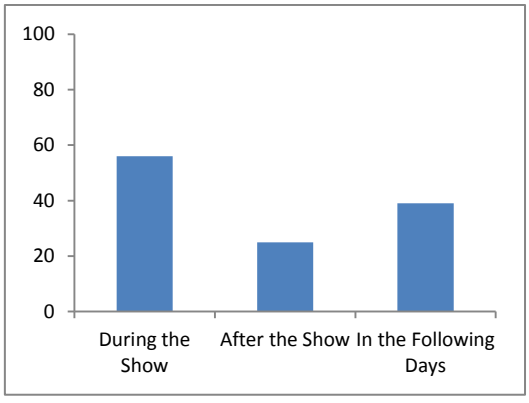


Figure 2- The moment when users recommend TV programs/shows

To send the recommendations different methods and gadgets are used. Respondents are mostly using SMS (61%), followed by phone (48%), in person (44%), through social networks (26%), and finally email (6%). This survey also reveals that 24% of respondents are watching shopping programs through TV. However, only a small percentage of them are buying products/ services.

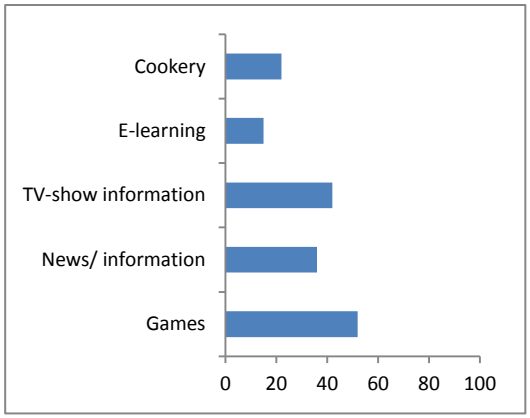


Figure 3- Types of (offline) Interactive TV applications received.

Considering interactive (TV) features, many channels and operators are now providing some type of interactive programs. This study shows that about 26% of respondents are receiving interactive programs and services. But this interactive applications are client installed, working offline, therefore limited in the type of interaction. Among them interactive games are in the top of preferences with (52%) followed by TV shows related information (42%), news and information (36%), cookery (22%) and e-learning related applications (15%).

6.4. Expectations for iTV

This survey reveals that about 3/4 of respondents (74%) are interested in having more interactive services in the TV. iTV may support many interactive features like communication, sharing content and recommendations, ordering products/services, interact with social networks or getting additional information about TV programs. The final section of the survey gathered the users' expectations and motivations towards these features.

Considering features related with the supporting communication situations with family members, friends and relatives through the TV, about 50% of respondents pointed that were most interested in this feature.

Recommending programs/channels with others and making comments are important activities in iTV. 26% of the respondents stated to be most interested in these features. Participatory TV shows and contests are having more audience, but still they are not attaining complete interactivity. iTV may perform an additional support for these shows. About 40% referred to be most interested in interacting with the shows and contests. About ordering some products and services, 10% of respondents showed to be most interested in this interactivity. Considering the interaction with social networks, some more significant data was gathered. About 38% of respondents showed to be most interested in having access to social networks in the TV. Finally, the users were asked about their interest in having additional information about the shows - enhanced information. 17.6% of respondents were most interested in this issue.

Providing all these features may carry some costs to the users. Therefore, they were asked to say if they would pay to get these features. 36% expressed their willingness to pay for it.

7. SUMMARY & CONCLUSIONS

This study found that most of the respondents are active TV viewers. They spend a lot of time watching TV and are eager for innovative services and interactive features. Despite the size of the sample, these results provide some indications of the high potential of iTV services in India. The interactive applications provided nowadays by the operators are mainly offline applications, but they provide an important experience that could be boosted with true iTV applications.

As referred these findings need further validation and complementary data. For this, a second phase of this research project will include providing Indian users and operators with hands-on experiences with (social) iTV prototypes. It is expected that more qualitative data can enrich the study and provide more insights towards the development of the (social) iTV domain in India.

8. ACKNOWLEDGEMENTS

This paper is based on research for the project *'Interactive Television for Social and Interpersonal Communication – A study on India and Portugal'* and supported by FCT - Fundação para a Ciência e a Tecnologia.

9. REFERENCES

1. Abreu, J. F., & Almeida, P. (2010). *From Scratch to User Evaluation - Validating a Social ITV Platform.* In. Adjunct Proceedings of EuroITV 2008, Salzburg, Austria 2008.
2. Abreu, J. F., Almeida, P., & Ricardo Pinto, V. N. (2009). *Implementation of Social Features Over Regular IPTV STB*. In Proceedings of the Seventh European Conference on European interactive Television Conference EuroITV '09. ACM, New York, NY, 29-32.
3. Boertjes, E., Klok, J., & Schultz, S. (2008). *ConnecTV: Results of the Field Trial*. In. Adjunct Proceedings of EuroITV 2008, Salzburg, Austria 2008.
4. Burke, D. (2011). *A Guide to Interactive TV*. Retrieved from http://www.whitedot.org/issue/iss_story.asp?slug=shortspytv – 4-03-2012.
5. Changrani, J. (2011). *The Digital Cable TV Industry: Paving the Road to Success.* Retrieved from http://voicendata.ciol.com/content/ContributoryArticles/111030901.asp - 4-3-2012
6. FICCI, & KPMG. (2011). *Indian Media and Entertainment Industry, Report 2011.*
7. Harboe, G. (2009). *In Search of Social TV, Social Interactive Television: An Immersed Shared Experiences and Perspectives.* IGI Global Publications.
8. Harboe, G., Massey, N., Metcalf, C., Wheatley, D., & Romano, G. (2008). *The uses of social television*. Computers in Entertainment, 6(1), 1. ACM.
9. Kartik. (2011). *Mobile Internet Usage in India: Statistics, Facts & Opportunities.* Retrieved from http://www.goospoos.com/2011/05/mobile-internet-usage-in-india/ - 4-3-2012
10. Lull, J. (1990). *The Social Uses of TV, In Inside Family Viewing.* New York: Routledge.
11. Moni, V. (2010). *Cable TV Industry in India.* Retrieved from http://technology.ezinemark.com/cable-tv-industry-in-india-171b12f9d8c.html - 4-3-2012
12. PWC. (2011). *India Entertainment and Media Outlook.* Price Water Cooper.
13. TSI. (2011). *Telecommunications Statistics in India.* Retrieved from Wikipedia – 4-3-2012.

Reconfigurable Systems for Digital TV Environment

Rodrigo Ribeiro de Oliveira
Universidade Federal do Amazonas
Av. Rodrigo Octávio J. Ramos, 3000
69077000, Coroado I. Manaus, AM
+55(092)81143441
rodrigo@dcc.ufam.edu.br

Eddie Batista de Lima Filho
Centro de Ciênc., Tec. e Inov. do PIM
Rua Salvador, 391 69057040,
Adrianópolis, Manaus, AM
+55(092)21235833
eddie@ctpim.org.br

Vicente Ferreira de Lucena Jr.
Universidade Federal do Amazonas
Av. Rodrigo Octávio J. Ramos, 3000
69077000, Coroado I. Manaus, AM
+55(092)81586371
vicente@ufam.edu.br

ABSTRACT
This paper details the current status of the development of a complete methodology for reconfiguring a receiver through the Digital TV broadcast signal. This project aims the design of platforms with reconfigurable features, whose goal is to minimize the legacy (old devices) that arises when deploying a new transmitting system. The reconfiguration of the FPGA is performed through a bit stream containing the synthesized hardware, which is transmitted as generic data embedded in the digital TV multiplex. The methodology used in the reconfiguration process and the partial results obtained during the update process are presented in this paper. As a result, future revisions in digital television standards could take place without immediately replacing the reception devices.

Keywords
SBTVD, HDTV, DTV, FPGA, VHDL, VERILOG, NRE e DATACASTING.

1. INTRODUCTION
Currently, the Field Programmable Gate Array (FPGA) plays a more dominant role in integrated systems design, mainly due to higher demands in logic capacity and on-chip resources. The new generation of FPGAs overcame limitations presented by the Moore's law, and now these devices are capable of meeting most design requirements [12].

The use of FPGAs was boosted by the semiconductor industry, due to the development of several Hardware Description Language (HDL) examples codes, which are normally provided in modules and can be directly incorporated into other projects. Besides, there also are bit streams synthesized for different architectures and HDLs for circuit simulation, test and validation [3] [6].

Another important issue is that the current FPGA cost is similar to the ASIC one, which is essentially given by the amortization of the cost of Non-Recurring Engineering (NRE) among customers of each integrated circuit (IC) family [3]. It then reinforces the use of FPGAs in embedded systems design, in such a way that the lifecycle of the final product becomes easier to control.

Most digital TV Set-Top Boxes (STB) available in market use a common hybrid system strategy, with modules developed in hardware, like video decoders, silicon tuner and demodulators, and also in software, running on higher-capacity processors [2]. Thus, the preparation of the system for being capable of upgrade should be performed during the design phase, according to the intended system framework.

In order to synthesize a code, written in a hardware description language such as Very High Speed Integrated Circuit HDL (VHDL) or Verilog, it is necessary to use a Computer Aided Design (CAD) tool. The result of this process is a bit stream containing the hardware description.

In the context of a digital television (DTV) system, the bit stream can be regarded as data that is transmitted along with the HDTV content of the television program (Figure 1). A DTV system is composed of several blocks, which include data preparation for transmission and complete signal reception, with subsequent content filtering.

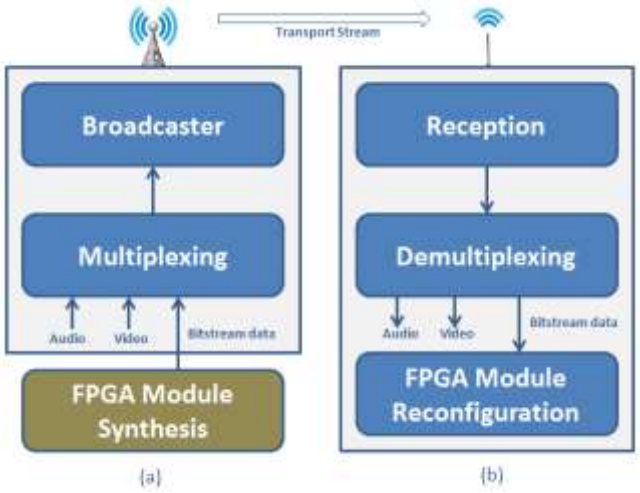


Figure 1 - Transmission of the hardware reconfiguration stream.

The transmitted information is usually composed of audio, video and data packets; the latter, in the present work, carry the hardware reconfiguration data. The generation of the reconfigurable stream is performed during the FPGA Module Synthesis Step (Figure 1(a)). In this step, the FPGA module is synthesized and, after passing test procedures, the resulting bit stream is ready to be sent.

The hardware reconfiguration stream, which is part of the generic data, may be encapsulated by a mechanism of datacasting (see Section 2). In the TS multiplexing step, audio, video and data (which include the hardware description) are interleaved, which gives rise to a Transport Stream (TS), which is necessary for transmitting information in a DTV system [1]. Finally, in the transmission step (Broadcaster), the resulting signal is then modulated and sent through the DTV channel.

On the other side, Figure 1(b), the receiver detects and decodes the signal. The result of this initial process then goes through the

demultiplexing step, which extracts audio, video and data packets, while part of the latter is used for reconfiguring the FPGA. The reconfiguration is performed through user interaction, with the aid of a resident control system, present in the STB, which notifies users when new updates are available. When an update is triggered, the system retrieves the reconfiguration stream and reassemblies it in a persistent storage module. Following that, the FPGA is automatically restarted, now with a new operation description.

This article is divided into five sections. The next one, which is section 2, contains a brief explanation about data broadcast and the mechanisms for signal transmission and reception. Section 3 discusses the reconfigurable modules and the technologies related to this process. All concepts presented earlier are then used in the formulation of the case study, in section 4. Section 5 contains the concluding remarks.

2. DATA BROADCAST

Data broadcast (datacasting) is one of the most interesting features of a digital TV system and offers the opportunity to make a more intelligent use of the channel bandwidth, allowing the delivery of generic data to receivers located in the coverage area. As the MPEG-2 is the standard adopted for multiplexing data in open DTV systems, the existing mechanisms make use of its facilities and structures for providing datacasting services. There are four known transport mechanisms: data piping [9], data streaming [9][10], multiprotocol encapsulation (MPE) [8][10] and Carousels, which are divided in data carousels and object carousels [9][10].

3. RECONFIGURABLE MODULES

The electronics industry is adopting hardware reconfiguration in several products [9]. In the digital TV segment, FPGAs has been used for implementing some functions of the receiver, like time control and output interfaces [7]. Recently, research projects seen in [4] and [5] have shown the use FPGAs for decoding video H.264 [11]. In [5], research efforts regarding H.264 video decoding aim to join efforts for developing intellectual property for the Brazilian Digital Television System (SBTVD[1]). Also, the target device (FPGA) used for the synthesis process was a XILINX Virtex-II Pro XC2VP30; resources required for this process include 21,200 LUTs, 8465 slices, 5671 registers, 21 internal memory modules and 12 multipliers. In [4], the target device used in the validation process and also for testing was a XILINX Virtex-II 2VP30FF896 – Pro. The results were satisfactory and indicate that these devices can be used in DTV receivers.

4. CASE STUDY AND RESULTS

In this case study, it was considered the update of the H.264 video decoder module, given that it is possible to transmit any synthesized FPGA module through datacasting. The base platform (reference design – Figure 2) used in this project was the NXP - STB236, which is based on the PNX8735 processor. The PNX has a Digital Signal Processor (DSP) responsible for decoding video, whose output is then routed to the video encoder output (Digital Encoder - DENC, which provides signals suitable to the standards adopted in a given country), before reaching the analog-to-digital conversion, and also to the digital outputs subsystem (e.g. HDMI).

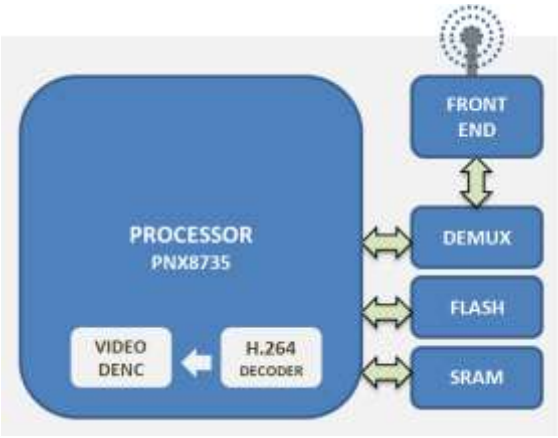


Figure 2 – Reference design platform STB236

The front-end, which is composed by a PLL, a demodulator and peripheral devices, performs the reception of the transmitted signal (air interface) and outputs a TS. Following that, when the TS reaches the back-end (processors and decoders), it is demultiplexed, separating audio, video and data. The demultiplexer filters the TS, through filters based on packet identifiers (PID[2]), and sends the resulting data to suitable processing modules.

In order to create a new architecture, the target device (receiver) must integrate a FPGA circuit and a FLASH memory (deployed in the grid area), which is illustrated in Figure 3. This architecture was prepared for reconfiguring the H.264 video decoder using a FPGA.

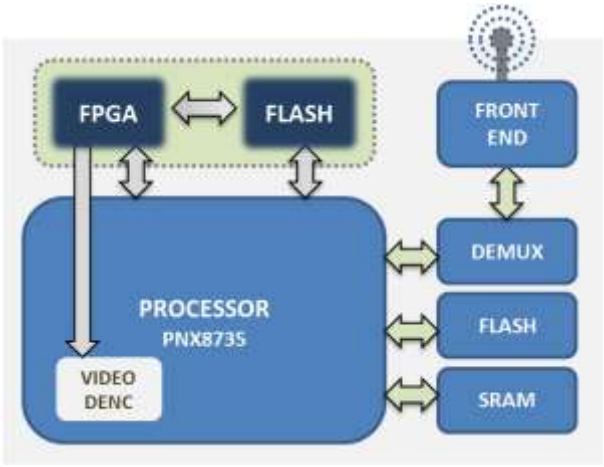


Figure 3 - Architecture with FPGA integrated

In order to perform this task, it is necessary to synthesize the H.264 video decoder [11] to a specific FPGA model and generate the bit stream. After that, the reconfiguration stream must be multiplexed with the digital TV content (audio and video). The transport stream multiplexing is performed using any multiplexing tool. For enabling the hardware reconfiguration (the upgrade of the H.264 decoder [11]) feature, which is the focus of this study, the transmitted data are regarded as bounded and can be divided into a finite number of slices of predetermined size, before being

[1] SBTVD – Brazilian System of Digital Television.

[2] PID – Packet Identifier.

fed to the transport mechanism. Besides, from the perspective of timing requirements, the data are asynchronous, because the bit stream does not need any synchronization with other data types or media elements.

After evaluating many possible mechanisms for transporting data, the streaming through private sections was chosen for the present work. The mechanism for data streaming via private sections is simpler and provides tools for error detection, ensuring system reliability. In addition, it also offers support to asynchronous and bounded data and uses less complex structures when compared to carousels. Among the selection criteria that were taken into consideration, it is worth noticing that some of the most important ones were the necessary computational resources, implementation complexity and packaging strategy (multiplexing) of the TS. By adopting the data streaming mechanism through private sections, the reconfiguration bit stream can be broken up into sections of 4084 bytes of payload, plus 8 bytes of header and 4 bytes of CRC, which are then tagged according to the slice insertion order (each 4084 bytes of data) of the bit stream. This mechanism also allows the adjustment of the section repetition rate. Thus, the reconfiguring application running on the receiver is able recover the transmitted information, in the original order established by the encapsulation process. After the transport stream generation procedure, it is necessary to employ a broadcast system for transmitting the related content, using a specific channel.

The resident software is responsible for controlling sections/tables filtering and finding hardware updates, which are accomplished in a systematic way. So, a new table was created to make the receiver aware of a reconfiguration stream, which is called update information table (UIT) (Table 1), whose goal is to broadcast the availability of an update regarding video decoding and the PID of packets carrying the reconfiguration bit stream.

Table 1 – UIT -- Update Information Table.

Data Structure	Bit Rate	Bit string
update_information_table(){		
table_id	8	Uimsbf
section_syntax_indicator	1	Bslbf
reserved_future_use	1	Bslbf
reserved	2	Bslbf
section_length	12	Uimsbf
synthesizable_core_type	8	Uimsbf
section_number	8	Uimsbf
last_section_number	8	Uimsbf
reserved_future_use	3	Uimsbf
reconfiguration_data_PID	**13**	Uimsbf
Reserved	4	Bslbf
update_loop_length	12	Uimsbf
for (i=0; i<N;i++) {		
update_identifier ()		
update_control_code	8	uimsbf
reserved_future_use	4	Bslbf
update_descriptor_loop_length	12	uimsbf
for (j=0; j<M; ;j++) {		
descriptor ()		
}		
}		
CRC_32	32	Rpchof
}		

The access point to newly created UIT tables is given by a PID located on the Program Map Table (PMT). Table 2 shows the list of module types allowed by the synthesizable core type field of a UIT.

Table 2 – Description the field synthesizable_core_type.

Synthesizable_core_type	Description
0x0000	reserved_future_use
0x0001	H.264 Syntetizable Decoder
0x0002...0x7FFF	Undefined

Table 3 shows the core description sent by the broadcaster. The field manufacture_id informs the manufacture of the FPGA (e.g. Xilinx-Virtex II, etc) intended to be updated. The HDL_used field holds the name of the HDL used during the module synthesis procedure (e.g. VHDL, Verilog, etc), and the target_device field presents the device type for which the core was synthesized (e.g. 2VP30FF896-Xilinx).

Table 3 – Description of the update_identifier field.

Data structure	Bit rate	Bit string
update_identifier (){		
manufacture _id	16	uimsbf
HDL_used	8	uimsbf
target_device	16	uimsbf
}		

After filtering this table, the PID of the reconfiguration stream is retrieved (see data flow, Figure 4). This then allows the download of the bit stream, through a PID filter allocated in the demultiplexer. The download is done directly in FLASH memory, which is then followed by a data reassembly procedure, ensuring the correct organization of the reconfiguration file. Finally, with the binary file (bit stream) correctly recorded in memory, the processor then restarts the FPGA-based system, which makes the new decoder operational. After completing the reconfiguration process, video decoding is restarted and then video packets are sent to the FPGA through a standard interface (e.g. DVB-CI – Common Interface or USB 2.0).

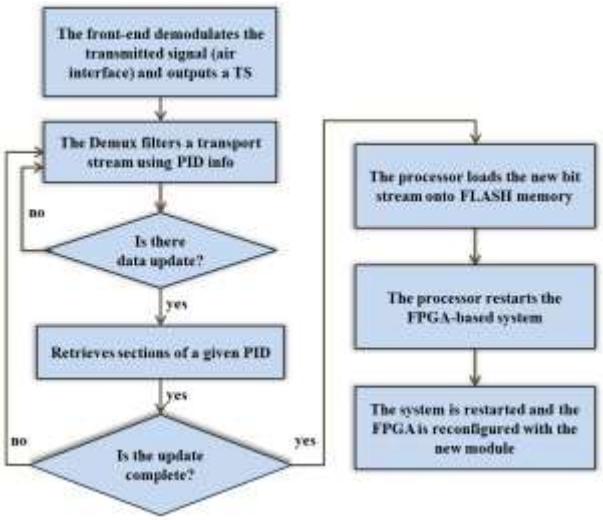


Figure 4 - Updating Flow

In order to implement the reconfiguration steps shown in Figure 3, a resident system (written in Qt/C++ and cross compiled to the PNX8735 main processor) is being developed, with the Graphical User Interface (GUI) depicted Figure 5. This GUI is divided into 6 steps: Lock, Checksum, Extracting, Flashing, Restarting and Playing. The first one (Lock) indicates when a channel is tuned. The current channel is then shown on the upper right corner of the screen and can be zapped using a remote control. In Figure 5, when channel 13 was tuned the LOCK gray button turned blue. After the Lock step, the demultiplexer is programmed and then the SI tables filtering procedure begins. If reconfiguration content is

found, the Checksum step starts and validates each table section, by evaluating the payload content with a Cyclic Redundancy Code (CRC) and comparing the result to the last 4 bytes of each section (CRC hash). The Checksum step ends when the last section arrives. During this procedure, the Checksum Progress bar shows the percentage related to the checked sections. If all sections are correct, the update process continues. In the Extracting step, all sections are reassembled to form a unique file. During this procedure, the Download Progress bar shows the percentage of sections that were already relocated. Finally, in the Flashing step, the flash memory is erased and then the new bit stream is stored. The next step (Restarting) reboots the system, that is, the FPGA retrieves the content stored in FLASH memory and restarts with a new module configuration. The last step (Playing) initializes the rest of the system and enables signal decoding, with video packages rerouted to the FPGA, which now provides the new video decoder.

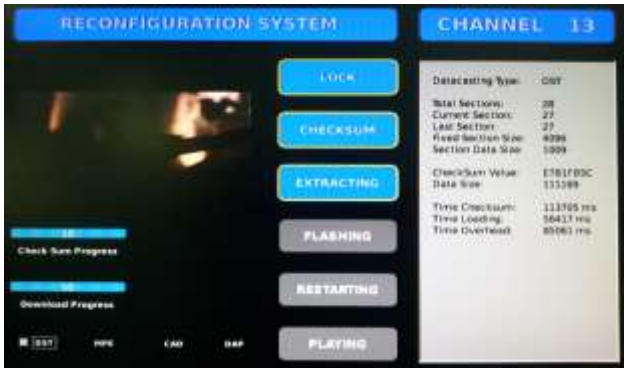


Figure 5 – Graphical User Interface of the resident system.

In addition, the GUI shows 4 check boxes, which identify the method used for multiplexing the reconfiguration data. PST represents the data streaming using private section, MPE the Multiprotocol Encapsulation, CAD the data carousel, and DAP the data piping. The other elements of the GUI show features of the downloaded content, like total, current and last section size, checksum value, total data size, and checksum, loading and overhead time.

In order to perform tests regarding the reconfiguration system, a test TS was generated. The data (a stream of 111KB) was encapsulated by the data streaming using private sections method. The bit rate used for transmitting the reconfiguration stream was about 1.57 Mbps. In order to transmit the test TS, a DTV test transmitter (SFU - Rohde Schwarz) was employed.

The result presented on Table 4 shows the average values to check and downloads the transmitted content. The checksum time, is the time required to check if the section data content is correct, the data loading time is the time necessary to extract the data content of the section payload. The overhead time is the total time to download the data content.

Table 4 – Results of transmission tests using the TS ex1.ts.

Feature	Results
Checksum time	7677 ms
Data loading time (extracting)	11317 ms
Overhead time	9236 ms

5. CONCLUDING REMARKS

This work presented a hardware reconfiguration procedure intended for H.264 decoders, however, the same concept could be used for updating others modules, like audio decoders or even channel decoders.

The same architecture could also provide update support for several different modules, which would create a completely reconfigurable system.

The expected results for this project are directly related to the context of its application. In a DTV system, this work could help to reduce the legacy caused by establishing a new DTV standard or improving an existing one. Thus, the overall cost involved in this process would also be considerably reduced. However, this would still depend on a close contact between broadcasters and receiver manufacturers, in order to update all devices in the DTV network.

6. REFERENCES

[1] ABNT NBR 15602-3, Brazilian Standard. Digital Terrestrial Television – Video coding, audio coding and multiplexing Part 3: Signal multiplexing systems. 2007.

[2] ABNT NBR 15604, Brazilian Standard. Digital Terrestrial Television - Receivers, 2007.

[3] Actel Corporation, "Flash FPGAs in the value-based market white paper," Tech. Rep. 55900021-0, Actel, Mountain View, Calif, USA, 2005. http://www.actel.com.

[4] Agostini, L. V., Porto, M., Güntzel, J. L., Porto, R. C., Bampi, S. High Throughput FPGA Based Architecture for H.264/AVC Inverse Transforms and Quantization. In: MWSCAS 2006 - 49th IEEE International Midwest Symposium on Circuits and Systems, San Jose, 2006.

[5] Agostini, L. V., Azevedo, A., Staehler, W., Rosa, V., Zatt, B., Pinto, A. C., Porto, R. C., Bampi, S., Suzin, A., Design and FPGA Prototyping of a H.264/AVC Main Profile Decoder for HDTV. Journal of the Brazilian Computer Society, v. 13, p. 25-36, 2007.

[6] Altera, "Standard Cell ASIC to FPGA Design Methodology and Guidelines." http://www.altera.com.

[7] Altera, "Supporting Digital Television Trends with Next-Generation FPGAs." http://www.altera.com.

[8] ETSI, "Digital Video Broadcasting (DVB): Specification for Data Broadcasting," ETSI standard, EN 301 192, V1.5.1, 2009.

[9] Garcia, P., Compton, K., Schulte, M., Blem, E., Fu, W., An Overview of reconfigurable hardware in embedded systems, EURASIP Journal on Embedded Systems, v.2006 n.1, p.13-13, 2006.

[10] Reimers U., DVB - The Family of International Standards for Digital Video Broadcasting. New York: Springer, 2004.

[11] Richardson, I. E., H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia. Wiley, 2003.

[12] Xilinx, "Xilinx Stacked Silicon Interconnect Technology Delivers Breakthrough FPGA Capacity, Bandwidth, and Power Efficiency." 2010. http://www.xilinx.com.

Workshop 3: Third Workshop on Quality of Experience for Multimedia Content Sharing

Quantitative Performance Evaluation Of the Emerging HEVC/H.265 Video Codec

Harilaos Koumaras
Computer Science Department
Business College of Athens (BCA)
Athens, Greece
+30 210 725 3783
koumaras@bca.edu.gr

Michail-Alexandros Kourtis
Department of Informatics, Athens University of Economics and Business
Athens, Greece
+30 210 650 3166
P3070085@dias.aueb.gr

Spyros Mantzouratos
Institute of Informatics and Telecommunications
NCSR Demokritos
Agia Paraskevi,Greece
+30 210 650 3107
smantzouratos@iit.demokritos.gr

Drakoulis Martakos
Department of Informatics and Telecommunications,
University of Athens
Zografou, Greece
+30 210 727 5217
martakos@di.uoa.gr

ABSTRACT

This paper, deals with the currently under-development video encoding standard H.265/HEVC. The main purpose of this paper is to perform a quantitative performance evaluation of the emerging standard in comparison to its predecessor H.264/AVC. The paper focuses on both the encoding and compression efficiency by measuring the PSNR scores and bit rate values of reference and non-reference signals. For the needs of the paper, the HEVC and AVC reference encoders were used. Scope of the paper is to research if the current working version of the new coding method is approaching or achieves its primary objective, which is to double the compression efficiency of the bit stream without significant degradation of the encoded quality.

Categories and Subject Descriptors

H.5.1 **[Multimedia Information Systems]**: Evaluation/methodology

General Terms

Measurement, Documentation, Performance, Design.

Keywords

HEVC, AVC, H.265, H.264, PSNR.

1. INTRODUCTION

The proliferation of the video services and applications has brought the contemporary world closer to digital videos than ever before, creating new demands for high performance video compression standards [1]–[2] in conjunction with efficient and resilient transmission systems [3]. This has increased the demand for high efficiency video coding methods and for this reason various standardized efforts run in order to derive new and more efficient encoding standards in terms of both performance (i.e. data compression) and video quality (i.e., Quality of Experience).

To meet the ongoing industry requirement for more efficient and standardized video coding techniques, ITU-T/VCEG and ISO-IEC/MPEG have recently formed the joint collaborative team on video coding (JCT-VC). The JCT-VC aims to develop the next generation video coding standard called High Efficiency Video Coding (HEVC) or H.265 as it is widely called, which is being developed as the successor to H.264/AVC.

The HEVC's main objective is to provide significant improvement in the video compression and performance efficiency compared to H.264/AVC, reducing bitrate requirements by half with comparable video quality, probably at the expense of increased computational complexity, which is expected to be three times higher. Thus, the new encoder is expected to satisfy the ever-increasing requirements for cost effective video encoding process in terms of better video quality compression efficiency, video resolution, frame rates and computational complexity.

In this framework, the authors of the paper, by tracking the development of the reference HEVC test model (HM), report on its encoding performance and compression efficiency by measuring the current performance and efficiency of HEVC in respect with its main objectives. More specifically, a set of YUV video signals (reference and non-reference ones) was used as input to both the reference encoders (i.e., HEVC and AVC). The set of test signals covers a variety of spatiotemporal activity, making the benchmarking framework of this paper appropriate for testing the performance efficiency of the new encoder under different signal complexity. For the performance assessment of the two encoders, the PSNR metric was selected.

Upon this introductory section, the rest of the paper is organized as follows: Section 2 introduces the main novel features of H.265/HEVC, providing a brief description. Section 3 describes the performance metric used in this paper. Section 4 describes the encoding process of H.264/AVC and H.265/HEVC and the test signals (reference and non-reference ones) used in the experimental part of this paper. In section 5 the benchmarking between the two codecs under test is performed in terms of compression and performance efficiency. Finally, Section 6 concludes the paper.

2. INTRODUCING HEVC/H.265

2.1 Block-Based Coding

The HEVC continues to implement the block-based hybrid video coding framework [4], with the exception of the increased macro-block size (up to 64x64) compared to AVC. However, three novel block concepts are introduced, namely: the Coding Unit (CU), the Prediction Unit (PU) and the Transform Unit (TU). The general outline of the coding structure is formed by various sizes of CUs, PUs and TUs in a recursive manner, once the size of the Largest Coding Unit (LCU) and the hierarchical depth of CU are defined. Given the size and the hierarchical depth of LCU, CU can be expressed as a recursive 64x64 quad-tree representation as it is depicted in Figure 1, where the leaf nodes of CUs can be further split into PUs or TUs, thus using variable block size (down to 4x4) for various regions of the frame. Since the leaf blocks of the CUs are either PUs or TUs, this means that they are used for prediction methods.

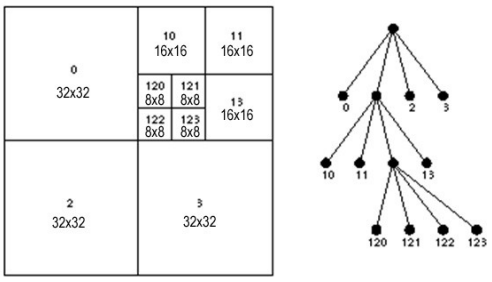


Figure 1. Example of 64x64 quadtree coding

When a prediction method is chosen, then a PU is created and matched to the CU, which contains the information about the type of the prediction method (e.g. intra or inter). The PU in turn may be further split the block to block-leaves and each leaf may apply a different prediction method. On the contrary, the TU is a block area for transform and quantization of data, without being necessary to be limited to coding tree leaf blocks as PU.

The introduction of this flexible sub-partitioning mechanism is one of the most important elements for higher compression performance in the encoded video signals, due to the adaptive accuracy of the encoding process on specific parts of the signal content. Thus, the more complex the video content is at specific areas, the better encoding performance is succeeded, due to the recursively break down coding of the parent CU block to four children blocks of smaller sizes (i.e., each one of them being the ¼ of the original CU block)..Therefore, this quad-tree approach provides smaller blocks which in turn build higher overhead, but it provides more efficient predictions.

2.2 Intra-Prediction in HEVC/H.265

The current intra prediction technique in HEVC unifies two simplified directional intra-prediction methods: the Arbitrary Direction Intra and the Angular Intra Prediction. The unified intra prediction technique enables a lower-complexity method in which parallel processing can be achieved. More specifically, samples of already decoded adjacent PUs are used in order to define the type of the intra prediction method (i.e., horizontal, vertical or depending on the block size up to 34 angular directions).

2.3 Inter-Prediction in HEVC/H.265

The inter prediction in HEVC uses the frames stored in a reference frame buffer, which allows multiple bi-direction frame reference. A reference picture index and a motion vector displacement are needed in order to select reference area. The merging of adjacent PUs is possible, by the motion vector, not necessarily of rectangular shape as their parent CUs. In order to achieve encoding efficiency, skip and direct modes similar to the AVC ones are defined, and motion vector derivation or a new scheme named motion vector competition is performed on adjacent PUs. Motion compensation is performed with a quarter-sample motion vector precision. At TU level (which commonly is not larger than the PU), an integer spatial transform (with range from 4x4 to 64x64) is used, similar in concept to the DCT transform. In addition a rotational transform can be used for block sizes larger than 8x8, and apply only to lower frequency components. In AVC scaling, quantization and scanning of transform are performed in a similar way.

At CU level, an Adaptive Loop Filter (ALF) can be applied prior to copying the frame into the reference picture buffer in order to minimize distortion relative to the original picture. Additionally a de-blocking filter is operated within the prediction loop (similar to the AVC de-blocking filter design). After applying these two filters the display output is written to the picture buffer.

2.4 Entropy Coding in HEVC/H.265

The HEVC defines two context-adaptive entropy coding patterns, one for the higher-complexity mode and one for the lower-complexity mode. The lower-complexity mode is based on a variable length code (VLC) table selection for all the syntax elements, while using a particular code table, which is picked in a context-based scheme depending on previous decoded values. This design is very similar to the CALVC pattern from AVC, but enables even simpler implementation according to its more systematic structure. A re-sorting of code table elements can be used as a supplementary compression improvement.

The higher-complexity design uses a binarization and context adaptation pattern similar to the AVC entropy coder, CABAC, but with the difference of using a set of variable-length-to-variable-length codes (indexing a variable number of bins into a variable number of encoded bits) instead of using an arithmetic coding engine. This is performed by applying a bank of parallel VLC coders – each of which is responsible for a certain range of odds of binary events (which area referred to as bins). The coding performance can be better parallelized and has higher throughput per processing cycle than CABAC, although being very similar to it. It must be noted that the compression performance of this design can be significantly higher than the lower-complexity VLC.

2.5 Encoding Profiles

The HEVC reference software has three main profile categories, namely: i) Intra Profile, which uses only I frames to code, aiming to studio usage, ii) Random Access Profile, which uses I frames with hierarchical bidirectional B frames, providing efficient compression but high computational power, and iii) the Low Delay profile, which uses only one I frame at the beginning and the rest frames are P or unidirectional B (i.e., restricted to previous frames), aiming to real time applications that require low delay in the coding process (like video telephony).

Each of the aforementioned profiles may be applied to a Low Complexity variant, where some of the encoding tools are disabled or switched to simpler modes that require less computational power, thus there are faster. Thus, six total profiles are available, with the low delay profile being further split to B or P frame mode.

3. PERFORMANCE METRIC

For the purposes of this paper the PSNR performance metric was used, providing a more profound and clear statement regarding the encoding efficiency of the HEVC encoder compared to AVC one.

$$PSNR = 10 \times \log_{10} \frac{MaxErr^2 \times w \times h}{\sum_{i=0,j=0}^{w,h} (x_{i,j} - y_{i,j})^2}$$

The PSNR metric is mostly used as a measure to calculate the error and noise introduced during the encoding process, to an encoded video signal compared to the original one. It is also used as an approximation to human perception of reconstruction quality, thus becoming a very useful tool for the purposes of these tests. It is defined via the mean squared error for two w (video width) x h (video height) monochrome images $x_{i,j}$ and $y_{i,j}$ where one of the images is considered a noisy approximation of the other, with *MaxErr* being the maximum possible absolute value of color components difference.

4. SIGNALS AND ENCODING PROCESS

For the evaluation process two reference video clips (Bubbles, Horse Race) [5] and two non-reference unique videos (Apocalypto Trailer and Batman Dark Night Trailer) were used, which represent various levels of spatial and temporal activity. A representative snapshot of each signal is depicted on Figure 2.

The test signals have spatial resolution 416x240 and 352x288 (reference and non-reference respectively), and for the experimental needs of this paper were encoded from their original uncompressed YUV format to ISO AVC Main Profile (MP) and to the following profiles of HEVC, namely: i. Random Access Profile (RAP), ii. Random Access Low Complexity Profile (RALCP), iii. Low Delay Profile (LDP), and iv. Low Delay P Profile (LDPP). Across all the encoding process, the reference software was used for both the AVC and HEVC coding. Especially for HEVC the HEVC Test Model (HM) Reference Software 5.1 was used [6].



Figure 2.The test signals

In order to maintain and achieve an ideal comparison between the various profiles, it is necessary all the profile configurations to have identical or very similar parameter values. For this reason, the GOP structure for all the encoding profiles and between the two encoding methods consisted of either I, P or B except in cases of LDP and LDPP where I and B only, and I and P only, patterns were used respectively, ensuring by this method the benchmarking of both Intra- and Inter-coding efficiency between AVC and HEVC standard profiles.

The Quantization Parameter (QP), for I and P frames, has a great impact on visual quality and compression rate, as it regulates how much spatial detail is maintained. For this reason, a group of QP values {12, 22, 32, 42, 51} was used, enabling the creation of an extensive evaluation results database and creating a more complete and rich set of measurements.

5. BENCHMARKING OF HEVC AND AVC

5.1 Performance Comparison

In this section, the video quality of the HEVC algorithm is examined, in comparison to the AVC for the video signals under test, when the same encoding parameters have been selected (i.e. the QP for both I and P or B frames were set to the value of {12, 22, 32, 42, 51} in both AVC and HEVC profiles).

The first comparison between HEVC and AVC is related to the average PSNR value vs. QP. Table 1 shows the average PSNR values and the corresponding QPs from all the test video signals, for the Main Profile (MP) of AVC and the four HEVC profiles (LDP, LDPP, RAP, RALCP).

Table 1. Average PSNR of H.264/AVC(MP) vs. HEVC(LDP,LDPP,RAP,RALCP)

QP	MP	LDP	LDPP	RAP	RALCP
12	52.08022	50.07781	50.14053	48.60677	48.04049
22	43.82217	42.21041	42.13275	41.87272	41.37249
32	36.71928	34.82381	34.79465	35.04391	34.73854
42	29.62316	29.04746	29.03886	29.51388	29.30517
51	11.73947	24.85184	24.86316	25.35809	25.20807

The graphical representation of the PSNR values from Table1, is shown in Figure 2

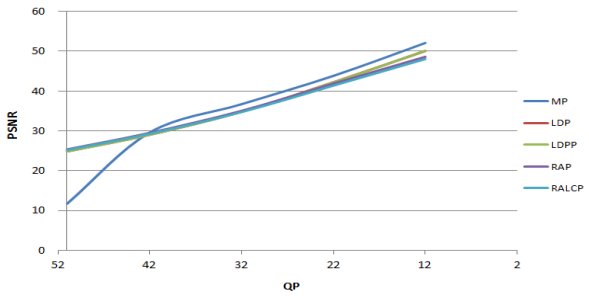


Figure 2 Average PSNR vs QP of HEVC and AVC signals

It can be observed that, the HEVC profiles have slightly degraded video quality, as compared to the Main Profile of AVC for QP values between 12 and 42. This degradation is between 1-2dB. Furthermore the AVC achieves slightly better video quality on low QPs, but on the other hand as the QP value gradually increases the AVC's encoding efficiency is reduced, reaching to an abrupt downfall on quality when the QP equals to 51. On the other hand HEVC performs much better than AVC as it scores PSNR at acceptable values, for QP higher than 42, while retaining its linearity. Additionally, the AVC's MP decoded video for QP equal to 51, proves not viewable (Figure 3), while HEVC RAP achieves acceptable video quality.

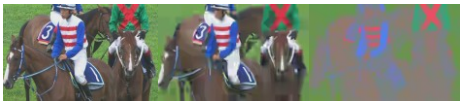


Figure 3. Original frame (on left) and decoded ones (HEVC in the middle, AVC on the right with QP=51)

5.2 Compression Efficiency Comparison

This section presents the experimental results of the comparison between HEVC and AVC encoded signals' bit rate. During the evaluation tests a set of QP values {12, 22, 32, 42, 51} was used in order to examine a variety of bit rate cases, and evaluate the compression efficiency of HEVC, when compared to AVC. The average values of the test results are shown in Table 2, which enlists the combined average bit rate for each profile of the four test video signals.

The graphical representation of Table 2 data is presented in Figure 3 and in Figure 4 zoomed in low bit rate cases. From Table 2, it is clear that the compression efficiency of HEVC is significantly better than AVC. For QP values from 12 up to 42, HEVC undoubtedly surpasses AVC, as it requires 32% to 62% less bitrate for all the video signals.

Table 2. Average Bit Rate vs. QP for H.264/AVC (MP) and HEVC (LDP, LDPP, RAP, RALCP)

QP	MP	LDP	LDPP	RAP	RALCP
12	10716.83	5080.182	5402.373	4176.731	4011.801
22	2054.2	1249.342	1316.253	1105.972	1083.826
32	582.2025	284.7245	286.4473	274.3813	277.2903
42	114.2425	78.63425	77.82075	78.31875	77.6965
51	6.8	34.1	33.64975	34.5805	33.2445

The huge margin between the performances of the 2 encoders can be clearly seen in Figures 3 and 4.

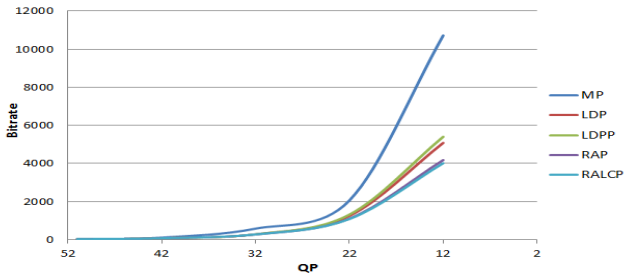


Figure 3. Average Bit Rate vs. QP for H.264/AVC (MP) and HEVC (LDP, LDPP, RAP, RALCP)

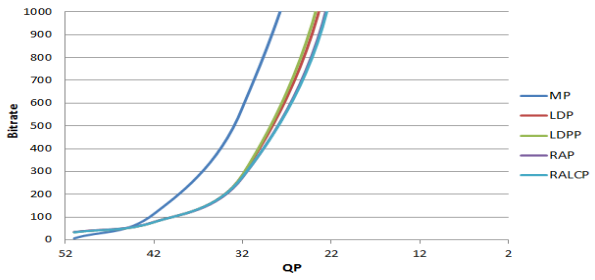


Figure 4. Average Bit Rate vs. QP for up to 1000 kbps.

However, for QP values between 42 and 51, AVC surpasses HEVC in compression rate, achieving significantly lower bitrate values. But as, it is thoroughly depicted in the previous section the video quality of AVC, in QP value equal to 51, is extremely low, making the decoded content not perceptually accepted.

In Figure 3, the apparent gap can be noticed separating the AVC average bit rate line from the HEVC ones, in a distant view including values up to 12kbps. Additionally in Figure 4 a more detailed and zoomed view of the same evaluation tests can be observed, portraying even better the compression efficiency of the two encoders in low bit rate cases.

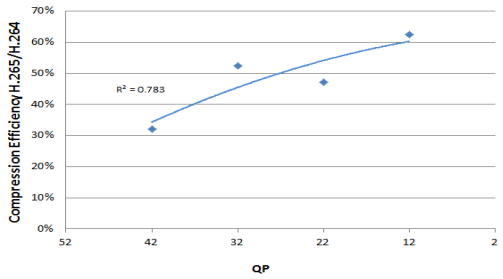


Figure 5. Average Compression Efficiency (%) vs. QP for HEVC RALCP and AVC MP.

A more comprehensive representation of the compression efficiency (%) of HEVC in comparison to AVC is depicted in Figure 5. The curve of this figure shows that for QP values from 12 to 42, the compression efficiency improvement of HEVC ranges from 32% to 62%. For values of QP between 42 and 51 the curve is intentionally not considered, as the quality of the decoded AVC video content is degraded to significantly low levels.

6. CONCLUSIONS

This paper performs a quantitative comparison between the HEVC and AVC encoders, in terms of video quality and compression efficiency. It is shown that the novel currently developing encoder achieves a 32% to 62% compression enhancement, while it attains 1 to 2dB reduced video quality, compared to its predecessor. Furthermore, HEVC in relatively low bitrates shows a linear performance in video quality, while AVC abruptly degrades to unsatisfactory video quality levels.

7. ACKNOWLEDGEMENT

Part of the work of this paper has been supported by the EC-funded GERYON project (SEC-284863)

8. REFERENCES

[1] ISO/IEC International Standard 14496-10, Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding, third ed., Dec. 2005, corrected version, March 2006.

[2] G. J. Sullivan, T. Wiegand, "Video compression from concepts to H.264/AVC standard," Proc. of IEEE, vol. 93(1), 18-31, Jan. 2005.

[3] F. Pereira and T. Alpert, "MPEG-4 Video Subjective Test Procedures and Results", IEEE Transactions on Circuits and Systems for Video Technology. Vol. 7(1), pp. 32-51, 1997.

[4] Kemal Ugur, Kenneth Andersson, Arild Fuldseth, Gisle Bjøntegaard, Lars Petter Endresen, Jani Lainema,Antti Hallapuro, Justin Ridge, Dmytro Rusanovskyy, Cixun Zhang, Andrey Norkin, Clinton Priddle,Thomas Rusert, Jonatan Samuelsson, Rickard Sj¨oberg, and Zhuangfei Wu "High Performance, Low Complexity Video Coding and the Emerging HEVC Standard", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 20, No.12, December 2010

[5] Link for the reference test video signals "ftp://ftp.tnt.uni-hannover.de/testsequences"

[6] HEVC reference toolkit available for download version directory https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware

Motion saliency for spatial pooling of objective video quality metrics

Marina Georgia Arvanitidou
Communication Systems Group
Technische Universität Berlin
Berlin Germany
arvanitidou@nue.tu-berlin.de

Thomas Sikora
Communication Systems Group
Technische Universität Berlin
Berlin Germany
sikora@nue.tu-berlin.de

ABSTRACT

In this paper we propose a novel motion saliency estimation method for video sequences considering the motion between successive frames and their corresponding parametric camera motion representation. Background motion is compensated for every pair of frames, revealing areas that contain relative motion. Considering that these areas will likely attract the attention of the viewer and in line with properties of the human visual system, regarding spatially invariant focus distribution, we augment their effect on the quality estimation. The generated saliency maps are thus incorporated in the spatial pooling stage of several video quality metrics, and experimental evaluation on the LIVE video database[1] shows that this strategy enhances their performance.

Keywords

video quality assessment, spatial pooling, motion saliency, global motion estimation

1. INTRODUCTION

The broad use of video in communication services, such as IPTV and video conferencing, resulted in the increasing interest of the image processing research community in the topic of objective Video Quality Assessment (VQA). As humans are the final judges of service quality, key issue is the development of algorithms that efficiently assess the quality experienced by users (QoE). The commonly acceptable way for assessing the video quality is to conduct a large scale subjective study where a group of observers are asked to provide their personal opinions of the video. This subjective Mean Opinion Score (MOS) can then be regarded as the ground-truth subjective quality of the video sequences. In practice subjective experiments are time-, effort- and resource- consuming, and therefore objective video quality metrics that can automatically evaluate the video's perceptual quality are appreciated.

[1]Publicly available at
http://live.ece.utexas.edu/research/quality/live_video.html

In this paper we focus on full reference objective quality assessment algorithms that employ both the reference and the distorted video sequences. Among the most widely used metrics in this category are the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR), which are straight-forward in implementation but do not consider any properties of the Human Visual System (HVS). Towards this direction a large variety of metrics have been proposed in the past. The Structural SIMilarity index (SSIM) [14] captures the structure information loss between the reference and distorted image to estimate the perceptual quality. Initially it was proposed in single scale, and multiscale versions of it [16] followed. The Visual Information Fidelity criterion (VIF) [11] is another powerful metric that employs the mutual information between the original and the distorted image to estimate the perceived quality. The Video Quality Model (VQM) [8], which is adopted by the American national standards institute, analyzes 3D spatio-temporal blocks to extract the salient features for estimating the video quality map, whereas the MOtion- based Video Integrity Evaluation (MOVIE) metric [9] utilizes properties of the visual cortex neurons to track perceptually relevant distortions along motion trajectories. The latter is a computationally complex metric since it relies in 3D optical flow estimation.

Existing works on the exploitation of temporal distortions in video sequences report improvements on the prediction performance of standard quality metrics by taking into account importance maps and appropriate pooling techniques. Moorthy *et al.* [6] propose a computationally efficient VQA algorithm that assesses the quality in block-based level and subsequently employs percentile pooling. Towards VQA enhancement Ma *et al.* [5] propose a visual saliency estimation algorithm, based on the quaternion Fourier transform of the motion vectors after block matching. Recently, the authors in [13] have proposed a video quality metric that employs the structural information contained in two descriptors extracted from the 3D structure tensors, and its corresponding eigenvector, whereas in [15] a model of human visual speed perception is incorporated to model visual perception in an information communication framework.

In line with the idea that the performance of VQA metrics can be improved by considering features along the temporal trajectories, in this paper we propose to exploit the motion information between successive video frames for estimation of motion saliency maps and employ them for spatial pooling of video quality metrics. By giving higher weighting to

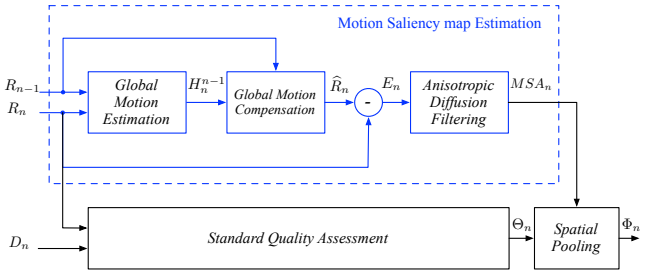


Figure 1: Proposed method overview

regions with detected motion, we expect these regions to attract humans' attention, and thus make distortion on these areas more perceivable. Experimental results on various video quality metrics evaluate the efficiency of our method.

The paper is organized as follows. Section 2 provides an overview of the proposed motion saliency estimation and describes the spatial pooling strategy that is employed for incorporation in objective video quality prediction algorithms. Section 3 presents and discusses the experimental evaluation of our algorithm and Ssection 4 summarizes and closes this paper.

2. PROPOSED MOTION SALIENCY MAPS AND VQA ENHANCEMENT

The main idea of our proposed method is to detect regions that contain significant relative motion between frames, and augment their effect to the image quality index in the spatial pooling stage that will follow. This means that if a distortion occurs in a region that contains motion, it is expected to attract the attention of the viewer and thus to have a negative impact on the quality assessment in comparison to a distortion that occurs in a region not containing motion.

2.1 Motion Saliency maps Estimation

Motion between successive frames is what differentiates a video sequence from a set of independent still images. Motion can be considered as a mixture of foreground object motion and camera motion. If we assume that the background (i.e. camera) motion is the dominant motion between two frames of a video sequence, then the foreground motion is likely to attract visual attention, according to the properties of the HVS that are explored with respect to this point of view in [15]. Based on this observation we propose the following strategy, which is illustrated in Figure 1.

First we estimate the perspective (eight-parameter) motion model that describes the background motion between two successive frames of the reference sequence $R_n - 1$ and R_n. This is realized as described in [3] where a set of feature-points are detected in each frame, the correspondences between the two sets of points are established for successive frames, and finally a motion model H_n^{n-1} is fitted to the correspondences. The RANdom SAmpled Consensus (RANSAC) method is employed for fast and accurate model fit based on features which are detected using the KLT feature tracker. Based on the estimated motion model H_n^{n-1} and the corresponding frame R_{n-1}, the estimated frame $\widehat{R}_n$ is computed and subsequently subtracted from R_n. This results in the (global motion) compensated absolute error frame E_n where high error energy corresponds mainly to motion of the foreground area.

Studies on the HVS properties have shown that the human retina is highly space variant in processing and sampling of visual information [2]. The accuracy is highest in the central point of focus, the fovea, and the peripheral visual field is perceived with lower accuracy. In our case, we consider the locations of the highest motion compensated error energy as the central points of focus, and to address the gradually decreasing focus, the error maps are low-pass filtered resulting in the motion saliency map MSA.

$$MSA(x,y,n) = \alpha * |\widehat{R}(x,y,n) - R(x,y,n)| \qquad (1)$$

Anisotropic diffusion filtering (α) [7] is employed here, as similarly employed for the scope of object segmentation in [1]. Anisotropic diffusion offers a non-linear and space-variant filtering of the error frame, that while having a low pass character preserves the edges of the image. In this way we give higher weighting to regions that have moved between two successive frames and we expect that they are more likely to attract visual attention in comparison to other areas that have not moved (or have moved with the background). As shown in Figure 2 our motion saliency estimation method can significantly detect the motion of the foreground (brighter areas) in the MSA map. Of course motion is not the only feature that attracts visual attention. Other features such as contrast, color and structural information will be considered implicitly through the incorporation in standard objective metrics that is following described.

2.2 Spatial Pooling

Standard image quality metrics tend to generate a quality index Θ between a reference and a distorted image (R and D respectively) and then consider that every pixel contributes equally to the overall image metric by averaging over all pixel locations. As we want to avoid this uniform spatial pooling, we employ the weighted mean spatial pooling strategy [2]. The estimated motion saliency maps are incorporated in MSE, SSIM [14], MS-SSIM [16] and VIF [11] metrics. The weighted mean for single scale metrics in (x,y) location of the nth frame [15] is formulated as

$$\Phi(x,y,n) = \frac{\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} w(x,y,n)\Theta(x,y,n)}{\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} w(x,y,n)} \qquad (2)$$

where Φ is the weighted metric and N_x, N_y are the frame dimensions. The squared error map serves as quality index in the case of MSE, and the SSIM index map is employed for SSIM. The proposed motion saliency maps $MSA(x,y,n)$ are used as weighting maps w, whereas in the following section, a local saliency map is additionally considered for comparison. For multiscale metrics, that use M scales, the weighting map is scaled correspondingly and the overall metric is calculated as following

$$\Phi(x,y,n) = \prod_{j=1}^{M} \frac{\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} w(x,y,n)\Theta(x,y,n)}{\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} w(x,y,n)} \qquad (3)$$

MS-SSIM incorporates SSIM evaluations in different scales. To this end, SSIM index maps are, weighted according to Eq. 2 in each scale and then the various weighted scaled indexes

are combined as described in [16]. For VIF [11], the mutual information (between the input and the output of the HVS channel) for the reference image and the mutual information (between the input and the output of the HVS channel) for the distorted image are separately weighted using Eq. 2, scaled to the corresponding scales and finally combined over multiple scales.

After the local weighted quality score of every frame $\Phi(n)$ is generated, by considering the motion saliency map, temporal pooling follows. The local scores are averaged over the T frames of the video sequence to yield the overall weighted quality score Φ.

3. EXPERIMENTAL EVALUATION

We evaluate the performance of our proposed algorithm on the above mentioned metrics on the LIVE video database [10], developed at the University of Texas at Austin. The LIVE database contains 150 distorted videos obtained from 10 uncompressed reference videos (768×432 pixels, 3206 frames totally) of natural scenes. The distorted videos are created using four commonly encountered distortion types. These include MPEG-2 compression, H.264 compression, simulated transmission of H.264 compressed bitstreams through error-prone IP networks, and through error-prone wireless networks. Each video was assessed by 38 human subjects in a single stimulus study with hidden reference removal, where the subjects scored the video quality on a continuous quality scale. The difference scores of a given subject are computed by subtracting the score assigned by the subject to the distorted video sequence from the score assigned by the same subject to the corresponding reference video sequence. Following the Difference Mean Opinion Score (DMOS) of each video is computed as the mean of the rescaled standardized difference scores (Z-scores) of statistically reliable subjects.

The quality prediction performance of the weighted metrics is evaluated using three performance indicators. As the Video Quality Experts Group (VQEG) recommends [12], we use the Pearson Linear Correlation Coefficient (LCC) and the Spearman Rank Order Correlation Coefficient (SROCC) between DMOS and the objective score after nonlinear regression. Additionally, we employ the Root Mean Square Error (RMSE) in the same manner. Non linear regression is applied in order to align each video quality metric output to the subjective rating scale using the logistic function described in [12]. Figure 3 illustrates an example scatter plot of predicted DMOS using standard MS-SSIM and weighted MS-SSIM using the proposed method (green and blue marks respectively) versus DMOS.

Table 1 shows the performance evaluation of various objective VQA algorithms. VS denotes the Visual Saliency model proposed in [5]. MSA is our proposed method, whereas local saliency denotes the employment of local saliency maps proposed by Itti & Koch [4] for weighting in the same manner as described in the previous section. For each evaluation metric we highlight the best results with boldface. Larger LCC and SROCC indicate better correlation between objective and subjective scores, while smaller RMSE is indicator of better performance. Saliency weighted metrics perform better compared to non-weighted metrics. Motion saliency spatial pooling proves to be more beneficial for objective VQA compared to local saliency pooling. This shows that motion saliency is beneficial for VQA metrics providing a better agreement with the subjective ground-truth scores. However, saliency weighted methods still cannot outperform MOVIE [9], which employs a complex HVS based model for exploiting temporal and distortions. Our proposed method does not seek to explicitly model properties of the HVS, however it is competitive with MOVIE, especially for the case of MSA-weighted MS-SSIM, even though it avoids computationally complex optical flow estimation or multi-scale filtering over large temporal trajectories. To examine the effect of our proposed weighting on different distortion types, we show in Table 2 the performance enhancement, in terms of Spearman Rank Order Correlation Coefficient, for our proposed method for each distortion class separately. As expected, our proposed method contributes on average more in cases of transient distortions (in the presence of packet losses, classes 1&2) compared to cases with uniformly distributed distortions (no packet losses, classes 3 & 4).

Table 1: VQA metrics performance on LIVE database. Data for VS from [5]

Algorithm	LCC	SROCC	RMSE
MSE	0.5614	0.5391	9.0839
MSE VS [5]	**0.6295**	**0.6268**	**8.5310**
MSE w=local saliency	0.5410	0.5267	9.2319
MSE w=proposed MSA	0.5669	0.5593	9.0427
SSIM	0.5411	0.5231	9.2315
SSIM VS [5]	0.6308	0.6187	8.5310
SSIM w=local saliency	0.6064	0.5825	8.7284
SSIM w=proposed MSA	**0.6470**	**0.6334**	**8.3698**
MS-SSIM	0.7556	0.7474	7.1911
MSSSIM VS [5]	0.7583	0.7468	7.1570
MS-SSIM w=local saliency	0.7623	0.7589	7.1042
MS-SSIM w=proposed MSA	**0.8009**	**0.7964**	**6.5726**
VIF	0.5322	0.5297	9.2936
VIF w=local saliency	0.6734	0.6566	8.1153
VIF w=proposed MSA	**0.6946**	**0.6959**	**7.8968**

4. CONCLUSIONS

In this paper we have provided a novel motion saliency estimation method for video sequences considering the motion between successive frames, and their corresponding parametric camera motion representation, for VQA. The proposed saliency maps are incorporated in the spatial pooling stage of several video quality metrics. Experimental eval-

Table 2: Performance enhancement in terms of SROCC for our proposed method on LIVE database

#	Distortion class	MSE	SSIM	MS-SSIM	VIF
1	H264 + wireless	-0.0291	0.1328	0.0638	0.1538
2	H264 + IP	0.1139	0.1166	0.0206	0.1326
	average (#1,#2)	0.0424	0.1247	0.0422	0.1432
3	H264	0.0251	0.1099	0.0901	0.1546
4	MPEG2	0.0238	0.1110	0.0662	0.0463
	average (#3,#4)	0.0245	0.1105	0.0782	0.1005
	All data	0.0202	0.1103	0.0490	0.1662



Figure 2: Example frames of four sequences of the LIVE database. Reference frames R_n and the corresponding motion saliency maps MSA_n (first and second rows respectivelly).

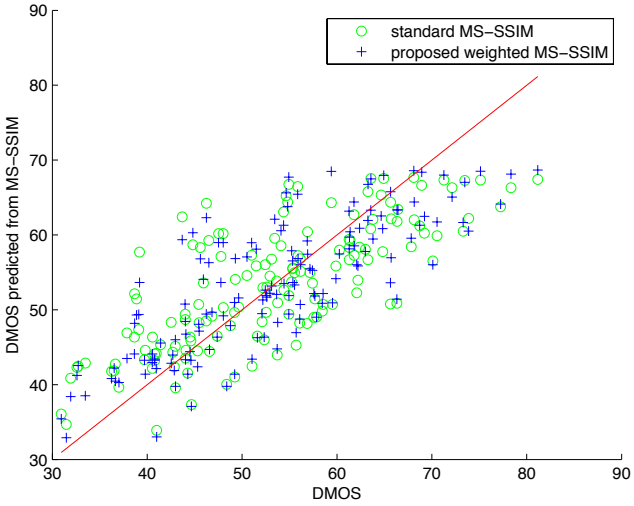


Figure 3: Scatter plot of predicted DMOS using standard MS-SSIM and weighted MS-SSIM using the proposed method (blue and green marks respectively) versus DMOS. Evaluation on LIVE video database.

uation on the LIVE video database has shown that thus, objective metrics are more closely in accordance with the subjective ground-truth scores, which is an indicator that motion saliency can be beneficial for VQA. Motion saliency aware temporal pooling and consideration of more neighboring frames remain interesting topics for future exploration.t

5. REFERENCES

[1] M. G. Arvanitidou, M. Tok, A. Krutz, and T. Sikora. Short-term motion-based object segmentation. In *Proc. of the IEEE Internl. Conf. on Multimedia & Expo*, Barcelona, Spain, Jul. 2011.

[2] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, and P. Ndjiki-Nya. Visual attention in quality assessment. *IEEE Signal Proces. Mag.*, 28(6):50–59, Nov. 2011.

[3] D. Farin and P. H. de Witha. Evaluation of a feature-based global-motion estimation system. In *SPIE Visual Comm. and Image Proces.*, volume 5960, 2005.

[4] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998.

[5] L. Ma, S. Li, and K. Ngan. Motion trajectory based visual saliency for video quality assessment. In *Proc. of the IEEE Internl. Conf. on Image Proces.*, pages 233–236, sept. 2011.

[6] A. Moorthy and A. Bovik. Efficient video quality assessment along temporal trajectories. *IEEE Trans. on Circuits and Systems for Video Technology,*, 20(11):1653–1658, Nov. 2010.

[7] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629 –639, Jul. 1990.

[8] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Trans. on Broadcasting*, 50(3):312–322, Sep. 2004.

[9] K. Seshadrinathan and A. C. Bovik. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans. on Image Proces.*, 19(2):335–350, 2010.

[10] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack. Study of subjective and objective quality assessment of video. *IEEE Trans. on Image Proces.*, 19(6):1427–1441, Jun. 2010.

[11] H. Sheikh and A. Bovik. Image information and visual quality. *IEEE Trans. on Image Proces.*, 15(2):430–444, Feb. 2006.

[12] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment, ph.II. http://www.vqeg.org, 2003.

[13] Y. Wang, T. Jiang, S. Ma, and W. Gao. Novel spatio-temporal structural information based video quality metric. *IEEE Trans. on Circuits and Systems for Video Technology*, PP(99):1, 2012.

[14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Proces.*, 13(4):600–612, Apr. 2004.

[15] Z. Wang and Q. Li. Video quality assessment using a statistical model of human visual speed perception. *Journal of the Optical Society of America*, 24(12):B61–B69, Dec 2007.

[16] Z. Wang, E. Simoncelli, and A. Bovik. Multiscale structural similarity for image quality assessment. In *Proc. of the Asilomar Conf. on Signals, Systems and Computers*, volume 2, pages 1398– 1402 Vol.2, Nov. 2003.

Profiling User Perception of Free Viewpoint Video Objects in Video Communication

Sara Kepplinger
Ilmenau University of Technology
PO Box 10 05 65
Ilmenau, Germany
sara.kepplinger@tu-ilmenau.de

ABSTRACT

This paper presents a study investigating influences on the subjective quality experience of free viewpoint video objects in video communication. The background is to use free viewpoint video objects to overcome the eye-contact problem. Therefore, different algorithms are used for the creation. This includes a variation in background and content presented in order to check the competition of the used algorithms in this case. Hence, the main aim is to find out which differences are experienced by the test participants when using different free viewpoint video algorithms to re-establish eye-contact. Herein, we assume that eye contact is mandatory for a video conversation leading to the perception of good quality. This work reports activities towards the evaluation of such visual representations. The work applies an open profiling of quality approach. In this early stage of development a fully operative video conversation system is missing (i.e. no audio, no real-time conversation). However, this work presents results on the monitored user perception of the created video representations.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems — *Evaluation / methodology*

General Terms

Measurement, Experimentation, Human Factors

Keywords

Free Viewpoint Video Object, Quality of Experience, Open Profiling of Quality

1. INTRODUCTION

New forms of video representation are developed for the facilitation of eye contact in video conversations. Figure 1 shows the eye contact conflict by displaying the presumed conversation window area with a dashed circle. The eye contact will not be perceived by the conversational partner as long as the person in front of it looks onto the display and not into the camera. Eye contact between the conversational partners will fail as long as the person in front of it looks onto the display and not into the camera. As pointed out

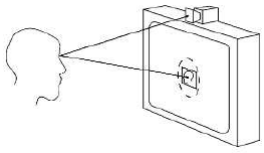


Figure 1: Eye contact problem (taken from [7])

in [14], new forms of representation make use of stereo 3D technology theory and create free-viewpoint video objects (3DVO) based on a pre-defined position. This visualization represents a video showing a recording (like an intended view at dashed circle in Figure 1) from a virtual camera position visualizing content similar to a real recording. Figure 2 exemplarily shows recordings (two upper pictures) and the resulting virtual view (the picture below). Weigel et al. [14]



Figure 2: Two recordings leading to a virtual view

create recordings by two web cameras within a parallel setup mounted on top of the display. They process the information with a modular creation chain containing the acquisition, segmentation, calibration, rectification, analysis and synthesis of available visual information. The aim is an optimal synthesis of natural video objects enabling a free viewpoint selection, or, as in this case, a pre-defined virtual view

as a result. New algorithms are developed within their work for a real-time synthesis on the variety of available video communication systems creating acceptable representations with a certain quality. This paper is about user perception of these representations. Therefore, it targets a parameterization on quality influencing factors within the perceptual domain in the area of 3DVO usage for video communication as introduced in [6]. The work intends the definition of so called quality features (see: [12], [5]) and their impact on the overall quality. It profiles users' perception by conducting an open profiling of quality (OPQ) study [13]. The goal is to define features influencing quality perceived subjectively by the user. This paper presents an overview about current methodological approaches, mandatory research problems, the OPQ method, results, and a conclusion.

2. STATE OF THE ART

Human perception and the human visual system (HVS) are the basis for the subjective quality definition and the metrics of visual representations. These existing measures, parameters, and assessment methods rate the general video quality objectively and focus on comparable reference values and technical feasibility. Problems occur concerning the involvement of pre-experiences, context of use, individuality of users, or the increasing complexity of the technical realization process. Guidelines (e.g. by the ITU [4] or EBU [1]) are available to guarantee a comparable procedure of subjective quality assessment in the context of multimedia visualizations and other different evaluation approaches (e.g. from the areas User Experience (UX), Quality of Experience (QoE), or both). However, the objective measurement of subjectively perceived quality influencing factors needs to be investigated further for example in terms of new forms of usage, context and user behaviour. Further, perceived quality may not change linear to the produced quality but similar to the perception of humans and androids as shown in [10] and face perception may be influenced by contextual factors as pointed out in [8]. Thus, and concerning new technical creation processes especially in the field of 3D technology, new developments ask for an inclusion of more sophisticated subjective measures in order to reach an adequate consideration of human perception. For 3DVO quality assessment, preliminary defined subjective quality features are derived, for instance, from occlusion, distortion, and shape, as outlined in [9]. However, the identification and the tighter definition of the extent of influences ask for further exploration. Quality estimation of emerging 3DVO technology in video communication has not been covered yet [11]. The question - beyond whether quality is good or bad - is *why* quality is rated like that. Towards this goal, we use OPQ for an evaluation of 3DVO as introduced in [13].

3. METHOD

OPQ is regarded as useful approach to construct a deeper understanding on subjective quality. Moreover, OPQ combines qualitative descriptive quality evaluation based on the individual's own vocabulary and quantitative psychoperceptual evaluation using standards (i.e. ITU-R BT.500 [4]). It is a mixed method approach targeted to naïve participants in experiments with miscellaneous stimulus material, for multimedia quality evaluation. The use of an OPQ approach helps to complement quality describing features and a definition of further influences including a valuation of impact. The test procedure starts with a visual screening of the participants and gives an explanation of the test procedure. The first quantitative part asks for general quality acceptance (yes or no) and the perceived overall quality based on a continuous quality scale (0-10) rating labeled with "bad" and "good" at its ends. Next, the sensory profiling takes place asking for individual quality describing attributes and their occurrence as suggested in [13]. The aim of this evaluation part is to offer the participants the opportunity to generate their individual descriptive vocabulary. Therefore, the exclusion of the evaluators' influences and the exclusion of a possible pre-definition of describing attributes are mandatory and can be ensured by the instructor. With this vocabulary, quality is rated in a next step. This specified individual attribute list is presented with an 11 point scale labelled with "min." and "max." per attribute where "min." means that the attribute is not perceived at all and "max." refers to its maximum pronounced form. Then, another quantitative part asks for eye contact perception (yes or no). Finally, the participants answer a questionnaire enquiring about demographical data, media experience, and social behavior. The test included 44 different video clips showing a speaking person which represents a possible conversational partner. These test items were created with six different algorithms and show different content variants to check wheather the different competences of the algorithms are perceived differently. Therefore, also the content varied with different persons: three men and one woman. Additionally, the eye contact was varied. For instance, one of the recorded men does not show any eye contact at all. The differences within the algorithms concern the disparity estimation, the post processing, and the virtual view synthesis. Out of 44 test items, four processed 3DVO did not undergo a horizontal correction. Additional distinctions were made concerning 1. the aggregation and 2. the pre-processing of the video material for the stereo analysis. This pre-processing includes a variation of the shown background: Once an artificial background is provided and once only a 50 % grey referred to as *no background*. These differently applied stereo analysis and synthesis processes are described in more detail in [6]. The test items were presented in randomized order, with a length of 10 seconds, a resolution of 640x480 pixels, and without audio. The presentation took place on a 19" liquid crystal display embedded in a graphic user interface intending to resemble a video conversation tool and replacing paper and pencil for answering the questions. The test was conducted under laboratory conditions considering ITU-R BT.500 [4]. This included the viewing distance for the participants, the ambient light conditions, the lighting of the display as well as the presented background colour on the display. 32 naïve participants (18 male, 14 female, age: 19-53) took part in the study.

4. RESULTS AND DISCUSSION

Overall, 40.5 % of the test participants rated yes, the presented quality is acceptable. With regard to the different provided algorithms, none of the test items was rated below 3.2 % and above 90.5 % of acceptance. The tendency was given as a low level of quality with 37.3 %, taking the median as a measure of central tendency. Taking into account the different stereo analysis and synthesis modes, these variations seemed to have an influence on the overall quality satisfaction when averaged over content (displayed men and

woman, with and without artificial background). Based on the continuous quality scale (0-10) figure 3 shows the mean experience score (MES) per used algorithm. This quality rating of the different algorithms includes all content variations. Although there seems to be a trend concerning algorithm 2 (MES: 3.61) and algorithm 6 (MES: 1.69) which is the one without horizontal correction, there are no significant differences beween the six algorithms. However, they are all in the low quality range.

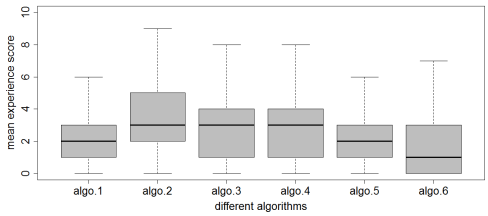


Figure 3: Mean experience scores

Within the sensory evaluation, all in all, the participants altogether developed 196 individual quality attributes valid for sensory analysis. The Generalized Procrustes Analysis (GPA) was applied, as suggested in [13] and described in [3], in order to find out how many components are needed to explain 100 % variance in the quality describing model in which components are useful for further interpretation. These dimensions form the perceptual space and can be seen as the description of the overall quality of experience. Attributes with a coefficient of determination (R^2) between 100 % and 50 % were emphasized and considered for further interpretation. A correlation plot can be derived from this showing the first two components (F1 and F2) which include the mentioned attributes from all participants and account for 64.79 % of explained variance (dimension 1, F1: 37.52 %, dimension 2, F2: 27.27 %) (see Figure 4. The positive polarity of dimension 1 included attributes like stripes, flicker representation, and fringed edges. Its negative polarity contained similar attributes like flicker and fringes. The positive polarity of dimension 2 had attributes related to holes, transparency, grey spots, and aliasing, whereas the negative polarity showed attributes like complete, mixed colour transitions, blurry and distorted. These attributes of these two dimensions can be seen as the description of the overall quality of experience. It was noticed that the positive polarity of both dimensions was stronger. The attributes positioned at the highest polarity with more than 75 % were about **holes, missing picture parts, disturbing pixel lines, lines, and stripes**. Here, too, this analysis was conducted for each of the mentioned algorithms seperately. Paying attention to the overall description and the single algorithms, outstanding quality describing attributes which appeared with high degrees of severity were **holes, occlusions in the picture** as well as **lines and stripes** in the picture. If locating the used algorithm within this perceptual space the GPA shows that algorithm 1 got a description with attributes of the highest positive polarity of dimension 2 and algorithm 5 got a description with attributes of the highest positive polarity of dimension 1. The measured eye contact perception over all data showed the indicated differentiation in perceived or not perceived eye contact (chi-square test, $< .0001$, df $= 43$). Regarding the different algorithms the measured eye contact perception varied within the algorithms (see Figure 5) and within the different contents (man presented without eye contact: 3.13 %, second displayed man: 59.5 %, third man: 85.8 %, woman: 86.25 %). However, there were also several differences between the test items created with the same stereo analysis and synthesis variants. Figure 4 shows that the acceptance rating (% of rated yes) was somehow connected to the perception of eye contact (% of rated yes).

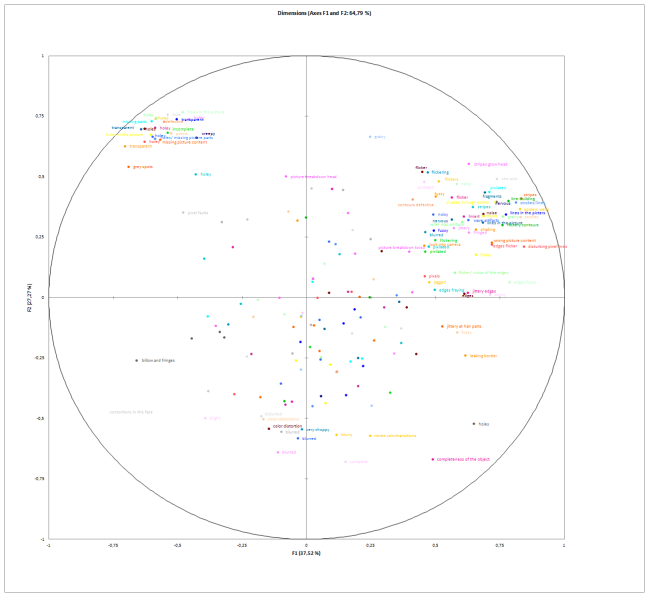


Figure 4: Correlation plot showing the first two dimensions F1 and F2

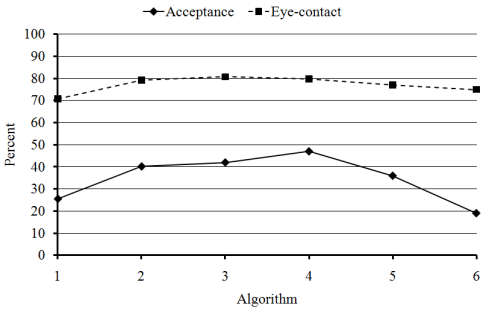


Figure 5: Acceptance vs. Eye Contact Perception

To sum up the results, information is available about the quality acceptance considering different creation processes and an overall quality rating. Quality describing attributes (e.g. holes, occlusions...) were collected and rated based on their severity, which as intended led to a tighter definition of quality features. Therefore, these strong or ever-present attributes are interpreted as quality features. Compared to already existing and defined influences (i.e. occlusion, distortion, shape) as described in [9], these features derive partly from the same creation processes. However, within this work detected subjectively perceived features additionally and especially concern not assignable areas named with lines or stripes. In order to evaluate the degree of depen-

dencies and to be able to design the appropriate quality of service, additional activities have to be undertaken. In a next step, combining the acceptance rating and the mean experience score may lead to an acceptance threshold. All in all, the assumption of rating the presented video clip as acceptable when perceiving eye contact has to be further investigated under consideration of this low quality range. Eye contact as a factor with a direct influence on quality acceptance was investigated and noticed as a trend. Compared to other detected quality features, it seems to constitute such a factor at the same or lower level. This, and the influence of background as well as the displayed persons' position may be a quality influencing matter, as changes of their appearance seem to be perceived differently. The differences beween with and without artificial background have to be regarded further as well as the factor horizontal correction (see algorithm 6). Combining the results from the sensory evaluation with the quantitative rating using external preference mapping may lead to more concrete results connecting the developed vocabulary with the single different algorithms in a further step. However, in general, the acceptance of the presented test items was not high. One problem was that these items were of no high diversity concerning their overall quality level. In spite of different creation processes, they were showing similar kinds of uncontrollable distortions without reference material. Probably another way of evaluation needs to be figured out in order to define the most severe and highly determining quality influences. To meet purposes and criteria which are not met by the OPQ method within this particular application, additional expert interviews with the developer and another evaluation with a higher variation in quality and used creation steps may be appropriate. To include the context factor "interactivity" and check for its influence, a Wizard of Oz study [2] may solve the problem of the real-time inability of the system. Additionally, a refinement of the usage environment to include influencing factors deriving from this should be considered in a next step.

5. CONCLUSION

This work provides insights into the subjective quality assessment of free-viewpoint video objects (3DVO) used for eye contact support in video communication. The work investigates the definition of quality influencing features as part of the perceptual domain. It intends the evaluation of effects of different algorithms on the perceived quality of 3DVO. In particular, it examines the assumption that perceived eye contact influences quality acceptance. Due to the tested system's complexity, the small diversity in a low quality range of test items, and the amount of influences, main concerns may refer to the study design and the validity and generality of the results. Nevertheless, information concerning possible quality determining features and their appearance could be collected and can be further investigated in the future. The advantage provided by the open profiling of quality (OPQ) method to combine quantitative quality rating with a sensory evaluation seems to be fruitful. Moreover, it is seen as a suitable alternative to evalution methods for subjective quality assessment actually not satisfying within this context of application.

6. ACKNOWLEDGMENTS

This work was conducted within the project no. BR1333/8-1 founded by the German Research Foundation (DFG).

7. REFERENCES

[1] T. Alpert and J.-P. Evain. Subjective quality evaluation - the SSCQE and DSCQE methodologies. Technical report, EBU Technical Review, 1997.

[2] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of oz studies Ů why and how. In *Intelligent User Interfaces*, pages 193–200, 1993.

[3] J. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[4] ITU. *Methodology for the subjective assessment of the quality of television pictures. ITU-R Recommendation BT.500-13.* International Telecommunication Union, Geneva, Jan. 2012.

[5] U. Jekosch. *Voice and Speech Quality Perception. Assessment and Evaluation.* Signals and Communication Technology, Springer-Verlag, Berlin Heidelberg, GER, 2005.

[6] S. Kepplinger and C. Weigel. Linking quality assessment of free-viewpoint video objects up with algorithm development. In *Proceedings of the 2012 IS&T/SPIIE Conference on Electronic Imaging 2012*, number 8293-29. IEEE Explore, January 2012.

[7] T. Korn. *Kalibrierung und Blickrichtungsanalyse für ein 3D Videokonferenzsystem.* Diplomarbeit, Technical University Ilmenau, Ilmenau, GERDiplomarbeit, Technical University Ilmenau, Ilmenau, GER, 2009.

[8] P. Mende-Siedlecki, C. Said, and A. Todorov. The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, January 2012.

[9] M. Rittermann. *Zur Qualit?tsbeurteilung von 3DVideoobjekten. Dissertation.* Technical University Ilmenau, Ilmenau, GER, 2007.

[10] A. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith. The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience*, 2011.

[11] O. Schreer, P. Kauff, and T. Sikora, editors. *3D Video communication: Algorithms, concepts and real-time systems in human centered communication.* John Wiley & Sons, Ltd., UK, 2005.

[12] A. Silzle. *Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey.* Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Dissertation, Shaker Verlag, Aachen, GER, 2008.

[13] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze. Open profiling of quality. a mixed method approach to understand multimodal quality perception. *Journal on Advances in Multimedia, Hindawi Publishing Corporation*, 2010(658980):28, July 2010.

[14] C. Weigel and N. Treutner. Establishing eye contact for home video communication using stereo analysis and free viewpoint synthesis. In *Proceedings of the 2012 IS&T/SPIIE Conference on Electronic Imaging 2012*, number 8290-2. IEEE Explore, January 2012.

Subjective evaluation of Video DNA Watermarking under bitrate conservation constraints

Mathieu Desoubeaux, Gaetan Le Guelvouit
Orange Labs, 4 rue du Clos Courtel 35512 Cesson-Sevigne, France

Yohann Pitrey
Entertainment Computing, University of Vienna, Austria

William Puech
France and Laboratoire LIRMM, 161 rue Ada, 34392 Montpellier Cedex 5, France

ABSTRACT

A popular and effective method to protect professional content against illegal distribution is to use watermarking, which inserts a robust and imperceptible fingerprint in the data using complex information hiding techniques. Video DNA watermarking is a state-of-the-art technique which enables shifting of the computational complexity to the server side. Despite the high level of protection it provides, this technique comes at a price in terms of storage space, as several versions of the same content have to be stored on the server. It is possible to limit the impact in terms of bitrate by slightly decreasing the quality of the final video. In this paper, we evaluate the impact of using Video DNA watermarking on the perceived quality by presenting the results of a standardized subjective experiment. We study various strategies to minimize the impact on the quality while maintaining the same overall bitrate, acting on the spatial resolution of the videos and bitrate repartition schemes.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Evaluation / methodology*

Keywords

DNA Watermarking, Video coding, subjective experiment

1. INTRODUCTION

As video data sharing becomes easier with current networks, protecting video contents against illegal redistribution is crucial for large-scale professional producers and providers. The construction of efficient and reliable content protection techniques is a difficult task, with systems dealing with hundreds of thousands of users. Cryptography and watermarking are two answers to this problem. Whereas cryptography releases an unprotected copy after the decrypting process, which is then easy to redistribute illegally, watermarking hides a fingerprint in each copy delivered to a client, which is never removed. The mark is inserted in a robust way, so that it cannot be erased from the data, except by applying severe treatments on it that make the content difficult to redistribute.

A serious question is how to embed the fingerprint at the time of a transaction without decreasing the service quality. "On the fly" digital watermarking can be processed either at the server side or at the client side. This is currently unacceptable for large-scale providing systems dealing with large numbers of users and very restrictive for client side approaches where the watermarking algorithm has to take the processing power of the client device into consideration. Video DNA is a server-side watermarking method which enables considerable savings in terms of computational complexity at the time of delivery for a protected video sequence to the client [2, 6]. The method relies on duplicated watermarked copies of the content, called video strands. Each strand contains a specific symbol from an alphabet, from which the fingerprint can be reconstructed. Then, the video distribution only consists of switching between strands before streaming. It allows using complex watermarking algorithms as the strands are calculated during a preprocessing step, while the needs in terms of network bandwidth do not increase and the technique is entirely transparent for the client.

However, the content provider needs to store as many versions of the original video content as symbols in the alphabet used in the fingerprints. This naturally has a significant impact on the required storage space and represents a major drawback of Video DNA watermarking. Nevertheless, it is possible to reduce the impact of this technique on the overall storage space, by accepting some limited loss in the quality. In this paper, we consider the use case of content providers who want to protect their content using Video DNA watermarking, but cannot afford a significant increase in storage space. We study the impact of different watermarking and encoding strategies on the perceived quality under the constraint that the overall bitrate should not increase with the DNA method. Additionally, we evaluate the increase in quality that can be expected when accepting a given increase in terms of global bitrate. Following the ideas presented in [3], we evaluate the possibility of using lower spatial resolution combined with upscaling to reach the same quality as full resolution, under the same bitrate constraint. The impact of bitrate repartition between the video strands is also studied, based on the results presented in [4]. A standardized subjective experiment was conducted to evaluate the influence of the studied parameters, following the international recommendations from the ITU. The

results are analysed in a systematic way to draw meaningful conclusions on the impact of Video DNA Watermarking on the perceived quality.

2. VIDEO DNA WATERMARKING

The video DNA watermarking [2] process can be divided in two parts. First, the video strands are generated by embedding symbols in each frame of the original uncompressed content coming from the producer and then encoded and stored on the server. The second step occurs upon user request on a protected content. The content provider associates a fingerprint to the client, either generated in real time or stored in a secured database, which contains a string of symbols refering to video strands. These symbols are used to build a watermarked video by switching between groups of pictures from the video strands. This process is illustrated in Figure 1, for two symbols in the case of binary fingerprint. Choosing a robust fingerprinting technique is

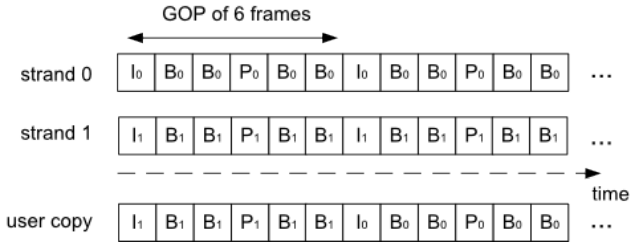


Figure 1: Watermarked video generation upon user request from two video strands. In this example, the user fingerprint contains the binary sequence '10', thus the final video contains one first GOP of frames marked with a '1', followed by a GOP marked with a '0'.

an important part of the process. It has to be robust to both video processing-based attacks and collusion attacks. Video processing-based attacks apply treatments on the video to make the mark undetectable, such as rotating, flipping, recoding, scaling or slightly modifying the colours or luminance of the video. To be robust to this kind of attack, the technique we use in this paper is the Broken Arrows technique [1], which is known for offering high levels of robustness, security and imperceptibility. Collusion is another common type of attack on a watermark, in which a group of dishonest users decides to forge an untraceable content by observing the differences between their personal copies. To be robust against collusion, we use Tardos's optimal codes [5], which is a state-of-the-art fingerprinting technique. These codes are based on a probalistic model that ensures that the fingerprint detected from an illegal copy is correlated with all users fingerprints that were used to generate it. Thus, if the correlation score between a colluded copy and a user's fingerprint is above a given threshold in a suspicious copy, the user can be accused with a probability of error fixed by the length of the code. The watermark embedding process comes with a high computational complexity if the mark is processed for each frame in the video. A way to reduce the complexity is to calculate a watermark over the average of several successive frames, then insert this watermark on each frame in the interval.

The two strategies are not only different in terms of computational complexity, but also in terms of robustness and visual impact. Typical watermarking techniques modify only high and mid-range frequencies in the images, which creates distortions around the edges of video objects. When the mark is calculated on each frame, flickering effects can appear due to the small changes in the video content. On the opposite, inserting a watermark calculated on an average frame can introduce ghosting effects, as the video content changes from one frame to another, while the mark remains identical. Both effects can be perceived as quite disturbing by human observers, which calls for subjective comparison between the two approaches. As the ghosting effect is simple to recognize in high motion video sequences, it becomes easy to remove. Therefore the interval of successive frames considered is this embedding strategy has to be kept as low as possible.

3. EXPERIMENTAL DESIGN

In this paper we consider binary DNA watermarking, in which only two video strands are generated per sequence because it appears that minimizing the number of strands is the best way to limit the increase in data to be stored, and therefore to limit the impact on the perceived quality.

We study two application scenarios over 12 processing conditions, or Hypothetical Reference Circuits (HRCs). Table 1 presents a description of these HRCs. HRCs 0, 1 and 2 are used as references in order to evaluate the impact of the tested parameters. HRCs 3 to 11 are designed in a systematic way and allow us to evaluate both the individual impact of the studied parameters, but also to get a view on the combined influence of these parameters.

We compare two main encoding scenarios. In the first scenario we constrain the overall bitrate to remain unchanged when switching from one unmarked video to two Video DNA strands (HRCs 3, 4, 5, 6, 7 & 8). This scenario allows us to evaluate the loss in quality that is to be expected in case the content provider can not afford an increase in overall bitrate. The second scenario relaxes this constraint by accepting a limited increase in bitrate (HRCs 9, 10 & 11). It aims at evaluating the necessary increase in bitrate in order to get an equivalent level of quality between the original content and the Video DNA-marked content.

Under a given constraint for the overall bitrate of a marked video, different strategies for the repartition of the bitrate between the two strands can be employed. An equal repartition (HRCs 3, 5, 7, 8, 9 & 11) between each strand ensures a consistent quality throughout the sequence, but means that the overall quality of each strand is relatively low. Unequal repartition (HRCs 4, 6 & 10) might be interesting as well (e.g.: 75% of the bitrate for one strand and 25% for the second one). One of the strands is then in better quality than the other, and the overall feeling of quality might be increased. It has been shown in [4] that upscaling a QVGA video sequence encoded with good quality to VGA can be well accepted by human observers. Such a video encoded under a given bitrate constraint can reach equivalent or higher quality when compared to the same content encoded directly in VGA format with an equal bitrate. Thus, we use this result to build a second type of encoding strategy (HRCs 5, 8 & 11) for Video DNA.

Finally, we want to compare the impact of the watermarking process when calculated on each frame (HRCs 3 & 5) and on a short interval of frames (HRCs 7 & 8). The length of

Table 1: Description of the HRC included in the experiment.

HRC	Resolution	Watermarking	Bitrate
0	VGA	none	uncompressed
1	VGA	none	1000
2	VGA	buffer	1000 / 1000
3	VGA	buffer	500 / 500
4	VGA	buffer	750 / 250
5	QVGA	buffer	500 / 500
6	VGA / QVGA	buffer	750 / 250
7	VGA	dynamic	500 / 500
8	QVGA	dynamic	500 / 500
9	VGA	buffer	750 / 750
10	VGA	buffer	1000 / 500
11	QVGA	buffer	650 / 650

the interval is set to 6 frames, which allows a significant reduction of the complexity of the watermarking step, while keeping ghosting artifacts at an acceptable level.

The input videos are generated from high quality uncompressed YUV 4:2:0 videos in VGA format (680 × 480 pixels) at 30 frames per second. Seven VGA reference video sequences were processed, containing a wide variety of contents and activity levels. Each video content was processed by the 12 HRCs, to produce 84 Processed Video Sequences (PVS). First, the original video contents were watermarked using the technique presented in section 2. Two video strands were generated for each content, one marked with symbol '0', the second marked with symbol '1'. This operation was repeated two times for single frame and 6 frames-interval strategies. Each video strand was then encoded using the x264 encoder with constant bitrate as described in Table 1. The GOP size was set to 12 frames, which corresponds to approximately half a second. To simulate the video delivery according to the DNA protocol, we generated a set of 32 binary Tardos codes containing a succession of 30 bits, which corresponds to the number of GOPs in one sequence. Each PVS was affected a random fingerprint, and the final video sequence was reconstructed following the Video DNA technique described in section 2.

To evaluate the visual quality of the presented scenarios, we used a standard Absolute Category Rating (ACR) with a 5-level scale, such as described in the ITU-T P.910 Recommendation. The test was presented to 11 non-expert viewers (age range: 18–25). Each test session took 35–40 minutes, including 15 minutes of introduction and 25–30 minutes of actual evaluation. Visual acuity was tested with a Sneller Plate and a Ishihara color plates test. A total of 84 videos of 12 seconds each were presented to each observer. They were displayed on a 24 inch diagonal consumer-range screen without any upscaling. After a short training to help the observers calibrate their judegement, they were asked to rate each video, presented in a semi-random order, making sure that the same content did not appear two times consecutively. The viewing distance between the observer and the screen was equal to 5 times the height of the displayed videos on the screen, according to the ITU recommendations for standard resolution videos. After each session, we asked the observers to give comments on their experience with the different levels of quality and types of artifacts. We summarized their comments and use them as additional information in the result analysis.

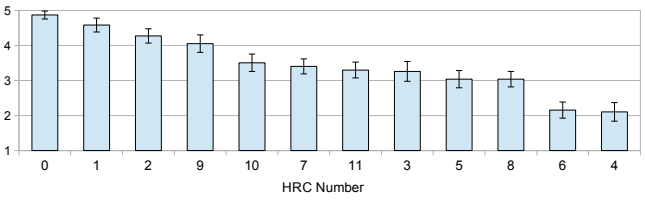


Figure 2: Condition MOS for the tested Video DNA strategies.

4. EXPERIMENTAL RESULTS

The results of the subjective experiment are presented in this section. We first propose a statistical analysis of the observers votes following a traditional procedure, and then provide a qualitative analysis of the comments gathered from the observers after each test session.

4.1 Statistical results analysis

We first analyse the influence of the three different varied factors in terms of Mean Opinion Score (MOS) per HRC and per source sequence. To give an overview of the variability of the results, we use 95% Intervals of Confidence together with the MOS per HRC. In Figure 2, the intervals of confidence are displayed as vertical line segments at the top of each bar. They represent the variability of the opinions of the viewers on the quality scale and can be used as a first approximation for the significance of the difference between MOS. If two intervals of confidence overlap, the difference between the corresponding MOS is considered non-significant. To verify the statistical significance of the difference between conditions, the following statements are based on the results of a student t-test.

As expected, HRC 0 obtains the best quality, as no coding or watermarking artifacts are visible. The loss of quality from HRC 0 to HRC 1 is very limited, which shows that coding artifacts at the high bitrate of 1000 kbps for a VGA video is well accepted by observers when compared to the other scenarios. The impact of the watermark alone when no constraint is applied on the overall bitrate can be evaluated by comparing HRCs 1 and 2 on Fig. 2. One can observe a slight difference in MOS between the two HRCs, however the results of the student t-test showed that this difference is not statistically significant. This is a first indication on the fact that the influence of the watermark is masked by the influence of other parameters such as upscaling and coding artifacts.

Additionally, the difference between calculating the watermark on each frame and on a buffer of 6 frames can be evaluated by comparing first HRCs 3 and 7, then HRCs 5 and 8. For these two pairs of HRCs, the MOS values are very close to each other. This is a second indication on the fact that artifacts due to pure watermarking do not have a significant impact on the visual quality, in the context of our experiment. This argument goes in favour of the buffered approach, which demands less calculation than the dynamic approach, while providing the same level of robustness.

We can compare the impact of upscaling artifacts to coding artifacts under the same bitrate constraint by observing HRCs 3 and 5. A slight advantage is in favor of HRC 3, which is encoded in VGA format. However, the difference is not statistically significant.

A second level of analysis to compare upscaling with coding is to look at the scenarios for which we tolerate

an increase in bitrate. Comparing HRC 1 with HRCs 9 and 11, one can notice that the increase in bitrate of the VGA version allows to enhance the quality and almost reach an equivalent quality as the coded reference. On the opposite, the QVGA version does not obtain a better quality than the other QVGA scenarios. This is due to the fact that upscaling produces artifacts independently from the quality of the upscaled video. Therefore the quality saturates when reaching high bitrates, whereas the blocking artifacts present in the native VGA version can be reduced significantly by raising the bitrate. As a result, strategies based on upscaling seem to be adapted when the overall bitrate conservation constraint is strong, whereas strategies using the native resolution of the videos are adapted when a slight increase in bitrate can be tolerated.

To compare equal bitrate repartition between strands to unequal repartition, HRCs 3 and 4 can be used. It appears quite clearly that unequal repartition is not well accepted by the observers. This result is further confirmed when comparing HRCs 9 and 10, with a significant preference for the scenario where the two strands are encoded with the same bitrate. This result can be explained by the high level of annoyance induced by quality changes during playback, which are perceived as very unnatural by the observers. As expected, the difference in quality is even higher between HRCs 3 and 6, in which upscaling/coding artifacts are cumulated with quality changes between the two strands. These observations show that the quality of the two strands should be identical in order to minimize the impact on the perceived quality.

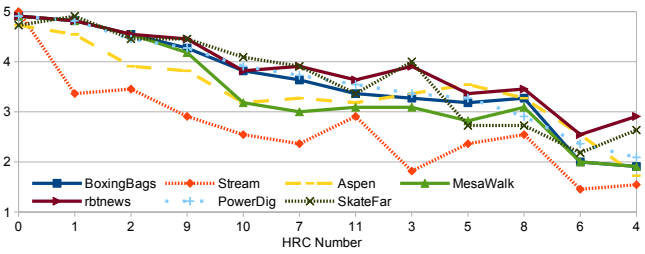


Figure 3: Detailed MOS values per HRC and for each source video (SRC).

Figure 3 provides more precise information about the behaviour of each video content under the 12 HRCs. One can observe the same global tendency as in Figure 2. However, the content named "Stream" reaches a noticeably lower quality than the other sequences. This content features a scene with complex fluid motion patterns and a high level of texture, which are known to suffer more from coding artifacts. For this particular content, only the reference HRC reaches high quality, while even the supposedly well accepted HRCs 1 and 2 only reach average quality. Additionally, HRC 11 reaches a significantly higher quality than HRC 3. This can be explained by the fact that the coding artifacts are less visible in the upscaled QVGA video, and probably masked by the blur introduced by the upscaling process. Surprisingly, an opposite phenomenon can be observed on the sequence named "SkateFar", as HRC 3 gets a significantly higher score than HRC 11. In this content, the background is static and relatively uniform, which could demonstrate that for simpler contents, coding artifacts appear as less severe than upscaling artifacts.

4.2 Observer comments

In their comments after taking the experiment, a significant amount of observers mentioned that they could see different types of distortions. When asked to draw a list of them, a large majority was capable of distinguishing between blockiness, bluriness and quality changes. Most of them were able to tell a difference between blockiness and blurriness, and mentioned that they would prefer seeing blurry videos than videos with visible blocks. This goes in favour of the watermarking strategies involving upscaled versions of the videos. This tendency disagrees with our statistical observations, which is relatively difficult to understand. A pair comparison experiment would allow us to learn more about this preference for upscaled videos. Some observers mentioned that the blur introduced by the upscaling step was relatively uniform and easy to accept, whereas coding distortions would create heratic artifacts and attract attention to them. Regarding quality changes, it was quite clear for most of the observers that they did not want to watch a video in which the quality would oscillate between relatively good and bad. When asked if they would prefer watching a video with a globally good but changing quality, or rather watch the same video with comparably lower but consistent quality, most of the observers stated a preference for the second video. This confirms our first observations on the MOS values and adds one more argument in favour of equal bitrate repartition between video strands.

5. CONCLUSION

Under bitrate conservation constraints, we showed that the Video DNA watermarking method has a significant impact on the perceived quality. As such, this technique might not seem applicable from a provider's point of view. However, accepting some increase in the bitrate can lead to acceptable quality. The results from our subjective experiment demonstrated that switching between different bitrates significantly decreases the quality and should not be used. Upscaling was observed as statistically equivalent to native full resolution, whereas observers mentioned a clear preference for the former type of distortion. Finally we showed that the watermarking algorithm has no significant impact to the quality compared to the encoding process. Further work will involve refining the subjective quality evaluation using a wider range of watermarking algorithms, higher resolution content and higher bitrates.

6. REFERENCES

[1] T. Furon and P. Bas. Broken arrows. EURASIP J. Inf. Secur., 2008.

[2] I. Kopilovic, V. Drugeon, and M. Wagner. Video-DNA: Large-Scale Server-Side watermarking. 2007.

[3] Pitrey, Y. and Barkowsky, M. and Pepion, R. and Le Callet, P. Subjective Quality Evaluation of H.264 HD Video Coding versus Spatial Upscaling & Interlacing. In QoEMCS Workshop, EuroITV, 2010.

[4] Pitrey, Y. and Engelke, U. and Barkowsky, M. and Pepion, R. and Le Callet, P. Subjective Quality Of SVC-Coded Videos With Different Error-Patterns Concealed Using Spatial Scalability. IEEE EUVIP, 2011.

[5] G. Tardos. Optimal probabilistic fingerprint codes. Journal of the ACM, 2008.

[6] F. Xie, T. Furon, and C. Fontaine. On-off keying modulation & tardos fingerprinting. In ACM workshop on Multimedia & security. ACM, 2008.

Subjective experiment dataset for joint development of hybrid video quality measurement algorithms

Marcus Barkowsky[1], Nicolas Staelens[2], Lucjan Janowski[3], Yao Koudota[1], Mikołaj Leszczuk[3], Matthieu Urvoy[1], Patrik Hummelbrunner[4], Iñigo Sedano[5], Kjell Brunnström[6]

[1]LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597, France,
[2]Ghent University - IBBT, Department of Information Technology, Ghent, Belgium
[3]AGH University, 30 Mickiewicza Av. PL-30-059 Krakow, Poland
[4]University of Vienna, Entertaiment Computing Research Group, Austria
[5]TECNALIA, Telecom Unit, Zamudio, Spain
[6] Acreo AB - NetLab: IPTV, Video and Display Quality, Kista, Sweden

ABSTRACT

The application area of an objective measurement algorithm for video quality is always limited by the scope of the video datasets that were used during its development and training. This is particularly true for measurements which rely solely on information available at the decoder side, for example hybrid models that analyze the bitstream and the decoded video. This paper proposes a framework which enables researchers to train, test and validate their algorithms on a large database of video sequences in such a way that the – often limited - scope of their development can be taken into consideration. A freely available video database for the development of hybrid models is described containing the network bitstreams, parsed information from these bitstreams for easy access, the decoded video sequences, and subjectively evaluated quality scores.

Categories and Subject Descriptors

H.1.2 **[User/Machine Systems]**: *Human factors*
H.2.4 **[Systems]**: *Multimedia databases*
H.5.1 **[Multimedia Information Systems]**: *Evaluation / methodology*

General Terms

Human Factors, Standardization, Algorithms, Measurement, Performance, Design, Experimentation, Verification.

Keywords

Subjective experiment, objective video quality measurement, freely available dataset, video quality standardization

1. INTRODUCTION

A wide variety of applications requires the assessment of video quality, ranging from the measurement of end user satisfaction to the inclusion of objective measurements in the encoding and distribution chain enabling an optimal allocation of bandwidth to perceptually important information.

Figure 1 shows a typical transmission chain from the camera capturing to the end user's perception. The individual steps may contain a large variety of features, ranging from a broadcast distribution chain with high quality camera capture and high bitrate transmission to user generated content shown on a mobile device. The expectation of the respective user in these two scenarios differs largely and, similarly, the video quality measurement needs to include different tools: From the prediction of just noticeable differences using, for example, the contrast sensitivity function [1] to the prediction of a frame freeze lasting for a whole second [2].

The development of video quality measurement algorithms often splits into the development of individual indicators which are targeted towards the measurement of a specific feature. These

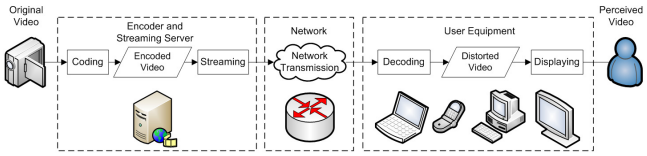


Figure 1 – Transmission chain (source: [3])

indicators may either measure content features and characteristics, such as the temporal activity [4] and the spatial activity (based on the color [5] or texture [6] layout), or they may detect typical degradations of videos. There is a wide variety of measurement algorithms, such as blockiness, blurriness, and noisiness [7], contrast metric [8], exposure metric [9], and measuring flicker [9]. In the context of multimedia applications additional indicators are necessary such as the annoyance introduced by a reduction of framerate or spatial resolution, the detection of skipping, and the influence of concealment algorithms used to circumvent outages due to lost packets or delays. Each such indicator has a certain scope. This scope describes the degradations which it is supposed to measure correctly. Out of scope degradations should not influence its results. However, most indicators are sensible to a number of degradations without necessarily providing a correct answer. For example, frame rate measurement often reacts also to video pauses due to delay in the transmission. When combining several different indicators into a single quality measurement algorithm, these side effects may have a large impact on the accuracy of the provided combined result.

The integration of the measured quality indications may become very complex due to their interaction. Typical examples for the creation of compound metrics are presented in [10] (global motion detection, spatial gradients, color, contrast) and [11] (spatial correlation, energy, homogeneity, variance, contrast).

The Joint Effort Group (JEG) of the Video Quality Experts Group (VQEG) invites researchers to perform joint work on the implementation and characterization of individually developed indicators as well as the combination of these indicators into objective measurement algorithms for the measurement of video quality based on the transmitted video bitstream and the decoded video sequence, so called hybrid no-reference models (HNR)[1]. The advantage of HNR models as compared to video only Full Reference (FR) models is their applicability on the client side. In

[1] Video quality models can be Full-Reference (FR), Reduced-Reference (RR) or No-Reference (NR) depending on how much of the reference information is utilized. In addition, the models could be bitstream, hybrid or video only. Traditionally FR, RR and NR is used for video only models. Hybrid models will be marked with an H before e.g. hybrid no-reference (HNR)

order to analyze the quality, they can rely on the decoded video, combined with the auxiliary bitstream information. This is also what sets the hybrid approach apart from the traditional video only no-reference approaches and has the potential to be more successful. The work targets proposals for ITU Recommendations, similar to the development of video coding algorithms such as ITU-T H.264 and HEVC.

One of the prerequisites for the development of hybrid video quality measurement algorithms is the availability of a large number of subjectively evaluated video databases in order to avoid overtraining. Several subjective experiment campaigns have already been conducted. In several cases, the decoded video sequences have been made available. One of the largest efforts performed recently was the HDTV evaluation, conducted by VQEG consisting of five publicly available datasets with 168 sequences each [12] for validating objective assessment methods for FR metrics. Unfortunately, for most of these sequences, the bitstreams are no longer available, hindering the development of bitstream models and hybrid models that would use the decoded video and the information transmitted over the network.

In this paper, a publicly available subjective dataset is described which includes various types of degradations so that different indicators can be evaluated. The data format for the bitstream data has been chosen by VQEG-JEG to be a simple XML file which allows for easy access to the relevant network and bitstream information. A method for evaluating the performance of an indicator is proposed as well, including an example evaluation.

The paper is organized as follows: Section 2 describes the subjective experiment, Section 3 discusses the example evaluation of an indicator and Section 4 concludes the paper.

2. Subjective experiment setup

The subjective experiment follows a full-matrix approach, 10 source reference sequences (SRC) were processed with each of 16 degradations. The SRC, listed in Table 2, were selected to spread a large variety of different content in Full-HD 1920x1080p25 format. The duration of each sequence was 10 seconds.

The different degradations, called Hypothetical Reference Circuits (HRC) in VQEG terminology, are listed in Table 1.

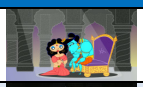
As an example for the second stage shown in Figure 1, the reference video sequences were encoded with either x264 in the version 0.120.x or the JM reference software encoder 18.2. Either, a fixed bitrate (R, in MBit/s) was used or the quantization parameter (QP) was chosen as a fixed value, resulting in a variable bitrate but an approximately constant quality over the duration of the video sequence. Several different GOP structures were configured, where the notation indicates the number of repetitions of the frame types and the GOP size is provided at the end. By using Immediate Decoder Refresh (IDR) images, closed GOPs were forced, allowing for an immediate error recovery at the start of each GOP.

Except for HRC14, one slice extended one macroblock line. Motion search range was set to 16, except for HRC7 where it was set to 8. In HRC8 the temporal resolution was reduced by skipping every second frame and duplicating frames at the decoder side. HRC9 contains a simulation of spatial downsampling during the transmission chain by using Lanczos-3 filtering. HRC10 simulates a transcoding scenario which re-encodes a strongly degraded sequence with a higher bitrate, thus spending bitrate to reproduce previously introduced coding artifacts.

Table 1 – List of HRCs considered in the subjective dataset

HRC	Remarks	Encoding			Packet loss	Deco-ding
		Enc.	R/QP	GOP		
0	(Reference)					
1		x264	16/-	IB7P64		JM
2		JM	-/32	IBBP32		JM
3		JM	-/38	IBBP32		JM
4		x264	8/-	IB3P16		JM
5		x264	4/-	IB7P64		JM
6		x264	1/-	IB7P64		JM
7		x264	-/32	IB3P16		JM
8	FPS ↓2	x264	8/-	IB3P16		JM
9	Res ↓2	x264	8/-	IB3P16		JM
10	Enc. JM IBBP32 Dec. JM	JM	-/44	IBBP32		JM
11		JM	-/32	IBBP32	Gilbert weak	JM
12		JM	-/32	IBBP32	Gilbert strong	JM
13		JM	-/32	IBBP32	Gilbert strong	ffmpeg
14		x264	8/-	IB3P16		JM
15		JM	-/32	IB3P16	Gilbert weak	JM
16		JM	-/32	IBBP32	Random strong	JM

Table 2 – List of SRCs

SRC	Thumbnail	Description
1		Sita Sings the Blues: Colorful animation with limited motion
2		Basketball court: Attention is on small objects moving fast (players)
3		Basket: Fast moving players with recognizable faces, fast camera pan
4		Cheetah: Diagonal structure in chainlink fence behind object of interest, slow camera pan
5		Lion: Strong contrasts due to sun on snow, scene cuts
6		Rotating collage: Objects with saturated colors spinning on a turntable, strong color and brightness contrasts
7		Lab: Highly structured due to small objects, camera adapts to illumination change
8		Manor house: Several shades of green with finely textured trees, helicopter shot with zoom-like motion
9		Zoo: Rapidly changing shots of animals in a zoo, wide variety of scene contents
10		Escalators: User generated content, Hall with three escalators with strong brightness contrasts due to point reflections, handheld camera

All sequences were streamed using RTP encapsulation without further multiplexing using the Sirannon software [13]. The packet trace was captured with tcpdump.

The network stage, shown as the third stage in Figure 1, may contain packet losses. These were introduced by simulating packet losses on the UDP level using Sirannon with either a random loss model or a Gilbert network channel model.

The decoding in stage 4 used mostly the JM decoder version 16.1, except for HRC 13 in which the ffmpeg decoder was used. In particular, HRC 12 and HRC13 differ only in the decoder, because they use different error concealment strategies: the frame copy concealment strategy was used with the JM decoder, while the motion copy strategy was used with ffmpeg. While the network data is therefore exactly identical, the decoded video sequence may include different types of artifacts.

2.1 Video quality scores

In order to build a dataset of subjective scores, an Absolute Category Rating with Hidden Reference (ACR-HR) test on a 5 point scale as described in ITU-T P.910 was performed in order to assess the video quality of each processed video sequence. The viewing environment corresponded to ITU-R BT.500. Screening tests were performed to ensure that observers have a (corrected-to) normal visual acuity (Snellen), and no color deficiencies (Ishihara plates).

A TV-Logic LVM-401W 40" display was used to display the sequences and calibrated to match ITU-R BT.500 and VQEG guidelines for TFT displays. The viewing distance was set to three times the height of the screen, which is 150 cm.

Table 3 – MOS obtained through the ACR-HR evaluation

SRCs	1	2	3	4	5	6	7	8	9	10	Mean	Std
REF	4.6	4.6	4.2	4.2	4.8	4.5	3.6	4.4	4.7	3.9	4.4	0.6
HRC 1	4.8	4.5	4.2	4.5	4.8	4.6	3.6	4.2	4.7	3.9	4.4	0.6
HRC 2	4.0	3.9	3.9	4.0	4.2	3.7	2.6	2.9	3.6	3.6	3.6	0.8
HRC 3	2.4	2.3	2.6	2.4	2.9	2.2	1.7	1.8	1.8	2.5	2.3	0.7
HRC 4	4.8	4.3	4.1	4.3	4.7	4.2	3.5	4.3	4.5	3.5	4.2	0.7
HRC 5	4.7	4.5	3.7	4.1	4.7	2.7	3.6	4.2	4.4	3.6	4.0	0.7
HRC 6	2.9	3.1	1.2	1.0	3.7	1.0	2.8	2.7	1.3	2.0	2.2	0.6
HRC 7	4.0	3.8	4.0	4.4	4.3	4.2	2.9	3.5	4.0	3.7	3.9	0.7
HRC 8	3.2	2.5	2.0	2.6	4.2	2.5	2.6	3.1	3.1	2.7	2.9	1.0
HRC 9	2.3	1.9	2.3	3.7	2.4	3.1	2.5	2.3	3.5	2.4	2.6	0.8
HRC 10	1.3	1.2	1.1	1.1	1.4	1.2	1.2	1.0	1.0	1.1	1.2	0.4
HRC 11	3.8	3.7	3.7	3.4	4.2	3.0	2.4	2.7	2.6	3.2	3.3	0.8
HRC 12	2.4	2.6	2.2	2.0	2.4	2.1	2.0	2.2	1.6	2.4	2.2	0.8
HRC 13	2.3	2.7	2.5	2.2	2.2	2.1	2.1	2.2	1.8	2.4	2.3	0.8
HRC 14	1.1	1.1	2.2	2.7	1.2	3.6	1.1	1.2	1.4	2.7	1.8	0.6
HRC 15	2.0	2.1	2.4	2.1	2.3	2.1	2.5	2.5	1.6	1.0	2.1	0.7
HRC 16	2.7	2.8	3.0	2.7	3.3	2.2	2.3	2.7	2.5	2.7	2.7	0.9

27 observers (14 males and 13 females), aged from 19 to 48 years old, viewed the 160 processed video sequences (PVS) in the experiment. First, five sequences were used as a training set, then all 160 sequences were shown in a semi-random order individually chosen for each observer with the restriction that the same source or the same HRC was never selected twice in a row. At the end of each sequence, a grey screen was displayed, and the observer was asked to evaluate the video quality with a score ranging from 1 (worst quality) to 5 (best quality). According to observers screening criteria from both ITU-BT.500 and VQEG Multimedia Test Plan, none of the observers was rejected. Table 3 shows the Mean Opinion Scores (MOS) for each PVS computed from the votes of the observers as well as the mean value and the standard deviation of each HRC.

2.2 Data format

The decoded video sequences are readily available in uncompressed AVI files. In order to facilitate the development of new objective video quality metrics, JEG proposes the use of an XML-based data exchange format as input to a hybrid metric. Using this approach, there is no need for writing a complete bitstream parser. All information can easily be extracted by simply parsing the XML file. As depicted in Figure 2, the Hybrid Model Input XML (HMIX) file contains all information extracted during the streaming and decoding of the video sequence.

The base for generating HMIX files lies in the availability of a network capture of the streamed video sequence by means of, for example, *tcpdump*. Different software tools, developed within JEG, are used to extract information from both the network and the video layer. All this information is then merged together into one HMIX file. Next to the generated HMIX file, also the PVS is made available to the hybrid objective quality metric. The interested reader is referred to [14] for more information on the different tools in the JEG processing chain.

All data files are available for download at http://www.irccyn.ec-nantes.fr/spip.php?article1033.

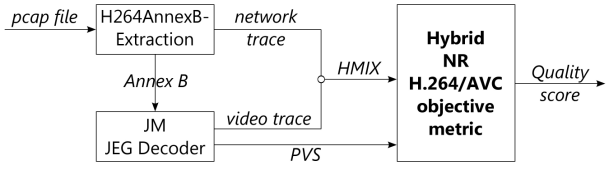


Figure 2 – Creation of the HMIX file containing information extracted during the streaming and decoding of the video.

3. Example evaluation

The following procedure is proposed to evaluate the performance of an indicator:

1. Identify the HRCs which are in the scope of the indicator
2. Perform a first order or a monotonous third order fitting, depending on the number of available data points
3. Report the performance of the indicator within its scope, for the extended scope and outside of its scope separately

As a simple example, the model that we described and implemented in [14] will be evaluated. The model contains only a single indicator as it estimates the Mean Opinion Score as a function of the encoded QP value. The algorithm is freely downloadable, was implemented in Python programming language, and takes an HMIX file as input.

The algorithm iterates over all the pictures in the video sequence and calculates the average QP value based on the QP Y values of each macroblock in that picture. QP_{mean}, the average QP computed across the entire sequence, is used to estimate the MOS from following equation:

$$\overline{MOS} = -0.172 \cdot QP_{mean} + 9.249$$

where $\overline{MOS}$ is the predicted Mean Opinion Score. The scope of the model is therefore limited to sequences coded with a constant QP. Only HRC2 and HRC3 fall into its designed scope. The mean value calculation for all QP implies that it may also provide a

rough estimation for HRC1, 4, 5, and 6 although it was not trained on this case, this will be identified as extended scope. The other HRCs are clearly out of scope for this model's indicator.

The fitting to the subjective data was therefore performed on HRC2 and HRC3, leading to

$$\overline{\text{fMOS}} = 1.314 \cdot \overline{\text{MOS}} - 1.246 .$$

Please note that for this very simple indicator, this fitting corresponds directly to a new fitting on the QP values. In Table 4 the Root Mean Square Error (RMSE), the epsilon-insensitive RMSE (RMSE*), the Pearson Linear Correlation Coefficient (PC) and the Spearman Rank Order Coefficient (SROCC) are provided.

Table 4 – Performance of an example model

Scope	RMSE	RMSE*	PC	SROCC
In scope	0.443	0.390	0.852	0.842
In extended	0.742	0.582	0.858	0.689
Out of scope	1.833	1.138	-0.380	-0.346

This analysis is only meant to illustrate the evaluation process that may be performed on a large video quality dataset. It cannot be used to validate the model's performance. The results indicate that the model may be useful for the conditions which are in the scope, and eventually for the extended scope with a different linear fitting as indicated by the contradiction between the high Pearson Correlation and the increased RMSE values. As expected, the out-of-scope results are not acceptable. A scatter plot is shown in Figure 3. A small random offset was added to each data point for visualization as the subjective and objective values are quantized. It is obvious that the indicator would often influence the results for the out-of-scope conditions which may reduce its value in a combined model. More accurate analysis can be easily performed as more subjective datasets become available, due to the usage of a common data format.

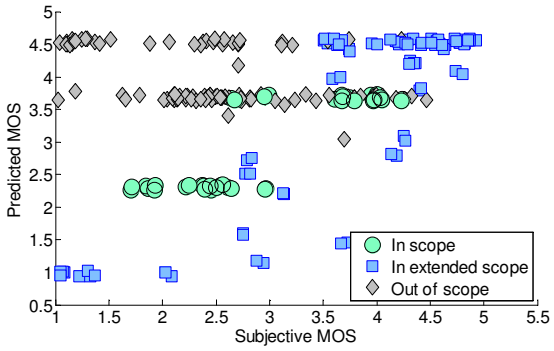


Figure 3 – Scatterplot relating the subjective MOS values to the predicted MOS values

4. Conclusions

The joint development of objective video quality measurement algorithms requires a combination of several indicators which should be evaluated in their appropriate scope. In this paper, a first subjective dataset is described towards the evaluation of such indicators and their combination in more complex models.

The subjective dataset is publicly available and contains easily accessible data, such as parsed bitstream data in XML files. This simplifies the adaptation of existing algorithms and provides a generic interface for the development of new hybrid algorithms.

For the performance evaluation of individual indicators or complete models, a method has been proposed that takes into account the designed scope of the model and its possible extensions.

The next steps will be to provide more video datasets that use the same data structure, to develop and compare indicators for individual degradations, and to combine these indicators into a reliable hybrid video quality measurement method.

5. ACKNOWLEDGMENTS

Mikołaj Leszczuk's work was financed by The National Centre for Research and Development (NCBiR) as part of project no. SP/I/1/77065/10.

6. REFERENCES

[1] Watson, A. B., Ahumada, J., Albert J. (2006). A standard model for foveal detection of spatial contrast. Journal of Vision, 5, 717–740.

[2] Huynh-Thu, Q., Ghanbari, M. (2006). Impact of Jitter and Jerkiness on Perceived Video Quality. VPQM

[3] Leszczuk, M., Romaniak, P., Głowacz, A., Derkacz, J., and Dziech, A. 2012. Large-scale works on algorithms of Quality of Experience (QoE). In *Proceedings of the fifth ACC Cyfronet AGH user's conference* KU KDM '12, Kraków, Poland, 33-34.

[4] Gillespie, W., and Nguyen, T. 2006. Classification of Video Sequences in MPEG Domain. *Mu. Sys. Appl.* 27, 1 (2006), 71-86.

[5] Kyungnam, K. and Davis, L. 2004. A fine-structure image/video quality measure using local statistics. In *Proceedings of the IEEE International Conference on Image Processing* (2004).

[6] Park, D. K., Won, C. S., and Park, S.-J. 2002. Efficient use of MPEG-7 Edge Histogram Descriptor. *ETRI. J.* 24, 2 (2002).

[7] Mu, M., Romaniak, P., Mauthe, A., Leszczuk, M., Janowski, L., and Cerqueira, E. 2011. Framework for the Integrated Video Quality Assessment. *Multimed. Tools. Appl.* (Nov. 2011), 1-31.

[8] Kusuma, T., and Zepernick. H.-J. 2005. Objective Hybrid Image Quality Metric for In-Service Quality Assessment. *Mu. Sys. Appl.* 27, 1 (2005).

[9] Janowski, L., Leszczuk, M., Papir, Z., and Romaniak, P. 2011. The Design of an Objective Metric and Construction of a Prototype System for Monitoring Perceived Quality (QoE) of Video Sequences. *Journal of Telecommunications and Information Technology*. 3 (2011), 87-94.

[10] Wolf, S., and Pinson, M. 2002. *Video Quality Measurement Techniques.* Technical Report. TR-02-392.

[11] Idrissi, N., Martinez, J. and Aboutajdine, D. 2005. Selecting a Discriminant Subset of Co-occurrence Matrix Features for Texture-Based Image Retrieval. In *Proceedings of the Advances in Visual Computing* (2005). Lecture Notes in Computer Science, 696-703.

[12] Pinson, M., & Speranza, F. (Eds.). (2010). Report on the Validation of Video Quality Models for High Definition Video Content. Video Quality Experts Group.

[13] INTEC Broadband Communication Networks, Homepage of Sirannon, Retrieved 2012, http://sirannon.atlantis.ugent.be

[14] N. Staelens, I. Sedano, M. Barkowsky, L. Janowski, K. Brunnström and P. Le Callet, "Standardized Toolchain and Model Development for Video Quality Assessment - The Mission of the Joint Effort Group in VQEG", Proceedings of the Third International Workshop on Quality of Multimedia Experience (QoMEX), September 2011

Instance Selection Techniques for Subjective Quality of Experience Evaluation

Yohann Pitrey, Werner Robitza, Helmut Hlavacs
Entertainment Computing Research Group, University of Vienna
firstname.lastname@univie.ac.at

ABSTRACT

This paper considers the problem of decreasing the number of video sequences to present in a subjective experiment so that the duration stays under the standard recommendations, while allowing to evaluate the perceived quality on a large set of configurations and contents. We present the constraints and related issues, then compare three selection algorithms through a series of subjective experiments. We analyse the correlation between the votes given to a set of videos in an initial large-scale test, and the votes given to three subsets of these videos in separate smaller experiments.

1. INTRODUCTION

Designing subjective experiments is a common and demanding task for every researcher in Quality of Experience. These experiments are the core of our activity as they allow us to sample the impact of different parameters and phenomenons on the visual quality, as perceived by human observers. A classical approach is to design a set of treatments, also refered to as Hypothetical Reference Circuits (HRC), expected to cover the whole phenomenon under study. These HRCs are used to process a set of source video sequences (SRC), in order to produce Processed Video Sequences (PVS). These PVSs are presented to a group of human observers, following a test methodology such as presented in the ITU-T P.910 recommendation [1]. This recommendation also specifies the maximum duration for one test session with one observer. Providing a subjective judgement on a set of video sequences is indeed quite demanding for observers, and the duration of a session should typically be kept under 25 to 30 minutes to avoid visual and mental fatigue. As a consequence, the acceptable number of PVSs to be presented in one test session is itself limited. This number depends mainly on two factors: the length of one PVS and the rating methodology employed for the experiment. Pair comparison methodologies, such as the Double-Stimulus Continuous Quality-Scale (DSCQS), allow a relatively low number of PVSs to be presented, as the observers have to give their opinion on each PVS twice. Single stimulus methodologies, such as the Absolute Category Rating (ACR), allow more PVSs to be presented, as the observers evaluate each PVS only once.

In a classical experimental design, all possible combinations between HRC and SRC are presented in the experiment. In order to remain under the maximum duration for a test session, the experimental designers usually have to reduce the number of HRCs, or the number of SRCs. As a result, it can be difficult to cover the whole range of parameters expected to have an impact on the perceived quality in the context under study. An alternative is to split the subjective experiment in several sessions for each observer, but this approach causes a significant increase in time to conduct the experiment and in cost to pay the observers. The experimental process is also made more complex, as a common set of configurations has to be included in all the sessions, so that the results from each individual session can be compared [5, 7].

Another solution for meeting the maximum duration in a subjective experiment without severely decreasing the number of HRCs or SRCs, is to select only some combinations out of the complete set of PVSs. However, some stringent constraints have to be addressed in order to construct a reliable selection. In this paper, we consider the problem of PVS selection for video QoE assessment experiments. First, we analyse the constraints and difficulties to address in order to solve this problem. Then, we oppose manual selection to automated selection and illustrate how both approaches can be used in practice. As automatic selection allows significant savings in terms of time and effort from the experimental designers, we further compare three common algorithms originally used by the data mining community for instance selection. This comparison relies on a large scale standardized subjective experiment which was conducted in four sessions per observers, such as presented in [6]. Our demonstration aims at constructing three subsets from the original experiment using the instance selection algorithms, so that each can be conducted in one session per observer. We analyse the reliability of the obtained votes, and discuss the advantages and drawbacks of each method.

2. CONSTRAINTS AND DIFFICULTIES

Experimental design is a complex task in which different types of constraints need to be respected, so that the provided evaluation is both reliable and easily reproducible. Several contributions specify these constraints, such as the ITU-T P.910 recommendation or the description VQEG Mutlimedia Test Plan [5]. One of the main constraints is that the subjective experiment must contain a wide and balanced range of quality levels. This constraint is particularly justi-

fied by a phenomenon called the context effect, which implies that human observers rate a given PVS relatively to all the other PVSs included in the experiment [2]. One common motivation for designing and conducting subjective experiments is to create perceptual quality metrics, which are algorithms capable of predicting the quality score given by a standard human viewer to a video sequence, based on environmental parameters [4]. The constraint on the distribution of quality scores is important for the design of such metrics, in order to train the models on a representative set of videos. Full reference quality metrics are a widely used type of algorithms which exploit the non-degraded video sequence as a basis for their estimation of quality. Therefore, including the non-degraded high quality reference version of each source video is another important constraint in the design of an experiment, regarding their further use for metric design.

The problem we address in this paper can be presented as selecting a given number of PVSs among a larger set that would require more than one session per observer to be fully evaluated. Such a situation is quite common when evaluating the impact of a particular phenomenon on the perception of quality by human viewers. A crucial aspect of the selection process is to ensure that the selected set of PVSs is representative of the original set. Indeed, meeting this constraint represents a major step towards extrapolating the results obtained on the selected set to the original set. To address this issue, each PVS needs to be described using parameters reflecting the treatments and the characteristics of the source video that were combined to generate it. Then, one goal of the selection process can be to conserve the original distribution of the parameter values in the selected set, while decreasing the number of PVSs. However, finding reliable descriptors for the PVSs can be a difficult task in itself. Such as mentioned in [3], using too many or unadapted descriptors can lead to misclassification in the context of data mining. While our problem is not directly related to data mining, one can easily transpose the mentioned problems to PVS selection.

Instance selection can also be used to design a balanced set in terms of quality. One of the main issues to address in this case is how to estimate the quality of PVSs before conducting the subjective test itself. In some cases, the experimenter has some prior knowledge or intuition about the impact of the different parameters, which can be used to estimate the level of quality of each PVS and make sure that the selection is balanced. However, taking manually into account all the parameters expected to have an influence on the quality might become very complex. An alternative can be using an adapted objective quality metric to estimate the quality of the PVS before the test. Naturally, this raises questions about the reliability of the measure. Some contexts, such as video coding distortions and scaling artifacts benefit from efficient metrics, however other domains are still missing accurate algorithms, such as illustrated in [8]. Nevertheless, adding a constraint on the distribution of quality in the selected set can bring a substantial contribution in the experimental design and deserves to be studied further.

It is well accepted that quality can be affected both by single parameters and by their combination. Such an example can be found in [8] in the context of lossy video transmission, where it is demonstrated that Quantization Parameter (QP) and packet loss distribution have a joint impact on the perceived quality. This experiment, which was conducted in four sessions per observer, combines specific patterns in the HRCs so that this joint impact can be evaluated. One danger of instance selection with such an experiment is that these configurations might not be selected, making it more difficult to evaluate the joint impact of parameters.

3. MATERIAL

To illustrate different instance selection techniques, we will rely on a subjective experiment which focuses on the impact of Scalabe Video Coding distortions under various combinations of QP values. More details on this experiment and the subjective results can be found in [6, 8]. This experiment was conducted in four sessions per observer, following a standard ACR procedure with a 5-level quality scale. The scalable streams contained one base layer in QVGA format ($320*240$ pixels) and one enhancement layer in VGA format ($640*480$ pixels). The QP value for each layer was set to 26, 32, 38 and 44, with a total of 16 scenarios. Four additional AVC scenarios upscaled from QVGA to VGA were included, as well as four AVC scenarios in native VGA resolution. A total of 25 scenarios including the non-degraded reference were therefore included in the test.

In the current paper, the goal of our demonstration is to decrease the number of required sessions to one per observer, by selecting subsets of PVSs using different algorithms. As mentioned in the previous section, a preliminary step to instance selection is to describe the data using HRC and SRC indicators. In the context we consider, the HRCs can be simply described using the QP values of both layers. In case an HRC does not contain two layers (*i.e.* AVC HRCs) and for the reference HRC, we use a fictive QP value set to 0. Regarding the SRC indicators, we use the descriptors inspired by the VQEG Multimedia Test Plan [5]: Spatial and Temporal Indicators (SI and TI), coding complexity, content type, number of scene cuts and shot type. The coding complexity is simply evaluated by encoding all the original sources with a common fixed QP value, and calculating the bitrate per second. The content type is set manually and comprises *news*, *sport*, *movie* and *nature*. The shot type is also set manually and comprises *short distance*, *medium distance* and *wide angle*.

4. INSTANCE SELECTION TECHNIQUES

In this paper, we mainly focus on automatic instance selection techniques. As mentioned in the previous section, manual selection can be very complex and requires prior knowledge about the studied phenomenon. However, in some cases such knowledge is available, and manual selection can be considered. In [6], we presented the results of a subjective experiment aimed at evaluating the influence of the source content under various SVC encoding conditions. This experiment is an extension of the experiment studied in this paper, in which a total of 21 HRCs varying the QP values for both layers are combined with 60 SRCs including a wide variety of content types. As a result, the whole data set would contain no less than 1260 PVS, which represents about 12 sessions per observer. As this scenario is obviously not acceptable in practice, we applied a semi-manual selection process to decrease the number of required sessions to four per observer. A quick calculation allowed us to determine that we could afford 5 HRCs per source content, for the total amount of PVSs to fit in the four sessions. We used the results from

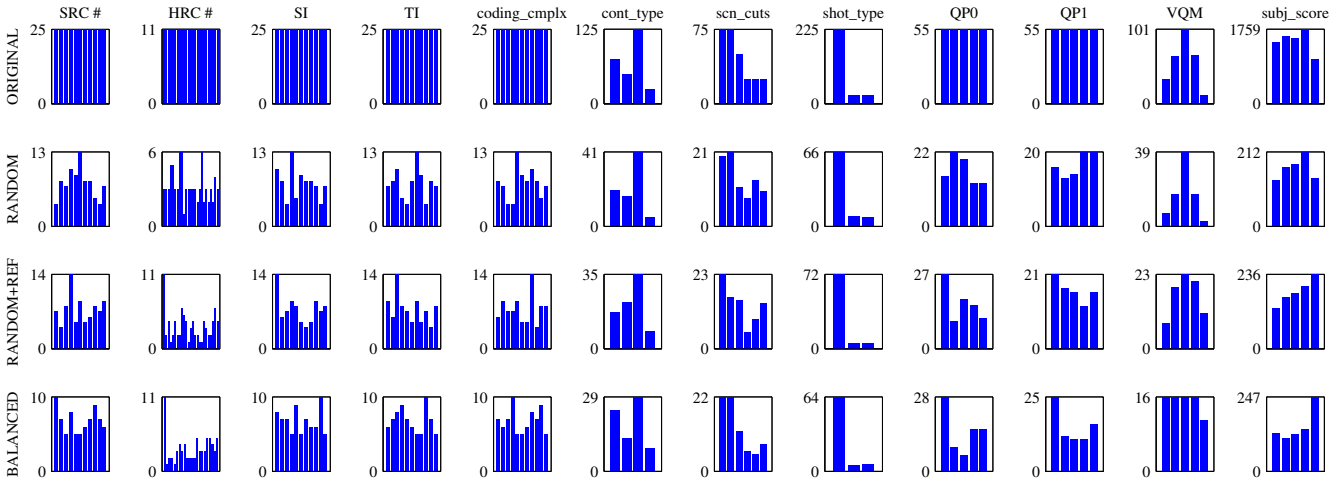


Figure 1: Parameter values distribution comparison.

the first experiment to estimate the MOS of the 21 HRCs, and classified them into five groups of comparable size, each group containing HRCs with similar quality. Then, for each SRC, one HRC from each group was randomly selected. One group contained only the non-degraded reference HRC, in order to meet the contraint mentioned in the previous section. The results of this experiment show that the quality is uniformly distributed on the quality scale, which proves the accuracy of the selection process. Nevertheless, this example is quite specific and such prior knowledge is not always available.

Here, we compare three instance selection algorithms available in the WEKA 3 data mining tool[1]. They represent different levels of control on the selection process, aiming either at conserving the original parameter distribution or designing a balanced test in terms of quality. To set the objective of the selection process, we first calculated the number of PVSs fitting in one test session. Each PVS is 10-second long, then the observers have to give their opinion of quality on a 5-levels scale, which we estimated takes about 5 seconds. Between two PVSs, a waiting time of 3 seconds is included. A test session should last at most 25 minutes, or 1500 seconds. As a result, $1500/18 \simeq 83$ PVSs can be included in one test session. The three considered algorithms are described in the remainder of this section.

Random Selection : The first approach uses a simple random selection with uniform distribution. In the WEKA framework, this method is called "Resample" and is available as a preprocessing filter after loading a data file. This approach is quite simplistic, its main drawback being that even the reference PVSs can be left out of the selection. This scenario is aimed at showing a relatively intuitive yet not-suitable approach for instance selection.

Random Selection with Reference : To overcome the drawback of the first approach, we manually pre-select all the reference PVSs and move them to the selection before running the random selection. As a result, only 72 PVSs are randomly selected from the remaining set. This approach slightly modifies the original parameter distribution but follows more closely the standard design rules and is suitable for full-reference metric extraction.

Balanced Design : The third method aims at selecting instances so that the overall distribution in terms of quality in the selected set is as uniform as possible. This method requires an estimation of quality for each PVS. To this end, we used the full reference Video Quality Model developed by Pinson and Wolf [4], which is known to give accurate predictions of quality for videos under coding and scaling distortions. For each PVS, the VQM scores were calculated and mapped to 5 categories using simple linear mapping and rounding operations. Each PVS was associated with an integer number comprised between 1 and 5, which was used as the class value for the selection process. In WEKA, we used the "Spread Subsample" method, which tries to minimize the difference between the least and most frequent classes after selection. The reference PVSs are forced into the selection in this approach as well.

Using the presented selection schemes, we designed three subsets of 83 PVSs each, which were used to conduct three subjective experiments. In the next section, we compare the results of these experiments with the results obtained on the original experiment.

5. EXPERIMENTAL RESULTS

Three subjective experiments were conducted based on the three selected sets of PVSs. The experimental conditions were identical to the conditions for the original experiment in terms of video presentation and voting protocol. The room conditions were close to the original experiment, although not conducted in the same laboratory. The most significant difference with the original experiment in terms of test environment was the use of a consumer-range conputer screen as display. A total of 33 observers were tested (11 observers per experiment). The three groups of observers were equivalent in terms of age and gender distribution (average: 23.4, min: 18, max: 42). Their visual acuity was tested using a standard Snellen chart and Ishihara plates.

We first analyse the three selected sets of PVSs by comparing their parameter distributions to the original test, such as presented by the histograms in Figure 1. The SRC and HRC numbers were not taken into account during the selec-

[1] http://www.cs.waikato.ac.nz/ml/weka/

	Random	Random+Ref	Balanced
SRC MOS	0.943	0.919	0.986
HRC MOS	0.988	0.986	0.986

Table 1: Linear correlation coefficients for SRC and HRC MOS, for the three selection algorithms.

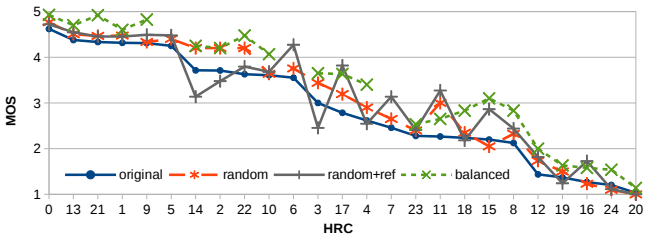


Figure 2: HRC MOS comparison for the four subjective experiments.

tion process, and were added afterwards to ease the reading. We can already observe some differences between the VQM values and the measured MOS values in the original test. The simplistic classification performed on the VQM values is not optimal, as it does not reflect the MOS repartition. This illustrates the dangers of relying only on an objective measure to estimate the quality of the PVSs before conducting the test.

Random selection follows the original distribution closer than the two other subsets. As no control is set on the selection process in this approach, randomly selecting PVSs overall produces a set that follows the same distribution as the original set. As previously mentioned, a main drawback of this approach is the absence of reference PVS for some SRCs. This issue is solved in the second selection process, in which the reference PVSs are added before runing the random selection on the remaining PVSs. As a result, the second selected set contains more high quality PVSs. One of the main drawbacks of this approach is that in case of an unbalanced original set of PVSs, the provided selection would also be unbalanced. The third selection approach grants more control on the distribution of the selected PVSs. As can be seen in Figure 1, the VQM scores in the selection are almost equally distributed among the five classes. The class with higher VQM scores only contains reference PVSs, which represents in total only 11 PVSs. This explains why this class is slightly less populated than the others. As a consequence of the modification of the VQM distribution in the selected set, the distributions of the other parameters are also slightly modified, as can for instance be observed for the number of scene cuts. However the overall tendency is followed quite closely. This approach allows the design of a balanced set of PVSs, which is a significant contribution for the experiment designer's work. However, it requires a reliable quality estimation of the PVSs.

Table 1 presents the Pearson linear correlation coefficients between the MOS from the original set and the three selected sets. Both the HRC MOS and the SRC MOS are presented, to illustrate how each condition and each content is rated. All three approaches reach very high correlation, which suggests that the overall tendency from the original set is respected in the MOS obtained on the selected sets. In order to get more insight regarding this statement, Figure 2 compares the HRC MOS values for the four sets. One can observe that the global tendency from the original test is indeed respected in the selected sets. However, some significant differences can be identified for some PVSs. For instance, with the random selection with reference, HRCs 14, 6, 17 and 11 obtain quite different MOSs from the original set. For each of these HRCs, only one or two PVSs were included in the selected set. The same observation holds for HRCs 22, 15 and 8 for the balanced approach. This illustrates the importance of repeating the evaluation of a HRC on a significant number of SRCs in order to obtain a reliable MOS.

6. DISCUSSION AND CONCLUSION

One of the main conclusions of this work is that automatically decreasing the number of PVSs in such a significant way (here we selected only 25% of the initial amount of PVSs) comes at a price in terms of consistency of the results, mainly due to the lack of control on the selection process. Substantial improvements can be expected in the development of more complex selection algorithms. Particularly, we showed that each HRC should be evaluated on a substantial number of PVSs in order for the observed MOS to match with the original MOS. A constraint to maximize the number of occurrences of a HRC could be added to the selection process to solve this issue. The selection could then be seen as finding a tradeoff between removing entrie HRCs from the experimental design, and applying each HRC only on a subset of the available SRCs. This approach would require treating the SRC and HRC indicators separately, conducting two parallel optimization tasks during the selection.

Additionally, a notion of distance between PVSs could be introduced, based on their similarity in terms of descriptors. An optimization criterion would then replace the random selection, based on the relative distance from one PVS to its neighbours.

7. REFERENCES

[1] *Subjective video quality assessment methods for multimedia applications*. ITU-T Recommendation P.910, 2008.

[2] M. Garcia and A. Raake. Normalization of subjective video test results using a reference test and anchor conditions for efficient model development. In *QoMEX*. IEEE, 2010.

[3] P. Gastaldo and J. Redi. Machine Learning Solutions for Objective Visual Quality Assessment. In *VPQM*, 2012.

[4] ITU-R. Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference. *ITU-R BT.1683*, 2004.

[5] M. Pinson and S. Wolf. *Techniques for evaluating objective video quality models using overlapping subjective data sets.* US Dept. of Commerce, Nat. Telecom. and Information Admin., 2008.

[6] Y. Pitrey, M. Barkowsky, R. Pépion, P. Le Callet, and H. Hlavacs. Influence of the source content and encoding configuration on the preceived quality for scalable video coding. In *IS&T/SPIE Electronic Imaging*, 2012.

[7] Y. Pitrey, U. Engelke, M. Barkowsky, R. Pépion, and P. Le Callet. Aligning subjective tests using a low cost common set. In *QoEMCS*, 2011.

[8] Y. Pitrey, R. Pépion, P. Le Callet, and M. Barkowsky. Using Overlapping Subjective Datasets To Assess The Performance Of Objective Quality Metrics On SVC And Error Concealment. In *Accepted for QoMEX*, 2011.

Lessons Learned during Real-life QoE Assessment

Nicolas Staelens
Ghent University - IBBT
G. Crommenlaan 8 bus 201
Ghent, Belgium
nicolas.staelens@ugent.be

Wendy Van den Broeck
VUB - IBBT
Pleinlaan 9
Brussels, Belgium
wvdbroec@vub.ac.be

Yohann Pitrey
University of Vienna
Dr.-Karl-Lueger-Ring 1
Vienna, Austria
yohann.pitrey@univie.ac.at

Brecht Vermeulen
Ghent University - IBBT
G. Crommenlaan 8 bus 201
Ghent, Belgium
brecht.vermeulen@ugent.be

Piet Demeester
Ghent University - IBBT
G. Crommenlaan 8 bus 201
Ghent, Belgium
piet.demeester@ugent.be

ABSTRACT

Subjective video quality experiments are usually conducted inside controlled lab environments, where stringent demands are imposed in terms of lighting conditions, screen calibration and position of the test subjects. However, these controlled lab environments might differ from a more natural setting such as watching television at home. This, in turn, will influence subjects' Quality of Experience. In previous research, we conducted a series of subjective experiments, mimicking the natural environment of the targeted use cases as much as possible. In this paper, we provide an overview of these conducted studies and summarize our main research findings. Our results show that the environment and overall experimental setup influence the primary focus of the test subjects. Consequently, a significant difference in impairment visibility, tolerance and annoyance is observed compared to conducting experiments in pristine lab environments. Our results highlight the importance of considering the targeted use case and assessing video quality in more natural environments.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human factors

General Terms

Measurement, Human factors

Keywords

Quality of Experience (QoE), Subjective video quality assessment, Real-life

1. INTRODUCTION

According to the definition of Quality of Experience (QoE) [3], *"the overall acceptability of an application or service, as perceived subjectively by the end-user"*, might be influenced by user expectations and context. Subjective video quality assessment methodologies as specified in, for example, International Telecommunication Union (ITU)-R Recommendation BT.500 or ITU-T Recommendation P.910 describe in detail how to set up and conduct experiments for obtaining real human ratings on perceived quality of (degraded) video sequences. These experiments must be conducted in a controlled lab environment subject to stringent demands in terms of, amongst other, lighting conditions and distance between the test subjects and the screen. Furthermore, observers participating with the experiment usually receive specific instructions on how to assess and evaluate the quality of the different video sequences. As such, users' expectations and users' context might already be influenced prior to the start of the experiment and hence will impact their QoE.

Due to the instructions given to the test subjects, observers' primary focus is on (audio)visual quality evaluation. Also, most of the existing subjective assessment methodologies focus on the evaluation of short video sequences (typically between 10 and 15 seconds long)[1]. Moreover, the standardized test environment in which the experiments must be conducted is not necessarily representative for a more realistic setting such as watching television in a living room environment [8].

In previous research [7, 8], we conducted a series of subjective video quality experiments in more realistic environments targeting specific use case scenarios. Consequently, we considerably changed the context and expectations of our observers compared to those during more standardized quality assessment. The results of these studies were also compared with results obtained during quality assessment in controlled lab environments as specified by the ITU Recommendations.

In this paper, we briefly present our research on assessing video quality in more realistic environments. By summarizing the main research findings of our real-life experiments, we also discuss the benefits and importance of mimicking realistic environments/settings during subjective quality assessment. Last, we identify some of the pitfalls encountered

[1]Except for the Single Stimulus Continuous Quality Evaluation (SSCQE) methodology as specified in ITU-R Rec. BT.500.

during real-life video quality assessment.
The remainder of this paper is structured as follows. In section 2, we present some related work on experiments conducted in less stringent assessment environments and methodologies enabling more natural quality assessment.
Next, section 3 provides more information on two subjective studies we performed to assess the influence of primary focus on perceived quality when conducting experiments in a more natural setting. Section 4 summarizes the main research findings of these experiments and highlights the importance of mimicking realistic environments during quality assessment. Finally, we conclude this paper.

2. RELAXING THE STANDARDIZED ASSESSMENT METHODOLOGIES

As pointed out in the introduction, several stringent demands are imposed on the controlled lab environment in which the subjective experiments should be conducted. On the one hand, this can make subjective experiments hard to set up and expensive. However, on the other hand, the standardized assessment methodologies facilitate repeatability of experiments.
Recently, a subjective experiment has been conducted to assess the influence of subjects' country and native language, playback/display device and overall test room conditions on audiovisual quality perception [6]. The audiovisual experiment was conducted several times in different environments and labs located in different countries. Furthermore, the experiments were conducted inside a controlled laboratory and inside a public area (e.g. cafetaria). Prior to the start of the experiment, subjects still received specific instructions and were thus focused on audiovisual quality evaluation. The results of this study showed that the environment in which the experiment is conducted does not greatly influence quality perception nor quality ratings. However, in the case of quality evaluation in a public area, a higher number of subjects is required for gathering stable results. This is a first indication that the stringent demands posed on the assessment environments can be relaxed to some extent.
The overall experiment duration and length of the video sequences are also two limiting factors of existing subjective quality assessment methodologies. In order to avoid viewer fatigue, experiment duration should not exceed 30 minutes. In [1], the authors propose a novel subjective methodology enabling quality evaluation of long duration audiovisual sequences. Instead of providing an actual quality rating in case of a degradation, subjects are asked to tune the quality of the sequence, by means of a knob, to the desired level. Based on feedback received from the test subjects, the proposed methodology requires less attention from the participants which allows them to focus more on the actual content.
A test bed for augmenting user experience by simulating sensory effects is presented in [10]. Using this test bed, the influence of, for example, wind, light or any other sensory effect on QoE can be evaluated. This also enables to simulate more natural environment conditions during quality assessment.
Research has also been conducted towards replacing the traditional quality rating scales by other scoring techniques such as providing feedback using a glove [2] or a gaming steering wheel [4]. In [5], a comparison was also made between different devices (mouse, joystick, throttle and sliding bar) for providing feedback in the case of time-variant video quality.
As such, active research is being conducted on relaxing the stringent demands imposed on the assessment environment and on creating more realistic environments during quality assessment.

3. MIMICKING REAL-LIFE CONDITIONS DURING SUBJECTIVE QUALITY EVALUATION

As user expectations and context influence QoE, we wanted to investigate the impact of conducting experiments in more realistic/natural environments. In particular, we are interested in discovering the influence of quality degradations when test subjects are not primarily focused on active (audio)visual quality evaluation.
In a first study [8], we worked out a novel subjective quality assessment methodology enabling quality evaluation in real-life environments. Our proposed methodology is based on injecting impairments in full length DVD movies. Next, test subjects were asked to take the DVD home and watch it in their most natural environment. In order to collect feedback concerning the perceived quality and the detection of degradations during playback, a questionnaire was provided to the test subjects in a sealed envelope containing, amongst other, the following questions:

- Did you perceive any kind of visual degradation during movie playback? If yes, which kind?
- Describe the scenes in which these degradations occurred.
- Indicate, on a scale from 1 to 5, impairment annoyance.
- Rate, on a scale from 1 to 5, the overall visual quality of the movie.

The participants were not instructed about the possible occurrence of visual degradations and they were asked not to open the envelope before having watched the entire movie. This way, we encouraged subjects to watch the DVD for its content without actively evaluating perceived quality. An additional questionnaire was used to obtain information on the environment in which they watched the DVD.
In a second experiment [7], we recently assessed the influence of audio/video (A/V) synchronization issues in the case of simultaneous translation of video sequences. Therefore, we mimicked the typical environment used by expert interpreters as shown in Figure 1. In this case, the experiment



Figure 1: Environmental setup used for assessing A/V sync issues during our subjective quality assessment experiment.

was conducted using both real interpreters (expert users) and non-experts. The main difference between the evaluations performed by the expert and the non-experts is the fact that the experts users were specifically instructed to

perform a simultaneous translation of the video sequences (as they would do in a real-life scenario). By doing so, the interpreters were primarily focused on the translating part instead of detecting A/V issues.

4. LESSONS LEARNED WHILE MIMICKING REAL-LIFE CONDITIONS

By analysing the results obtained during the two experiments described in the previous section, we found some significant differences compared to results gathered using a standardized quality assessment methodology.

4.1 Influence of assessment environment and task on primary focus

Using our two subjective experiments, we changed the typical assessment environment used during subjective video quality assessment and tried to influence the focus of our test subjects by giving different instructions prior to the start of the experiment.

Concerning the experiment with the full length DVD movies, subjects were not aware of the fact that quality degradations could occur during video playback. Furthermore, the test subjects watched the complete movie in their preferred environment. Also, no impairments were injected in the first or last 15 minutes of the movie. Considering all these factors, participants were stimulated to watch the DVD primarily for its content. Based on the feedback collected from the subjects, we found that our novel methodology based on full length movies is capable of mimicking the typical lean-backward TV-viewing experience. As opposed to the standardized assessment methodologies using short video sequences, we are now also able to increase the immersion of our test subjects.

In our second experiment on assessing the influence of A/V synchronization issues, we explicitly instructed our expert users to perform a simultaneous translation of the audiovisual sequences. On top of that, the interpreters were aware of the fact the synchronization problems could occur during playback. After each sequence, the experts were asked to indicate whether they observed any synchronization issue and how they would rate its annoyance. The results clearly showed that performing a translation of the video sequences significantly changes the ability of the subjects to detect quality degradations. The non-experts participating with the experiment were only instructed to detect the A/V synchronization problems.

As such, by mimicking more realistic conditions during subjective video quality evaluation or by giving slightly different instructions to the test subjects, we are able to significantly change the primary focus of our observers. In turn, this greatly influences their judgements concerning impairment visibility and tolerance as will be explained in the next section.

4.2 Influence of primary focus on impairment visibility and tolerance

During both experiments, we changed the context of quality assessment and, hence, influenced QoE. Afterwards, we also compared these results with results obtained using a standard subjective quality assessment methodology. Due to the change in primary focus of our observers, we see some significant changes concerning impairment visibility and tolerance when mimicking more realistic environments.

While the test subjects watched the DVDs, their focus was mainly on the actual content of the movie. This greatly influenced tolerance towards certain types impairments. As depicted in Figure 2, frame freeze impairments are significantly less detected during real-life QoE assessment. However,

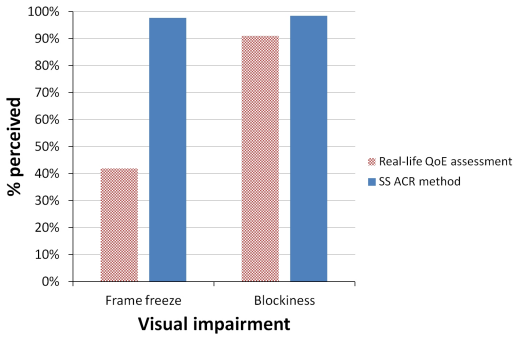


Figure 2: Impairment visibility of frame freezes and blockiness while conducting experiments in real-life and controlled lab environments.

when questioning the observers towards the tolerance of frame freezes or blockiness impairments, the majority of the subjects do not tolerate frame freezes. This is caused by the fact that freezes break the natural flow of the movie and have a greater impact on the immersive experience. As stated before, this immersive experience cannot be achieved using a standardized subjective methodology.

In the case of detecting A/V synchronization problems, we see that the expert users detect these sync issues much less compared to the non-experts as depicted in Figure 3. In general, the interpreters do not detect lipsync problems except in the case delay between audio and video increases up to -240ms[2]. As such, due to a change of primary focus going

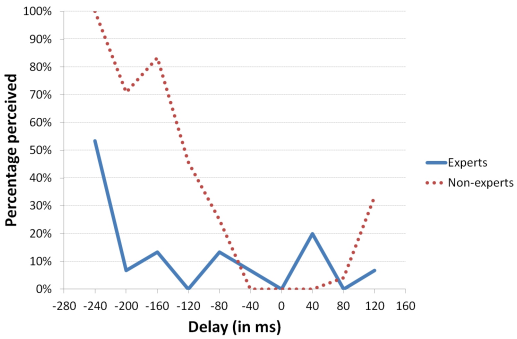


Figure 3: Percentage of the experts and non-experts who detected the A/V synchronization delay.

from active video quality evaluation to simultaneous translation of video content, observers' attention is taken away from the quality aspect of the content.

These results indicate that the specific task, immersion and

[2]A negative offset implies that audio is delayed with respect to the video.

primary focus significantly impact quality perception and detection.

4.3 Points of special interest to consider

By running subjective experiments in more natural environments/settings, we are able to mimic realistic conditions which influence subjects' QoE. These studies revealed some important differences compared to performing quality assessment inside a controlled lab environment. However, conducting experiments by mimicking natural environments requires some points of special attention.

From nature, subjective quality experiments are known to be time-consuming and expensive. In our case, we also conducted face-to-face interviews after the experiment with each of the participants. This aids in contextualizing the experiments and can provide more insights on how quality is assessed in natural environments. Unfortunately, these interviews complicate and further increases the time needed for conducting such experiments.

Another point of special interest to consider is the (highly) uncontrolled environment. During real-life QoE assessment, a lot of environmental factors such as lighting conditions, screen contrast ratio, distance to the screen, etc. cannot be controlled and can differ significantly from one subject to another. Therefore, special attention should be given to collecting as much parameters as possible characterizing the environment by means of, for example, an additional questionnaire. Further research is needed to assess the influence of different environmental conditions on quality perception.

In case of our full length DVD methodology, subjects were only questioned after having watched the entire movie about perceived quality and the number of detected visual degradations. As such, some bias can be introduced. For example, it can happen that certain impairments are forgotten, due to recency or primacy effects. Therefore, other means might be investigated for providing instantaneous feedback on perceived quality. However, this immediate feedback should not change the focus from movie content to active quality evaluation.

5. CONCLUSIONS

In this paper, we presented two subjective studies conducted to assess the importance of mimicking more natural environments during subjective video quality evaluation. By summarizing the main research findings of these experiments, we highlighted the importance of considering the targeted use case when conducting subjective experiments as this significantly influences subjects' expectations and context.

Our results show significant differences concerning impairment visibility, tolerance and annoyance compared to results obtained using one of the standardized quality assessment methodologies. This shows that results gathered during quality assessment in a controlled lab environment can be relaxed when measuring quality in more realistic environments. This calls for additional research on mapping controlled lab results to more real-life scenarios.

6. ACKNOWLEDGMENTS

Part of the research activities described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT) and the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). This paper is the result of research carried out as part of the IBBT projects *Video Q-Sac* (consortium of industrial partners Alcatel-Lucent, Telindus, Televic and Fifthplay in cooperation with the IBBT research groups: IBCN & MultimediaLab (UGent), SMIT (VUB) and IMEC) and *OMUS* (consortium of the industrial partners Thomson, Televic, Streamovations, Excentris in cooperation with the IBBT research groups: IBCN, MultimediaLab & WiCa (UGent), SMIT (VUB), COSIC (KULeuven) and PATS (UA)).

7. REFERENCES

[1] A. Borowiak, U. Reiter, and U. P. Svensson. Quality evaluation of long duration audiovisual content. In *IEEE Consumer Communications and Networking Conference (CCNC)*, pages 337 –341, January 2012.

[2] S. Buchinger, W. Robitza, M. Nezveda, M. Sack, P. Hummelbrunner, and H. Hlavacs. Slider or glove? Proposing an alternative quality rating methodology. *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, January 2010.

[3] ITU-T Rec. P.10/G.100 Amd 2. Vocabulary for performance and quality of service, 2008.

[4] T. Liu, G. Cash, N. Narvekar, and J. Bloom. Continuous mobile video subjective quality assessment using gaming steering wheel. *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, January 2012.

[5] O. Nemethova, M. Ries, A. Dantcheva, and S. Fikar. Test equipment of time-variant subjective perceptual video quality in mobile terminals. In *International Conference on Human Computer Interaction (HCI)*, 2005.

[6] M. H. Pinson. The influence of environment on audiovisual subjective tests. VQEG_MM2_2011_044_audiovisual lab comparison, Hillsboro, US, December 2011.

[7] N. Staelens, J. De Meulenaere, L. Bleumers, G. Van Wallendael, J. De Cock, K. Geeraert, N. Vercammen, W. Van den Broeck, B. Vermeulen, R. Van de Walle, and P. Demeester. Assessing the Importance of Audio/Video Synchronization for Simultaneous Translation of Video Sequences. *Springer Multimedia Systems*. Accepted for publication.

[8] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën and, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester. Assessing quality of experience of IPTV and Video on Demand services in real-life environments. *IEEE Transactions on Broadcasting*, 56(4):458–466, December 2010.

[9] N. Staelens, B. Vermeulen, S. Moens, J.-F. Macq, P. Lambert, R. Van de Walle, and P. Demeester. Assessing the influence of packet loss and frame freezes on the perceptual quality of full length movies. *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-09)*, January 2009.

[10] M. Waltl, C. Timmerer, and H. Hellwagner. A test-bed for quality of multimedia experience evaluation of sensory effects. In *International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 145 –150, July 2009.

Measuring the Quality of Long Duration AV Content – Analysis of Test Subject / Time Interval Dependencies

Adam Borowiak
Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology
Trondheim, Norway
adam.borowiak@q2s.ntnu.no

Ulrich Reiter
Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology
Trondheim, Norway
reiter@q2s.ntnu.no

Oliver Tomic
Nofima
Ås, Norway
oliver.tomic@nofima.no

ABSTRACT
This paper is an extension of our previous work describing a novel methodology for quality assessment of long duration audiovisual content. In this article we concentrate on results obtained from an experiment conducted using the new methodology. A mixed-effects ANOVA model was used to explore dependencies between subjects and particular time intervals. We found that the time dimension does not influence participants' expectations with respect to perceived quality and that a possible increase or decrease in acceptable quality level is rather directly related to the presented material itself. Moreover, we found that participants are less sensitive to quality changes when the process is controlled externally than when they are in charge of the quality adjustment.

Categories and Subject Descriptors
I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement – feature representation.

General Terms
Measurement, Experimentation, Human Factors.

Keywords
Subjective audiovisual quality assessment, data analysis, quality of experience (QoE).

1. INTRODUCTION
Recently we have proposed a novel methodology designed for quality evaluation of audio, video and audiovisual material of long duration. The method represents a novel approach towards evaluation of quality of experience (QoE) in mono and multimodal systems.

Initial results obtained with this method, including the detailed description of the experimental design and test procedures, have been described previously in [1]. In this paper we want to focus on the data analysis and interpretation of the results as well as discuss what additional insights into the quality perception this method can bring.

The remainder of this paper is organized as follows: in section 2, a brief description of the novel method is presented. In section 3, the experimental setup and procedures are discussed. Section 4 consists of a detailed description of the results and method for the data interpretation. Finally, conclusions are drawn and future work is discussed.

2. METHOD DESCRIPTION
Instead of giving numerical scores on predefined scales, the method proposed in [1] allows participants to adjust the quality to a desired level in case degradation occurs. As this is a purely perceptually based judgment, there is no need for assessors to translate their impressions into a numerical value or semantic designator, thus avoiding most of the problems typical for the MOS related approach [2]. To select the desired quality level a rotary controller device is used (e.g. knob, scroll wheel, etc.). Only by appropriate adjustments, based on perceptual judgment, the optimal (highest) quality level can be achieved. During the assessment task, automatic changes in the quality are introduced by the system periodically and stepwise (e.g., the degradation process starts every three minutes and then quality levels decrease every 10 seconds). Assessor's input stops the automatic procedure of quality degradation and provides him/her with full control over the quality adjustment. Rotating the knob clockwise increases the quality level of the presented stimulus up to the point where highest quality is achieved. Turning the device further clockwise causes a gradual decrease in the quality. This is a sort of penalty introduced when the maximum quality level is crossed. The process is reversible and by rotating the knob in the opposite direction the assessor can return to the highest quality level. The quality levels are prepared beforehand, and each of them has assigned a numerical value (e.g. lowest quality - 0, highest - 11). Acquisition of the user's responses is performed automatically, at a time resolution of one millisecond and with values which lie within the range of the corresponding quality levels. A more detailed description of the method can be found in [1].

The proposed methodology can be regarded as a sort of extension of the 'Method of Limits' originally developed by Fechner [3]. The main purpose of Fechner's method is to determine a threshold which represents the level of the stimulus property at which it becomes detectable. It is achieved by gradually increasing (decreasing) the intensity of the studied property in discreet steps until it is detectable (not detectable). In our case, quality represents the property which is gradually decreased (by system- in order to find the level at which assessor notice the change in the quality) and increased (by participants - in order to find the maximum level of quality). The aim is to find the acceptability threshold over time with respect to the quality of the displayed material.

3. DETAILS OF SUBJECTIVE STUDY

An audiovisual clip extracted from the first episode of a nature documentary series titled 'Life' (produced by BBC television) was used. With respect to the experimental design the original material was cut to 30min and 55sec while still preserving a semantic structure. The video contained a variety of camera angles and shots, including ultra-slow motion captures as well as action scenes with a lot of movement, details and close-ups. High quality material (1080p version) extracted from a Bluray disc served as a starting point for further processing using H.264/AVC encoder (x264). In total, 12 different quality levels were prepared and variations in-between the levels were achieved by using different settings of quantization parameter (QP). The test material was prepared such that the steps between quality levels were similar with respect to the just noticeable difference (JND) threshold. The values of QP were set according to [7, 8] and also based on our own investigations. The highest quality was achieved by setting the QP in H.264/AVC encoder to 0. To produce the lowest quality used in this experiment, QP was set to 48. Subsequently, all video clips were decoded to YUV format (4:2:0). The original audio track (stereo, 48 KHz) was used for the audio playback.

3.1 Participants

22 subjects (8 female, 14 male) took part in the experiment. The average age was 33 years (age range 24 – 61 years). The subject pool consisted of students and employees from the Norwegian University of Science and Technology in Trondheim / Norway. Participants were screened for visual acuity using Snellen chart, resulting in 2 subjects having slightly lower visual acuity (20/30) than the others. Consequently, results provided by these two subjects were removed from data analysis. Seven participants had worked with HD content on a professional ground and / or had watched movies in HD quality on a regular basis. They can be regarded as experienced viewers with respect to HD audiovisual material.

3.2 Test Procedure

The test procedure was divided into pre-test, test and post-test session. In the pre-test session, each subject received instructions (written and oral) and watched two 5min long training clips. The purpose of the training was to familiarize participants with the experimental setup, types of distortions, and with the sensitivity of the quality adjustment device. Assessors were allowed to ask questions during the training. The training clips were selected to span the same quality range as the test clip. In the second part, participants evaluated the quality of an over 30min long audiovisual material. To familiarize subjects with the highest quality of the presented content, the maximum quality was displayed for the first three minutes. Participants were told that the first degradation in the quality would occur at 1 to 5 minutes into the clip. In the post-test session, qualitative data was gathered. Participants were asked questions regarding difficulties (if any), and positive/negative aspects of the evaluation method. We were also interested in assessors' attention to the task/content as well as in their interest in the presented material. The clips were presented on a 50" full HD plasma screen. Two professional grade active monitor loudspeakers were used for sound reproduction. The experiment was conducted in a room particularly designed for video and audio quality tests, with appropriate lighting and room acoustic treatment. Viewing, listening and lighting conditions were set according to ITU recommendations BP.500-11 and P.911 [5, 6]. Participants were sitting in a comfortable, cinema style chair with a slide-out tray, which was used for placing the adjustment device on.

4. RESULTS

Besides the numerical data (time and values corresponding to quality levels), also information about the source of the quality change (system or user) was gathered. Participant's responses were collected at time resolution of 1 millisecond which resulted in a huge set of data being recorded.

We were particularly interested in investigating relations between the three-minute time slots (from one to another automatic degradation procedure) and users' responses to quality changes. The objective of this work was to find answers to the following questions:

- Do the quality expectations decrease over time and with increased involvement in the content?
- How fast can participants notice quality changes and at which quality level does this happen?
- Is the quality level at which the change is noticed similar to the quality level set by the participant?

To proceed with data analysis which would allow us to investigate above inquiries we created three subsets of data. These subsets consist of:

1) quality levels averaged over last minute of each three minute time section
2) reaction time to quality changes introduced by the system
3) quality levels corresponding to such reaction time

Each of these data subsets had the same size 20x10, where 20 correspond to the number of participants and 10 to the number of three – minute time periods. For the subset 1) we decided to focus on the last 60 seconds of each time section as it should be the period when users have established their preferences with respect to the perceived quality. To reduce unwanted noise the results provided by assessors during these periods were averaged. Analysis of Variance (ANOVA) was used as a tool for extraction of information from this data. ANOVA is a well established statistical method that compares the deviation between means of several groups to the random deviation within groups [8]. With respect to the assumptions required to use ANOVA the data was checked for normality and homogeneity of variance across assessors as well as across time slots. All data sets showed normal distribution and homogeneous variance.

A mixed-effects model of ANOVA was applied to each of the mentioned subsets of data with participants representing random-effect type factor and time slots representing fixed-effect type factor. Using such a model we can better understand which factor is responsible for most of the variation in the data and also compare the main effect of each of them. The ANOVA test used for the following analysis was conducted at the 0.05 significance level.

Results for the subset data 1) are presented in Table 1.

Table 1. Results of ANOVA test for the first data subset.

Source	DF	Seq SS	Adj SS	Adj MS	F	*p*
User	19	152.074	152.074	8.004	4.12	0.000
Time slot	9	12.339	12.339	1.371	0.71	0.702
Error	171	331.847	331.847	1.941		
Total	199	496.260				

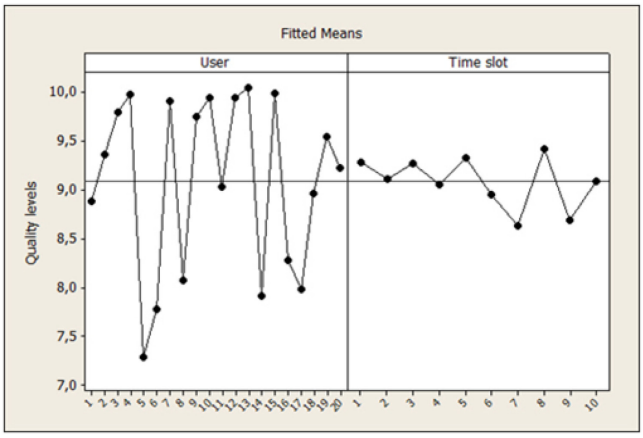


Figure 1. Main effects plot for quality levels averaged over last minutes of each 3 min time section.

We can see that the mean quality levels for time slots are not significantly different from each other ($F = 0.71$; $p > 0.05$) whereas the mean quality levels across users greatly differ in the statistical sense ($F = 4.12$; $p < 0.05$). The results indicate that indeed there were no significant changes in the quality expectations over the entire length of the clip. Assessors were quite consistent in their choices during the whole period of time and variation across time slots (see SS of time slots in ANOVA table) in quality setup may be due to presented content/the particular scene. On the other hand we can conclude that there were big variations in means of quality levels across participants which is presumably due to differences between the ways assessors were using the adjustment device to improve the quality. From the Fig. 1 we can observe that variation due to user differences is much larger then variation due to time slot differences. We can also notice that the mean values for time slots are relatively alike to each other.

Similar results were obtained for the data subset 2). Table 2 shows that there were no significant differences in reaction time to quality changes across all time sections ($F = 0.62$; $p > 0.05$). The average difference between the start of each degradation process and the time at which users noticed a change in the quality was around 26000 milliseconds. This corresponds to a 3 levels drop in the quality before assessors reacted. From Table 2 it can also be seen that mean reaction times for users are significantly different ($F = 2.03$; $p < 0.05$). Looking at Fig.2 we can see more clearly how means for reaction times across users as well as across time slots were distributed.

Table 3 shows again that there is no significant difference between means for time sections in subset data 3) ($F = 1.75$; $p > 0.05$), whereas we can see statistical significance with respect to means comparison across the users ($F = 5.14$; $p < 0.05$). One could notice that this time the p value in the first case is relatively low compared to subset 1) an 2) and very close to the significance level at 0.05. This is due to high values in one particular time slot

Table 2. Results of ANOVA test for the second data subset.

Source	DF	Seq SS	Adj SS	Adj MS	F	*p*
User	19	13320424016	13300287372	700015125	2.03	0.010
Time slot	9	1928817039	1928817039	214313004	0.62	0.779
Error	168	58068052784	58068052784	345643171		
Total	196	73317293839				

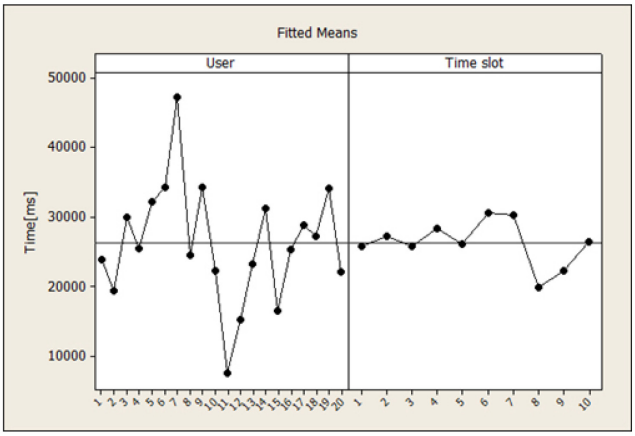


Figure 2. Main effects plot for reaction time.

(see Fig. 3). Fig. 4 shows comparison between average quality values corresponding to the time at which participants noticed a change (QLRT), and average quality values set by them during the last minute of each time section (AQL). The bottom plot (Diff) represents the actual differences between them. It can be seen that these differences are alike for each time period and in average correspond to three quality levels. This clearly suggests that it is easier for assessors to distinguish between neighboring quality levels when they exercise control over the displayed quality themselves (e.g. during the adjustment procedure) than when the process is independent from them and happens at random. We could also conclude that quality level set by assessors towards the end of each time slot is not necessarily the one which represent the acceptable quality level for most of the participants.

Table 3. Results of ANOVA test for the third data subset.

Source	DF	Seq SS	Adj SS	Adj MS	F	*p*
User	19	265.255	265.255	13.961	5.14	0.000
Time slot	9	42.705	42.705	4.745	1.75	0.082
Error	171	464.195	464.195	2.715		
Total	199	772.155				

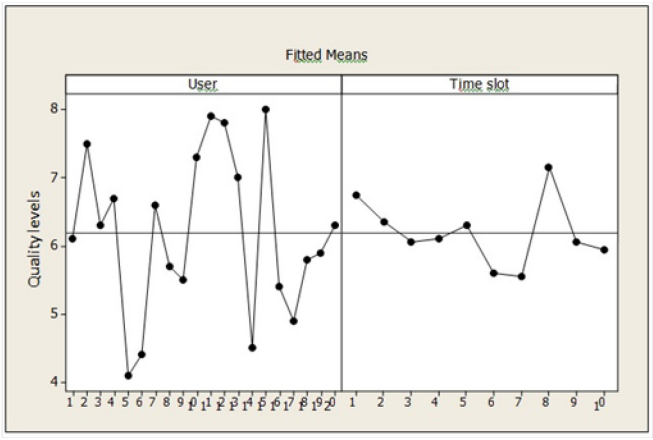


Figure 3. Main effects plot for quality levels corresponding to reaction time.

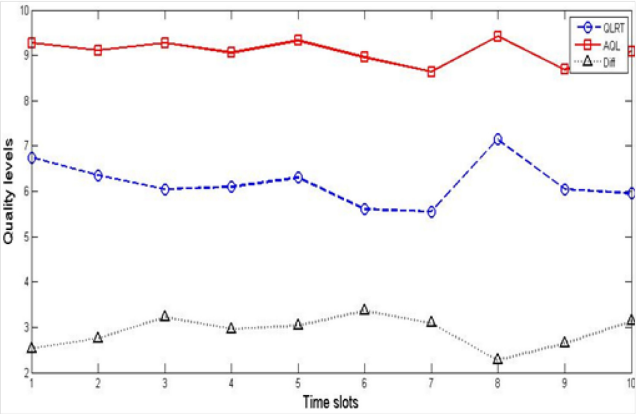


Figure 4. Comparison of sensitivity to quality changes under different conditions.

The averaged acceptable quality level is rather related to the one corresponding to reaction times.

5. CONCLUSIONS AND FUTURE WORK

The experiment carried out using our novel subjective quality assessment methodology resulted in a set of data that was analyzed using the analysis of variance method. Specifically, we used the mixed-effects ANOVA model to test our data. This approach is often convenient to gain insight into the relationships between means of specific factors. Such procedures can help to find out about the presence or lack of differences across users and particular time intervals.

The results obtained confirm some of our previous conclusions and thoughts, and also deliver new findings. We discovered that quality expectations over extended period of time are rather constant and that the same holds for the reaction time to quality changes. These results suggest that the time dimension is not necessarily a factor influencing participants' quality expectations. Therefore, the reason behind fluctuations in the quality perception might be directly related to the test material itself and/or personal involvement in the content. In general, subjects reported great interest in the presented stimuli which might have an impact on the obtained results, but this needs to be verified. More interestingly, the data analysis showed that participants are less sensitive to quality changes when the process is independent from them than when they have the possibility to adjust the quality themselves. Based on the above we can clearly see that our novel methodology can produce results which contribute to a better understanding of assessor's behavior regarding quality selection, expectations and reactions to quality changes over extended periods of time. Furthermore, it can improve our knowledge about QoE providing results which cannot be obtained using popular MOS based approaches.

Our future plans can be specified in two domains: i) further validation of the method, and ii) more in-depth statistical data analysis. For the first domain the aim is to check usability of the method in other modal contexts, such as audio or mixed media. The influence of content representing different types of audio/video properties (and of even longer duration) on users' performance and results will also be investigated. Other types of artifacts, e.g. quality degradations caused by packet loss, might be considered at a later time. In addition, we are planning to record assessors' emotional state in order to check how (or if at all) emotions induced by content influence the process of quality level selection.

6. ACKNOWLEDGMENTS

This work was performed within the PERCEVAL project, funded by The Research Council of Norway under project number 193034/S10.

REFERENCES

[1] Borowiak A., Reiter U., and Svensson U.P., "Quality Evaluation of Long Duration Audiovisual Content", *The 9th Annual IEEE Consumer Communications and Networking Conference – Special Session on Quality of Experience (QoE) for Multimedia Communications*, pp. 353-357, 2012.

[2] Chen K. T., Wu C. C., Chang Y. C., and Lei C. L., "A crowdsourceable QoE evaluation framework for multimedia content ", *In Proceedings of the 17th ACM international conference on Multimedia* , pp. 491-500, 2009.

[3] Fechner G.T, "Elements of psychophysics" (Vol. 1). Original work published in 1860. Translated by H.E. Adler Holt, Rinehart and Winston, New York, 1966.

[4] ITU-R, Methodology for the subjective assessment of the quality of television pictures BT.500-12, 2009

[5] ITU-T, Subjective Audiovisual Quality Assessment Methods for Multimedia Applications P.911, 1998.

[6] Wang H., Qian X., and Liu G., "Inter Mode *Decision Based on Just Noticeable Difference Profile" Proceedings of 2010 IEEE 17th International Conference on Image Processing*, Hong Kong, 2010.

[7] Yang X., Tan Y., and Ling N., "Rate control for H.264 with two-step quantization parameter determination but single-pass encoding", *EURASIP Journal on Applied Signal Processing,* pp. 1-13, 2006.

[8] Bech S., Zacharov N., "Perceptual Audio Evaluation – Theory, Method and Application", 2006.

Workshop 4: UP-TO-US: User-Centric Personalized TV ubiquitOus and secUre Services

SIP-Based Context-Aware Mobility for IPTV IMS Services

Victor Sandonis
Departamento de Ingeniería Telemática
Universidad Carlos III de Madrid
Avda. de la Universidad, 30
28911 Leganés – Madrid (Spain)
vsandoni@it.uc3m.es

Ismael Fernandez
Departamento de Ingeniería Telemática
Universidad Carlos III de Madrid
Avda. de la Universidad, 30
28911 Leganés – Madrid (Spain)
ismafc06@gmail.com

Ignacio Soto
Departamento de Ingeniería de Sistemas Telemáticos
Universidad Politécnica de Madrid
Avda. Complutense, 30
28040 Madrid (Spain)
isoto@dit.upm.es

ABSTRACT
In this paper we propose a solution to support mobility in terminals using IMS services. The solution is based on SIP signalling and makes mobility transparent for service providers. Two types of mobility are supported: a terminal changing of access network and a communication being moved from one terminal to another. The system uses context information to enhance the mobility experience, adapting it to user preferences, automating procedures in mobility situations, and simplifying decisions made by users in relation with mobility options. The paper also describes the implementation and deployment platform for the solution.

Categories and Subject Descriptors
C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design – network communications, packet-switching networks, wireless communications

General Terms
Design, Experimentation, Standardization.

Keywords
Mobility, IMS, SIP, IPTV, Context-awareness

1. INTRODUCTION
Access to any kind of communication services anytime, anywhere and even while on the move is what users have become to expect nowadays. This poses incredible high requirements to telecom operators' networks to be able to serve the traffic of those services. Mobility requires wireless communications, but bandwidth is a limited resource in the air interface. To be able to overcome this limitation the operators are pushing several solutions such as the evolution of the air interface technologies, the deployment of increasingly smaller cells, and the combination of several heterogeneous access networks.

Offering connectivity through a combination of several access networks requires the ability to manage the mobility of terminals across them, while ensuring transparency to service providers. Changing access network implies a change in the IP address of the terminal. To prevent this change from having an effect on the terminal open communications, we need to use a mobility support solution to take care of it.

In telecom networks, operators are pushing the adoption of the IP multimedia subsystem (IMS) to provide, through their data IP-based networks, multimedia services, such as VoIP or IPTV. In this paper we propose a mobility support solution based on SIP, the signalling protocol used in IMS. The solution is completely integrated in IMS, does not require modifications to the IMS specifications, and supports mobility transparently to service providers.

The proposed solution supports terminal mobility, the movement of a terminal between access networks without breaking its communications. Additionally, it also supports session mobility, allowing a user to move active communications among its terminals. The last key point of our proposal is the use of a context information server to enhance the users' mobility experience, for example automating some configuration tasks.

The mobility solution proposed in this paper is valid for any service based on IMS, both for unicast and multicast communications. Nevertheless in the context of the Celtic UP-TO-US project we are focusing on the IPTV service: multicast-based LiveTV and unicast-based Video on Demand (VoD), which will be our use cases through the paper.

The rest of the paper is organized as follows: section 2 reviews the terminology and analyses works in the literature related to the paper; section 3 describes the proposed solution for transparent SIP-based context-aware mobility; section 4 covers the implementation of the solution, with a description of the used tools and platforms, and their behaviour; finally section 5 highlights the conclusions that result from our work.

2. BACKGROUND
The first part of this section presents an overview of the various types of mobility; then the related work relevant to the solution proposed here is discussed.

For the case of IPTV Services two types of mobility have been considered: session mobility and terminal mobility.

Session mobility is defined as the seamless transfer of an ongoing IPTV session from the original device (where the content is currently being streamed) to a different target device. This transfer is done without terminating and re-establishing the call to the corresponding party (i.e., the media server). This mobility service makes possible, for example, that the user can transfer the current IPTV session displayed in her mobile phone to a stationary TV (equipped with its Set-Top-Box).

Session mobility has two modes of operation, namely, push mode and pull mode:

- In the *push mode*, the original device (where the content is currently being streamed) discovers the devices to which a

session mobility procedure potentially could be initiated, selects one device and starts the mechanism.

- In the *pull mode*, the user selects in the target device, the on-going session he wants to move. Thus, the target device needs previously to learn the active sessions for this user that are currently being streamed to other devices. After selecting a session, the target device initiates the session mobility procedure.

Terminal mobility allows a terminal to move between access networks (the terminal changes its IP address) while keeping ongoing communications alive despite the movement. This type of mobility is aligned with current network environments that integrate different wireless and fixed access network technologies, using terminals with multiple network interfaces and enabling users to access IPTV service via different access networks.

The mobility solution presented in this paper considers three different ways to initiate the mobility:

- *Manual*: The user decides that he wants to initiate a mobility procedure.
- *Semi-Automatic*: The UP-TO-US system proposes initiating a mobility procedure. To perform this proposal the system takes into consideration the different context information related to the user, the terminals and the networks. The user always has the opportunity to accept or reject this suggestion.
- *Automatic*: The UP-TO-US system taking into consideration all the context information initiates a mobility procedure.

The user has the opportunity to configure as part of his preferences the mobility mode he prefers (manual/semi-automatic/automatic).

Terminal mobility support in IP networks has been studied for some time. In particular, the IETF has standardized the Mobile IP (MIP) protocols both for IPv4 [1] and IPv6 [2] to provide mobility support in IP networks. Mobile IP makes terminal mobility transparent to any communication layer above IP, including the applications and, therefore a node is able to change the IP subnetwork is being used to access the Internet without losing its active communications. Mobile IP is a good solution to provide terminal mobility support but its integration with IMS is far from trivial [3][4][5]. This is essentially because MIP hides from the application layer, including the IMS control, the IP address used by a mobile device in an access network, but IMS requires this address to reserve resources in the access network for the traffic of the services used in the mobile device.

Another alternative is to use the Session Initiation Protocol (SIP) to support mobility [6][7][8] in IP networks. In this respect, 3GPP has proposed a set of mechanisms to maintain service continuity [9] in the event of terminal mobility or session mobility. Using SIP to handle mobility presents the advantage of not requiring additional mechanisms outside the signalling used in IMS. But the traditional SIP mobility support is not transparent to the IPTV service provider.

Making mobility transparent to the IPTV service provider is a desirable feature that is feasible to provide. TRIM architecture [10] provides terminal mobility support in IMS-based networks without requiring any changes to the IMS infrastructure and without entailing any upgrades to the service provider's application. The main component of the TRIM architecture is a SIP Application Server (AS) located in user's home network. This AS stays on the signalling path between the two communicating end-points. It receives all the SIP signalling messages corresponding to the multimedia sessions and controls a set of Address Translators located also in the user' home network. A particular mobile user is served by one of these Address Translators. The Address Translator is configured by the SIP Application Server to address the media received from the multimedia service provider (i.e., IPTV service provider) to the appropriate location where the mobile terminal is at that moment. Analogously, it forwards the traffic in the opposite direction, i.e. from the mobile terminal to the IPTV service provider. In particular, the address translator simply changes the IP addresses and ports contained in the packets according to the configuration provided by the SIP Application Server. Therefore, the service provider always observes the same remote addressing information for the Mobile Terminal, no matter where the latter is located.

The UP-TO-US mobility solution uses the functionality explained above from the TRIM architecture[1] to provide not only terminal mobility but also session mobility (not considered in TRIM). Additionally, the UP-TO-US mobility solution integrates the concept of context-awareness mobility to enhance the user's mobility experience, for example proposing to the user a session transfer according to his preferences and the available nearby devices.

3. TRANSPARENT SIP-BASED CONTEXT-AWARE MOBILITY

In this section we describe the UP-TO-US IPTV Service Continuity solution that provides transparent SIP-based context-aware mobility.

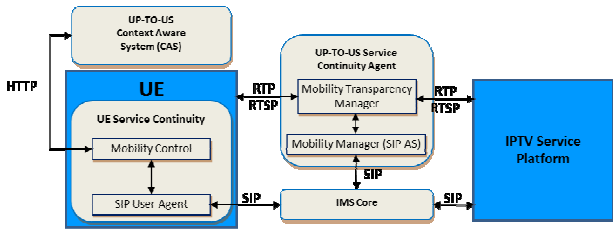


Figure 1. Architecture

3.1 Architecture

The architecture of the UP-TO-US IPTV Service Continuity solution (see figure 1) is composed by the modules explained in next subsections.

3.1.1 UP-TO-US Context-Awareness System

The UP-TO-US Context-Awareness System (CAS) is the module in charge of storing the context information of the user environment, the network domain and the service domain. It provides context information about mobility user preferences, available devices, available access networks, user's session parameters, etc. This information is useful to manage the different mobility procedures.

[1] The TRIM architecture also adds functionality (modifications) to user terminals that we have chosen to avoid in the UP-TO-US mobility solution.

3.1.2 UP-TO-US Service Continuity Agent

The UP-TO-US Service Continuity Agent is located in the home network. In order to provide mobility support functionalities, it is inserted both in the signalling path and in the data plane media flows of user's IPTV sessions. The UP-TO-US Service Continuity Agent is divided in two sub-modules: the Mobility Manager and the Mobility Transparency Manager.

3.1.2.1 Mobility Manager

The Mobility Manager is a SIP Application Server (AS) [6] located in the signalling path between the User Equipment (UE) and the IPTV Service Platform, such that receives all SIP signalling messages of users' multimedia sessions (the IMS initial filter criteria is configured to indicate that SIP session establishment messages have to be sent to the Mobility Manager). The Mobility Manager makes mobility transparent to the control plane of the service provider side. Additionally, it modifies SIP messages and Session Description Protocol (SDP) payloads to make media packets of multimedia sessions travel through the Mobility Transparency Manager. The Mobility Manager controls the Mobility Transparency Manager according to the information about IPTV sessions it extracts from SIP signalling messages. This information concerns the addressing parameters of the participants in the session and the session description included in the SDP payload of SIP messages.

3.1.2.2 Mobility Transparency Manager

The Mobility Transparency Manager is a media forwarder whose functionality is to make the mobility transparent to the user plane of the service provider side. This way, the content provider is unaware of UE movements between different access networks (terminal mobility) or movements of sessions between devices (session mobility) in the client side. The Mobility Transparency Manager is controlled by the Mobility Manager according to the information extracted from the SIP signalling messages of user's multimedia sessions. It is configured to properly handle the media flows of user's IPTV sessions[2]. This way, when the UE moves between different access networks (terminal mobility) or when the session is transferred between devices (session mobility), the Mobility Transparency Manager is configured to forward the media received from the content provider to the current location (access network) of the UE in case of terminal mobility or to the new UE in case of session mobility. In the opposite direction, user plane data received from the UE is forwarded to the content provider. This way, data packets are forwarded to the appropriate destination IP address and port according to the information provided by the Mobility Manager.

The result is that the content provider side always perceives and maintains the same remote addressing information regardless of the UE handoffs between different access networks (terminal mobility) or when the session is transferred between devices (session mobility). Figure 2 describes how the Mobility Transparency Manager forwards data packets in the user plane to the appropriate destination IP address and port.

[2] According with the ETSI TISPAN standard [11], the SDP offer of SIP signalling messages contains a media description for the RTSP content control channel. This way, RTSP packets exchanged between UE and the service platform are considered as belonging to a media flow in the data plane. The Mobility Transparency Manager handles RTSP packets as a RTSP proxy.

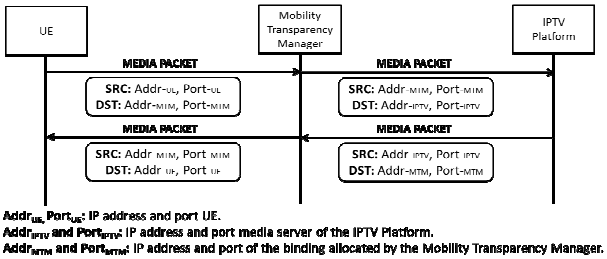


Figure 2. Mobility Transparency Manager forwarding

On the other hand, the Mobility Transparency Manager behaves as a RTSP proxy inserted in the RTSP signalling path between the UE and the IPTV Platform for Video on Demand service[2].

3.1.3 UE Service Continuity

The UE Service Continuity module is an element of the User Equipment. The UE Service Continuity interacts with the UP-TO-US Context Awareness System (CAS) and with the UP-TO-US Service Continuity Agent to manage mobility support functionalities. The UE Service Continuity module is divided in two sub-modules: the Mobility Control and the SIP User Agent.

3.1.3.1 Mobility Control

The Mobility Control sub-module has all the logic of mobility and makes mobility decisions taking into account the available context information and user mobility preferences obtained from the interaction with the CAS. This way, the Mobility Control sub-module subscribes to the mobility triggers in the CAS which monitors user context and notifies to the Mobility Control sub-module when a mobility event occurs. Possible mobility events are (1) trigger informing that the user is close to other devices different from the current one to initiate push session mobility, (2) trigger informing about the user proximity to the device to initiate pull session mobility and (3) trigger informing that the terminal has to perform a handover to another access network. On the other hand, the Mobility control sub-module retrieves from the CAS context information about mobility user preferences, available devices, available access networks, user's session parameters, etc. that is useful for handling mobility procedures.

Based on the combination of the existing context information that is available locally in the UE (for instance IP address and access network type of the interfaces of the UE, Wi-Fi signal strength level, parameters of the session that is active on the UE, etc.) and the information retrieved from the CAS, the Mobility Control sub-module makes mobility decisions on which mobility procedure is appropriate under a specific context (push session mobility, pull session mobility or terminal mobility procedures) and requests the SIP User Agent to exchange the appropriate SIP signalling to handle the different mobility procedures with the UP-TO-US Service Continuity Agent. This way, the UP-TO-US Service Continuity Agent is maintained updated with the current addressing information of the UE when it moves between access networks (terminal mobility) or the session is transferred between devices (session mobility).

3.1.3.2 SIP User Agent

The SIP User Agent is in charge of the SIP signalling exchange between the UE and the UP-TO-US Service Continuity Agent to handle the different SIP procedures for registration, session establishment, session modification, session release, etc. SIP

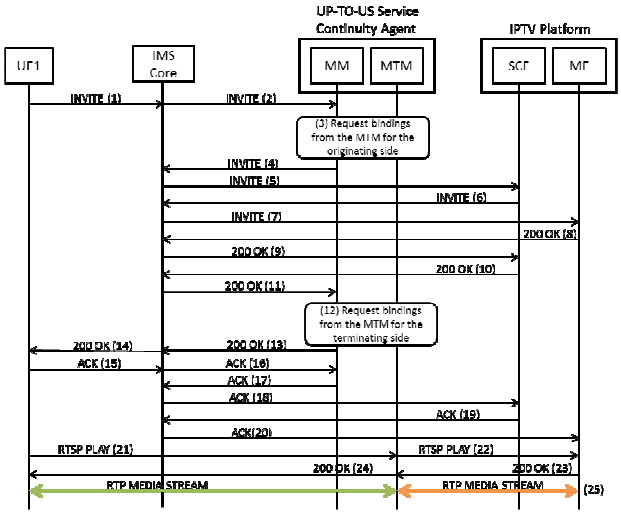


Figure 3. VoD session establishment

signalling messages travel through the IMS infrastructure to arrive to the UP-TO-US Service Continuity Agent.

3.2 Procedures of the proposed solution

The IPTV service continuity solution considers the Video on Demand (VoD) service and the LiveTV service. Next subsections explain the different procedures of the IPTV service continuity solution for VoD and LiveTV services. These procedures are based on TRIM [10] and IETF, ETSI TISPAN and 3GPP specifications [7][8][9] and [11]. The figures presented in the next sections show the exchange of the main messages between modules assuming that the user is already registered in the IMS on UE1 and UE2. The modules described in section 3.1 are shown in the figures (the MM sub-module is the Mobility Manager and the MTM sub-module is the Mobility Transparency Manager), but also the IPTV Service Control Function (SCF), the IPTV Media Function (MF) [12] of the IPTV Platform and the Elementary Control Function and the Elementary Forwarding Function (ECF/EFF) [13] appear in the figures. The SCF is the module of the IPTV platform responsible of controlling the sessions of all IPTV Services (service authorization, validation of user requests based on user profile, MF selection Billing & Accounting, etc.). The MF is in charge of the media delivery to the users. The ECF/EFF can be described as the access router that processes Internet Group Management Protocol (IGMP) [14] signalling and forwards multicast traffic.

3.2.1 Video on Demand Service

This section describes the procedures of the solution for the VoD Service.

3.2.1.1 Session Establishment

The establishment of a VoD session is illustrated in figure 3. After the user selects the content he wants to watch, UE1 begins the session establishment issuing an INVITE request (1) to the IMS Core that according with the IMS initial filter criteria is routed (2) towards the Mobility Manager. The SDP offer of the INVITE message contains the media descriptions of the RTSP content control flow and the content delivery flows, and it is processed by the Mobility Manager (3). For each media component included in the SDP offer:

a) The Mobility Manager obtains the addressing information where the UE1 will receive the data traffic of the media component. This is the IP address and port included in the media component description. The Mobility Manager requests the Mobility Transparency Manager to create a binding for this addressing information. As result, the Mobility Transparency Manager allocates a new pair of IP address and port, and returns it to the Mobility Manager.

b) The Mobility Manager modifies the media description of the SDP payload replacing the IP address and port by the binding obtained from the Mobility Transparency Manager. This way, the data traffic of the media component will be sent by the MF to the Mobility Transparency Manager.

After this processing, the Mobility Manager issues a new INVITE message (4) that includes the modified SDP offer that according to the IMS initial filter criteria reaches the Service Control Function of the IPTV platform (5). The SCF performs service authorization and forwards the INVITE message (6) to the appropriate MF. Since the SDP payload received by the MF includes the IP addresses and ports of the bindings allocated by the Mobility Transparency Manager, the data traffic of the different media components will be sent by the MF to the Mobility Transparency Manager. On the other hand, the MF generates a 200 OK response that includes a SDP payload with the media descriptions of the content delivery flows and the RTSP content control flow for the MF side. The MF includes in the SDP payload the RTSP session ID of the content control flow and the RTSP URI to be used in RTSP requests. The 200 OK is routed back and arrives to the Mobility Manager (8 -11). The Mobility Manager processes the 200 OK message from the MF and its SDP offer. For each media component included in the SDP offer:

a) The Mobility Manager obtains the addressing information of the MF for the data traffic of the media component. This is the IP address and port included in the media component description. The Mobility Manager requests the Mobility Transparency Manager to create a binding for this addressing information. As a result, the Mobility Transparency Manager allocates a new pair of IP address and port, and returns it to the Mobility Manager.

b) The Mobility Manager modifies the media description replacing the IP address and port by the binding obtained from the Mobility Transparency Manager. This way, the data traffic of the media component will travel through the Mobility Transparency Manager.

c) The RTSP URI parameter included by the MF in the SDP payload is modified in order to insert the Mobility Transparency Manager in the RTSP signalling path between the EU and the MF.

This way, both the content delivery flow and the RTSP content control flow travel through the Mobility Transparency Manager. The result is that the MF always perceives and maintains the same remote addressing information (the Mobility Transparency Manager binding) regardless the session mobility or terminal mobility. Therefore, mobility is transparent to the MF. Then, the

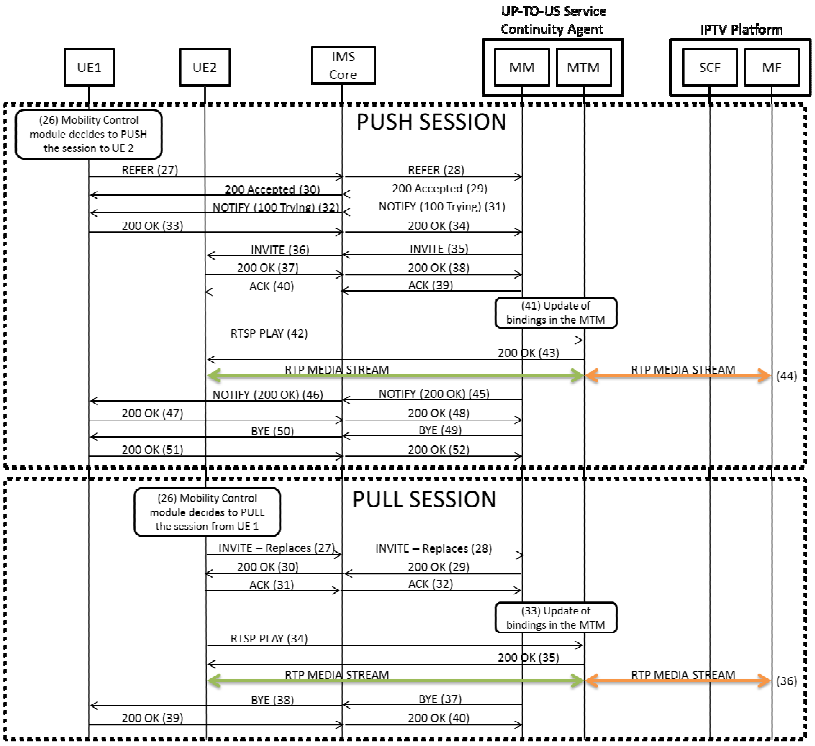


Figure 4. VoD push/pull session mobility

Mobility Manager issues a new 200 OK message with the modified SDP payload (13-14). The UE1 sends an ACK message to complete the SIP session establishment (15-20). Finally, the UE1 sends an RTSP PLAY message (21) to play the content. The RTSP PLAY message is issued to the RTSP URI using the RTSP session ID extracted from the SDP payload of the previous 200 OK message (14). This way, the RTSP PLAY message reaches the Mobility Transparency Manager, which generates a new RTSP PLAY message (22) and sends it to the MF. The RTSP 200 OK response is routed back to the UE 1, travelling through the Mobility Transparency Manager (23-24). At this moment (25), the MF begins streaming the RTP flows that travel through the Mobility Transparency Manager before arriving to the UE1. The Mobility Transparency Manager makes the mobility transparent to the MF by forwarding data packets according to the bindings created (steps (3) and (12)).

3.2.1.2 Push Session Mobility

Push session mobility procedures for VoD service are shown in figure 4. The figure represents a scenario where the UE1 transfers a multimedia session to the UE2. It is assumed that the VoD session is already established on UE1. The Mobility Control module of UE1, after receiving a trigger from the UP-TO-US CAS or under a user request, decides to push the session to UE2. UE2 addressing information to route subsequent messages is obtained by the Mobility Control module from CAS (26). In order to transfer the multimedia session to the UE 2, UE1 sends a REFER message [15] that arrives to the Mobility Manager after being routed by the IMS Core (27-28). The REFER message includes:

- Refer-To header that indicates the destination of the session transfer. In this case the addressing information of the user in UE2 (SIP URI of the user in UE2).
- Target-Dialog header that identifies the SIP session that has to be transferred from UE1 to UE2.
- Referred-By header that indicates who is transferring the session (SIP URI of the user in UE1).

The Mobility Manager processes the REFER method and informs to UE1 that it is trying to transfer the session (29-34). In order to transfer the session to the UE2, the Mobility Manager initiates a new SIP session with the UE2. The SDP payload of the INVITE message (35) includes the bindings generated by the Mobility Transparency Manager, the RTSP session ID of the content control flow and the RTSP URI to be used in RTSP requests obtained when the session was established on the UE1 (see figure 3)[3]. After the session is established with the UE2, the Mobility Manager updates the Mobility Transparency Manager (41) with the addressing information of UE2 extracted from the 200 OK message (38) (media descriptions of the RTSP content control flow and the content delivery flows for UE2). This way, the Mobility Transparency Manager can forward the content delivery flows and the RTSP content control flow to the new device, UE2. Note that the MF always perceives and maintains the same remote addressing information (the Mobility Transparency Manager binding) regardless the session is transferred between devices

[3] Note that if the UE2 does not support the codecs used in the session, the Mobility Manager could re-negotiate the session (by means of a re-INVITE) with the MF to adapt the multimedia session to the codecs supported by the UE2.

(session mobility), and it is also not involved in the mobility SIP signalling. Therefore, the session mobility is transparent to the MF.

Finally, in order to play the content, UE2 sends a RTSP PLAY (42) message to the MF. The RTSP PLAY message is issued to the RTSP URI using the RTSP session ID extracted from the SDP payload of the previous INVITE message (35). This way, the RTSP PLAY message reaches the Mobility Transparency Manager that answers with a RTSP 200 OK message (43) to UE2. Note that the Mobility Transparency Manager does not issue a new RTSP PLAY message to the MF because the session is already played. Since the Mobility Transparency Manager has been updated with the addressing information of the UE2 (step 41), it can forward the content delivery flows and the RTSP content control flow to the new device, UE2 (44). Once the session has been transferred to the UE2, the Mobility Manager informs the UE1 about the success of its requested session transfer, NOTIFY message (45-46), and terminates the session with the UE1 (49-52).

3.2.1.3 Pull Session Mobility

Pull session mobility procedures for VoD service are shown in figure 4. The figure represents a scenario where the UE 2 obtains a multimedia session that is active on the UE 1. It is assumed that the VoD session is already established on UE1. The Mobility Control module of UE2, after receiving a trigger from the CAS or under a user request, decides to pull the session that is established on the UE 1 (26). Information about the session established on UE 1 needed for performing the pull session mobility is obtained by the Mobility Control module from the CAS. In order to transfer the multimedia session from UE 1 to UE 2, UE 2 SIP User Agent triggered by the mobility control module sends an INVITE request (27-28) that includes:

a) Replaces header [16] that identifies the SIP session that has to be replaced (pulled from UE 1).

b) SDP payload with the media descriptions of the RTSP content control flow and the content delivery flows for UE 2.

The Replaces header allows the Mobility Manager to identify the SIP session that the UE 2 wants to pull. This way, the Mobility Manager sends a 200 OK response (29) to UE 2 that includes the bindings generated by the Mobility Transparency Manager, the RTSP session ID of the content control flow and the RTSP URI to be used in RTSP requests obtained when the session was established on the UE1 (see figure 3)[2]. Once the session establishment between UE2 and the Mobility Manager has finished (29-32), the Mobility Manager updates the Mobility Transparency Manager with the addressing information of the UE2 extracted from the INVITE-Replaces message (28) (media descriptions of the RTSP content control flow and the content delivery flows for UE2). This way, the Mobility Transparency Manager can forward the content delivery flows and the RTSP content control flow to the new device, UE2. Note that the MF always perceives and maintains the same remote addressing information (the Mobility Transparency Manager binding) regardless the session is transferred between devices (session mobility), and it is also not involved in the mobility SIP signalling. Therefore, the session mobility is transparent to the MF.

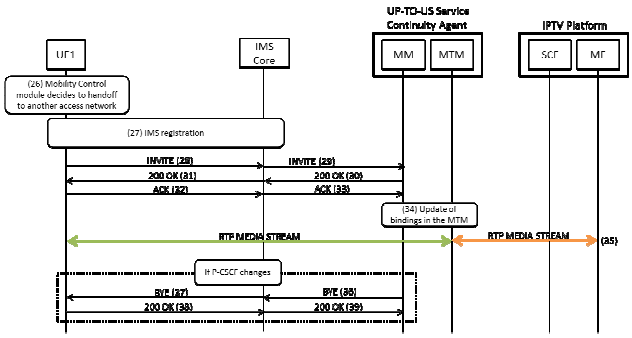


Figure 5. VoD terminal mobility

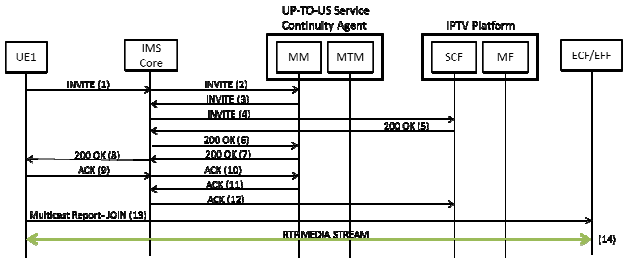


Figure 6. LiveTV session establishment

Finally, in order to play the content, UE2 sends a RTSP PLAY (34) message to the MF. The RTSP PLAY message is issued to the RTSP URI using the RTSP session ID extracted from the SDP payload of the previous 200 OK message (30). This way, the RTSP PLAY message reaches the Mobility Transparency Manager that answers with a RTSP 200 OK message (35) to UE2. Note that the Mobility Transparency Manager does not issue a new RTSP PLAY message to the MF because the session is already played. Since the Mobility Transparency Manager has been updated with the addressing information of the UE2 (step 33), it can forward the content delivery flows and the RTSP content control flow to the new device, UE2 (36). Once the session has been transferred to the UE2, the Mobility Manager terminates the session with the UE1 (37-40).

3.2.1.4 Terminal Mobility

Terminal mobility procedures for VoD service are shown in figure 5. The figure represents a scenario where the UE1 performs a handover between two access networks. It is assumed that the VoD session is already established on UE1. The Mobility Control module of the UE 1, after receiving a trigger from the CAS or under user request, decides to perform a handover to another access network (26). This way, after obtaining IP connectivity in the new access network, the UE 1 registers in the IMS to be reachable in the new IP address (27). In order to inform the Mobility Manager about the new addressing information in the new access network, the UE 1 sends an INVITE message (28) that includes the SDP payload with the media descriptions of the RTSP content control flow and the content delivery flows updated to the new addressing information in the new access network. This INVITE is a re-INVITE if the Proxy-Call Session Control Function (P-CSCF) is the same regardless of the handover between access networks or a new INVITE with a Replaces header in case that the P-CSCF has changed with the handover between access networks (a new dialog has to be established because the route set of proxies traversed by SIP messages changes). When the Mobility Manager processes the INVITE

request, it answers with a 200 OK response (30) to UE 1 that includes the bindings generated by the Mobility Transparency Manager, the RTSP session ID of the content control flow and the RTSP URI to be used in RTSP requests obtained when the session was established (see figure 3). Once the Mobility Manager receives the ACK message (33) that completes the session establishment, it updates the Mobility Transparency Manager with the new addressing information of the UE 1. This way, the Mobility Transparency Manager can forward the content delivery flows and the RTSP content control flow (35) to the new location of the UE 1 (new access network). Note that the MF always perceives and maintains the same remote addressing information (the Mobility Transparency Manager binding) regardless the UE handoffs between different access networks. Therefore, the terminal mobility is transparent to the MF. Note that the UE 1 does not issue a new RTSP PLAY message to the MF because the session is already playing on the device. Finally, in the case the P-CSCF changes with the handover between access networks, the Mobility Manager close the session of the old SIP dialog to release the reserved resources in the old access network (36-39).

3.2.2 LiveTV Service

This section describes the procedures of the solution for the LiveTV Service.

3.2.2.1 Session Establishment

The establishment of a LiveTV session is illustrated in figure 6. After the user selects a BroadCast (BC) service, UE1 begins the session establishment issuing an INVITE request (1) to the IMS Core that according with the IMS initial filter criteria is routed (2) towards the Mobility Manager. The SDP offer of the INVITE message contains the media descriptions of the BC session that is processed by the Mobility Manager. Since the request is about a BC service, the Mobility Manager does not modify the SDP offer because the data traffic of the BC session will be sent by multicast. Therefore, it is not needed that the data traffic travels through the Mobility Transparency Manager. Then, the Mobility Manager forwards the unaltered INVITE message (3), that is routed towards the SCF of the IPTV platform (4) according with the IMS initial filter criteria. After performing service authorization, the SCF checks the SDP offer and generates a 200 OK response that includes in the SDP payload the same media descriptions of the received INVITE. The 200 OK is routed back towards the Mobility Manager (5-6). Then, the Mobility Manager sends the 200 OK message without any modification in its SDP payload because it is not needed that the data traffic travels through the Mobility Transparency Manager (it is multicast traffic). When UE1 receives the 200 OK message, it completes the SIP session establishment (9-12). Finally, UE 1 joins the multicast group of the BC session by sending an IGMP JOIN message (13) to the ECF/EFF which starts forwarding the multicast traffic of the BC session to the UE 1 (14).

3.2.2.2 Push Session Mobility

Push session mobility procedures for LiveTV service are shown in figure 7. The figure represents a scenario where the UE1 transfers a multimedia session to the UE2. It is assumed that the LiveTV session is already established on UE1. Regarding the SIP signalling, it is similar to the VoD service case (see previous explanation of SIP signalling for VoD service push session mobility for more details). The main difference between the procedures for VoD service and LiveTV service is that the Mobility Manager does not modify the SDP payload of SIP messages because the data traffic of the BC session is sent by multicast. Therefore, it is not needed that the data traffic travels

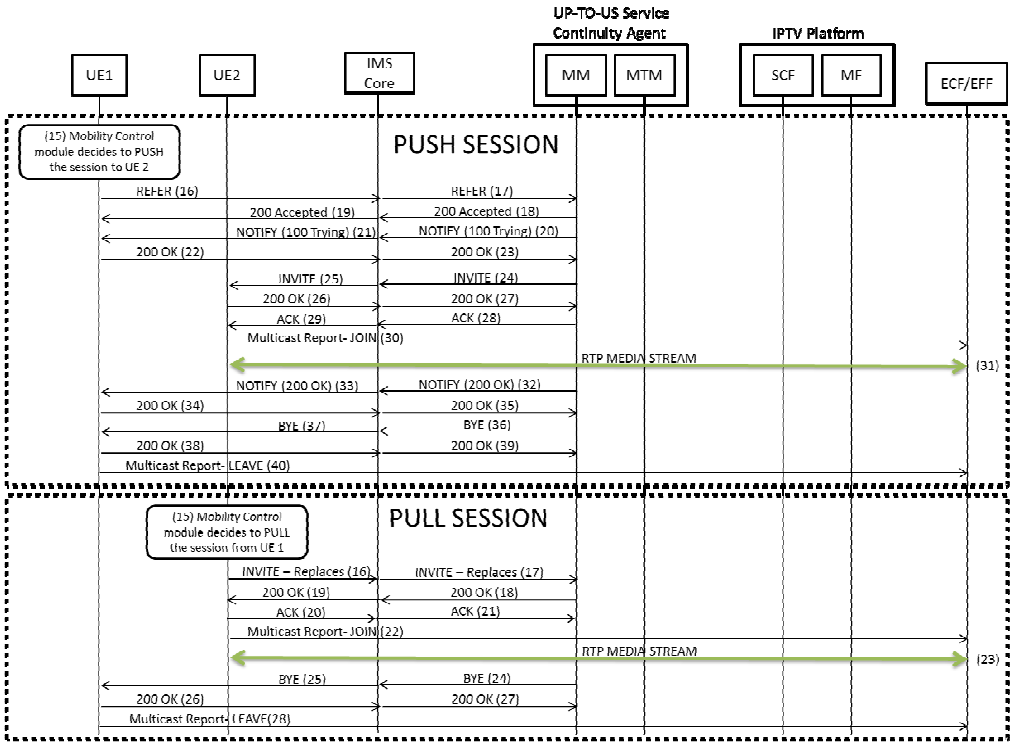


Figure 7. LiveTV push/pull session mobility

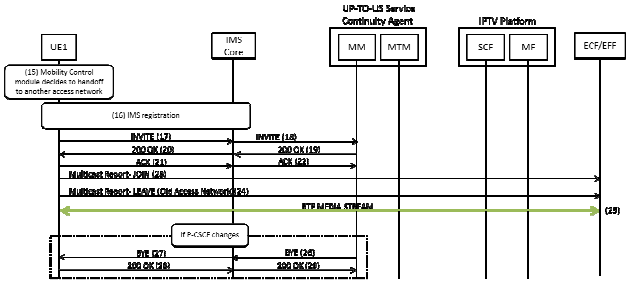


Figure 8. LiveTV terminal mobility

through the Mobility Transparency Manager. Another difference in the procedure is that after the new SIP session establishment, between the Mobility Manager the UE2, is completed (24-29), the UE 2 joins to the multicast group of the BC session by sending an IGMP JOIN message (30) to the ECF/EFF to receive the multicast traffic of the BC session (31). On the other hand, once the session has been transferred to the UE 2, the UE 1 leaves the multicast group of the BC session by sending an IGMP LEAVE message to the ECF/EFF (40).

3.2.2.3 Pull Session Mobility

Pull session mobility procedures for LiveTV service are shown in figure 7. The figure represents a scenario where the UE 2 obtains a multimedia session that is active on the UE 1. It is assumed that the LiveTV session is already established on UE1. As with the push session mobility, the SIP signalling is similar to the VoD service case (see previous explanation of SIP signalling for VoD service pull session mobility for more details). The main difference is the same, it is not needed that the data traffic travels through the Mobility Transparency Manager because it is sent by multicast, thus the Mobility Manager does not modify the SDP payload of SIP messages. Besides, the UE 2 joins the multicast group of the BC session by sending an IGMP JOIN message (22) to the ECF/EFF to receive the multicast traffic of the BC session (23) after completing the new SIP session establishment between the UE2 and the Mobility Manager. Finally, the UE 1 leaves the multicast group of the BC session by sending an IGMP LEAVE message to the ECF/EFF (28) once the session has been transferred to the UE2.

3.2.2.4 Terminal Mobility

Terminal mobility procedures for LiveTV service are shown in figure 8. The figure represents a scenario where the UE1 performs a handover between two access networks. It is assumed that the LiveTV session is already established on UE1. Again, the SIP signalling is similar to the VoD service case (see previous explanation of SIP signalling for VoD service terminal mobility for more details). The difference between the VoD service case and LiveTV service case is that the Mobility Manager does not modify the SDP payload of SIP messages because the data traffic of the BC session is sent by multicast and it is not necessary to make it travel through the Mobility Transparency Manager. Moreover, the UE1 joins to the multicast group of the BC session in the new access network and leaves it in the old access network by sending the appropriate IGMP messages (23-24) to the ECF/EFF to receive the multicast traffic of the BC session.

4. DEPLOYMENT PLATFORM

In this section we present the implementation of the UP-TO-US mobility solution. We have deployed a testbed (see figure 9) with all the components of the architecture already presented: an IMS Core, a Mobility Manager Application Server which controls the Signalling Plane, a Transparency Manager that controls the Media Plane, the CAS with only simplified functions to perform mobility, an IPTV Server, and the user clients.

We have used the Fokus Open IMS Core[4] as our IMS platform; this is a well-known open source implementation. In our testbed, all the components of the OpenIMS core have been installed in a ready-to-use virtual machine image.

We have provisioned two public identities in the HSS, UE1 and UE2, which represent two terminals of a user. We have also configured the Mobility Manager as an Application Server in the Core, and when the user calls the IPTV server from one of his terminals the corresponding INVITE message is routed to our Mobility Manager AS, so mobility support can be provided.

The Mobility Manager has been developed as an application deployed in a Mobicents JAIN SLEE Application Server[5]. Mobicents JAIN SLEE is the first and only open source implementation of the JAIN SLEE 1.1 specification[6]. With Mobicents JAIN SLEE we have a high throughput, low latency event processing application environment, ideal to meet the stringent requirements of communication applications, such as network signalling applications, which is exactly our case.

The Transparency Manager is a standalone Java application, which offers an RMI Interface to the Mobility Manager and uses Java sockets to redirect media flows.

In order to simulate the user terminals and also the IPTV Server, we have used SIPp[7] scripts. SIPp is an open source test tool and traffic generator for the SIP protocol, which can also send media traffic, and is capable of executing shell commands.

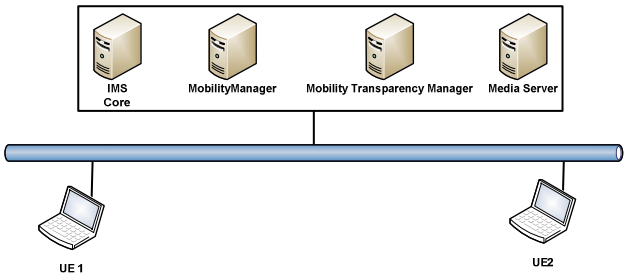


Figure 9. Testbed

[4] Open IMS Core, Fraunhofer Institute for Open Communication Systems (http://www.openimscore.org/).

[5] http://www.mobicents.org/slee/intro.html

[6] http://www.jcp.org/en/jsr/detail?id=240

[7] http://sipp.sourceforge.net/

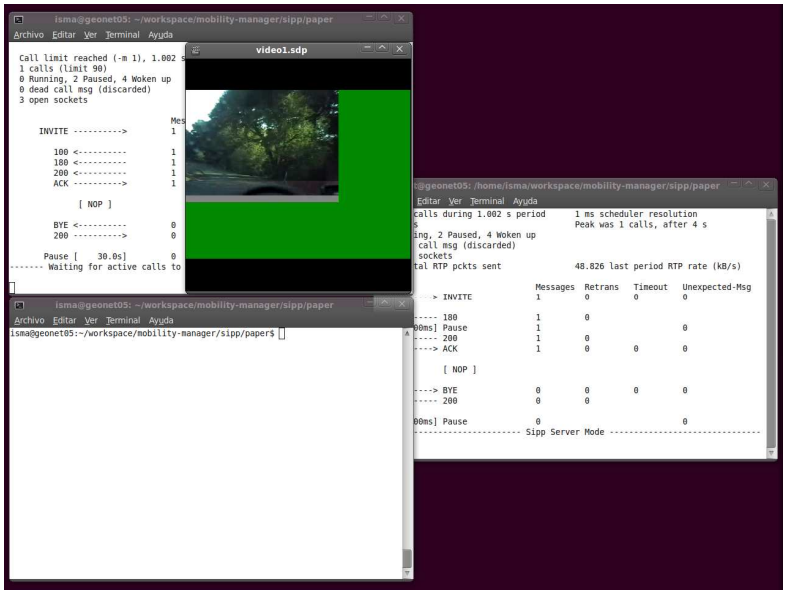


Figure 10. Pull session transfer: Initial snapshot

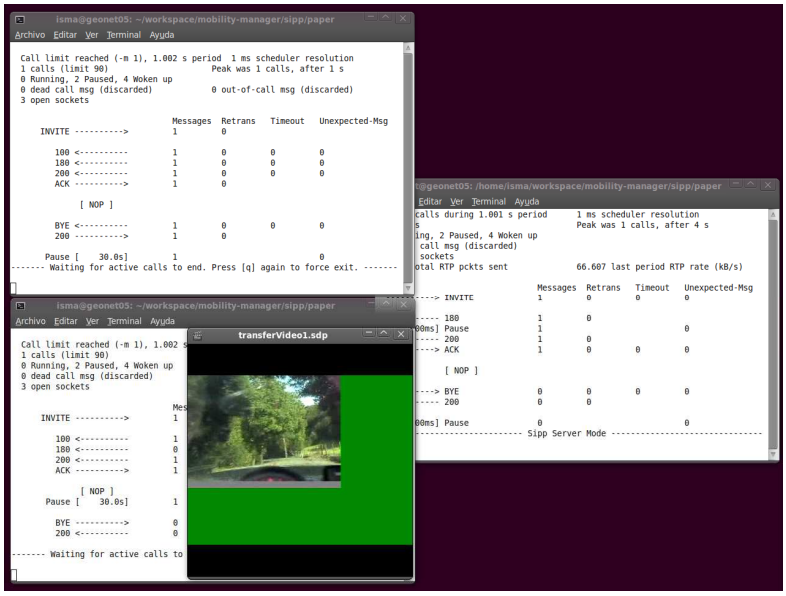


Figure 11. Pull session transfer: Final snapshot

The IPTV Server is a SIPp script which receives and accepts SIP session requests, and uses the ability of sending media traffic to deliver content to the clients.

Regarding the clients, we have developed several SIPp scripts which show the SIP flows, and call shell commands to reproduce the content being received. For that last purpose it uses Totem, the official movie player of the Gnome Desktop Environment.

The client also writes in a file the necessary information to perform a pull session transfer. This is a simple way in which we simulate the CAS.

Our test platform is formed by three Linux Ubuntu computers. One of them runs the Mobicents JAIN SLEE with the Mobility Manager application deployed, the Transparency Manager, the SIPp scripts which emulate the IPTV Server and the IMS Core (installed in a virtual machine). The other two Linux Ubuntu computers are the clients UE1 and UE2.

We have tested pull session mobility and push session mobility in our testbed. In figures 10 and 11, which directly represent two snapshots taken in our test platform, we can see the pull session mobility process. In figure 10 we can see a SIPp script representing the first terminal of the user calling the IPTV server (which is also a SIPp script). The IPTV server is sending the content that is being played using Totem.

In figure 11, we can see another SIPp script representing the second terminal of the user, which initiates the pull session

mobility. The session with the first terminal is closed, and the content is now being played in the second terminal.

In order to do preliminary tests about the mobility performance, we have taken two different measurements, one for pull session mobility and the other one for push session mobility. For pull session mobility, we measure the time it takes since the INVITE message with Replaces header is sent till the first content packet is delivered to the new terminal. We repeated the measurement 30 times obtaining an average pull session mobility delay of 58 ms.

For push session mobility, we measure the time it takes since the first terminal sends the REFER message till the first content packet is delivered to the second terminal. We repeated the measure 30 times obtaining an average push session mobility delay of 73 ms, a bigger value than the pull session mobility case. This is logical because the procedure is more complex.

5. CONCLUSION

In this paper we propose a SIP-based mobility support solution for IMS services. The solution makes mobility transparent to service providers, it is integrated with the IMS infrastructure, it does not require modifications to the IMS specifications, and it works with unicast and multicast services. The solution supports both terminal mobility, a terminal that changes access network; and session mobility, a communication moved from one terminal to another. Context-awareness is used in the solution, through a context server, to enhance users' mobility experience: facilitating typical operations when a particular situation is recognized, highlighting to the user mobility-related options that are relevant in those situations, and providing information useful to take decisions regarding mobility.

We have implemented our mobility support solution using standard tools and platforms such as the Fokus Open IMS Core as IMS Platform and the Mobicents application server as the SIP application server. We have made preliminary trials showing that our solution works as intended supporting terminal and session mobility with standard servers that are not affected by the mobility of terminals communicating with them.

As future work we intend to do extensive testing of the implementation, also considering the analysis of performance metrics, for example related with terminal handover delays.

6. ACKNOWLEDGMENTS

The work described in this paper has received funding from the Spanish Government through project I-MOVING (TEC2010-18907) and through the UP-TO-US Celtic project (CP07-015).

7. REFERENCES

[1] C. Perkins, IP Mobility Support for IPv4, Revised, RFC 5944, *Internet Engineering Task Force* (Nov. 2010).

[2] C. Perkins, D. Johnson, J. Arkko, Mobility Support in IPv6, RFC 6275, *Internet Engineering Task Force* (Jul. 2011).

[3] T. Chiba, H. Yokota, A. Dutta, D. Chee, H. Schulzrinne, Performance analysis of next generation mobility protocols for IMS/MMD networks, in: *Wireless Communications and Mobile Computing Conference*, 2008. IWCMC'08. International, IEEE, 2008, 68-73.

[4] T. Renier, K. Larsen, G. Castro, H. Schwefel, Mid-session macro-mobility in IMS-based networks, *IEEE Vehicular Technology Magazine*, 2 (1) (2007) 20-27.

[5] I. Vidal, J. Garcia-Reinoso, A. de la Oliva, A. Bikfalvi, I. Soto, Supporting mobility in an IMS-based P2P IPTV service: A proactive context transfer mechanism, *Comput. Commun*, 33 (14) (2010), 1736-1751

[6] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, SIP: Session Initiation Protocol, RFC 3261, *Internet Engineering Task Force* (Jun. 2002).

[7] R. Shacham, H. Schulzrinne, S. Thakolsri, W. Kellerer, Session Initiation Protocol (SIP) Session Mobility, RFC 5631, *Internet Engineering Task Force* (Oct. 2009).

[8] R. Sparks, A. Johnston, Ed., D. Petrie, Session Initiation Protocol (SIP) Call Control - Transfer, RFC 5589, *Internet Engineering Task Force* (Jun. 2009).

[9] 3GPP, IP Multimedia Subsystem (IMS) service continuity; Stage 2, TS 23.237 v11.1.0 Release 11, *3rd Generation Partnership Project* (3GPP) (Jun. 2011).

[10] Ivan Vidal, Antonio de la Oliva, Jaime Garcia-Reinoso, and Ignacio Soto. 2011. TRIM: An architecture for transparent IMS-based mobility. *Comput. Netw.* 55, 7 (May 2011).

[11] ETSI TS 183 063: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS-based IPTV stage 3 specification".

[12] ETSI TS 182 027: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; IPTV functions supported by the IMS subsystem".

[13] ETSI ES 282 001: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture".

[14] W. Fenner, Internet Group Management Protocol, Version 2, RFC 2236, *Internet Engineering Task Force* (Nov. 1997).

[15] R. Sparks, The Session Initiation Protocol (SIP) Refer Method, RFC 3515, *Internet Engineering Task Force* (April 2003).

[16] R. Mahy, B. Biggs, R. Dean, The Session Initiation Protocol (SIP) "Replaces" Header, RFC 3891, *Internet Engineering Task Force* (Sep. 2004).

MTLS: A Multiplexing TLS/DTLS based VPN Architecture for Secure Communications in Wired and Wireless Networks

Elyes Ben Hamida, Ibrahim Hajjeh
INEOVATION SAS
37 rue Collange, 92300 Levallois Perret - France
{elyes.benhamida, ibrahim.hajjeh}@ineovation.net

Mohamad Badra
Computer Science Department
Dhofar University, Salalah, Oman.
mbadra@du.edu.om

ABSTRACT

Recent years have witnessed the rapid growth of network technologies and the convergence of video, voice and data applications, requiring thus next-generation security architectures. Several security protocols have consequently been designed, among them the Transport Layer Security (TLS) protocol. This paper presents the Multiplexing Transport Layer Security (MTLS) protocol that aims at extending the classical TLS architecture by providing complete Virtual Private Networks (VPN) functionalities and protection against traffic analysis. Last but not least, MTLS brings enhanced cryptographic performance and bandwidth usage by enabling the multiplexing of multiple application flows over single TLS session. The effectiveness of the MTLS protocol is investigated and demonstrated through real experiments.

Categories and Subject Descriptors

C.2.0 [**Computer Communication Networks**]: General – *Data Communications, Security and Protection.*

General Terms

Design, Experimentation, Performance, Security, Standardization.

Keywords

SSL, TLS, DTLS, MTLS, VPN, Network Security, Performance Evaluation, Experimentation.

1. INTRODUCTION

Recent years have witnessed the rapid growth of network technologies and the convergence of video, voice and data applications, requiring thus next-generation security architectures. In this context, Virtual Private Networks (VPN) have been designed to fulfill these requirements and to provide confidentiality, integrity and authenticity between remote communicating entities. Confidentiality aims at protecting the exchanged data against passive and active eavesdropping, whereas integrity means that the exchanged data cannot be altered by an unauthorized third party. Finally, authenticity proves and ensures the identity of the legitimate communicating parties. Several security protocols have so far been designed and adopted to be used within VPN technologies, in particular the Transport Layer Security (TLS) protocol [1] and IP Security (IPSec) [2] that are widely used to secure transactions between communicating entities. However, this paper puts the focus on TLS rather than on IPSec for interoperability, deployment and extensibility reasons. Indeed, the IPSec protocol faces a number of deployment issues and fails in properly supporting dynamic routing, as already discussed in [3].

The TLS protocol is an IETF standard which provides secure communication links, including authentication, data confidentially, data integrity, secret keys generation and distribution, and negotiation of security parameters. Despites the fact that TLS is currently the most deployed security solution due to its native integration in browsers and web servers, this security protocol presents several limitations. Indeed, TLS was not initially designed as a VPN solution and is therefore vulnerable against traffic analysis, which motivated the design of the new Multiplexing Transport Layer Security (MTLS) protocol [4-5].

The MTLS protocol aims at extending the classical security architecture of TLS, and brings several new security features such as the client-to-site (C2S) and site-to-site (S2S) VPN functionalities, the multiplexing of multiple video, voice and data application flows over single TLS session, application-based access control, secure of both TCP and UDP applications, etc. Thanks to its modular architecture, MTLS puts a particular focus on strong security features, boosted cryptographic performance, interoperability and backwards-compatibility with TLS/DTLS.

The contributions of this paper are many fold. First, the latest version of the MTLS security protocol is presented in details, including the new MTLS handshake sub-protocol and the *Client-to-Site* and *Site-to-Site* MTLS-VPN architectures. Second, a brief comparative study between the MTLS-VPN and existing VPN technologies is provided. Finally, a real-world implementation of the MTLS-VPN protocol is discussed, and a *client-to-site* VPN case study is presented in order to evaluate the performance of MTLS against the well known TLS security protocol.

The remainder of this paper is organized as follows. Section II discusses the MTLS design motivations, the corresponding MTLS protocol specifications and a brief comparison with existing VPN solutions (IPSec, SSL/TLS, OpenVPN and SSH). Section III describes the implementation of MTLS using the OpenSSL library. Section IV provides an experimental performance evaluation of MTLS. Finally, Section V concludes the paper and draws future research directions.

2. THE MULTIPLEXING TRANSPORT LAYER (MTLS) SECURITY PROTOCOL

MTLS is an innovative transport layer security protocol that extends the classical TLS/DTLS protocol architecture and

provides strong security features, security negotiation between multiple communicating entities, protection against traffic analysis and enhanced cryptographic performance. Last but not least, MTLS brings new security features such as Client-to-Site (C2S) and Site-to-Site (S2S) VPN functionalities, Single Sign On (SSO), application-based access control and authorization services. This section describes the main design concepts that are behind the MTLS protocol; presents the MTLS-VPN functionalities and architectures; and finally provides a high level comparison with existing VPN technologies.

2.1 TLS/DTLS Protocols Overview

The TLS protocol is an IETF standard which is based on its predecessor, the Secure Socket Layer 3.0 (SSL) protocol [6], developed by Netscape Communications. TLS provides three main security features to secure communication over TCP/IP: 1) Asymmetric encryption for key exchange and authentication, 2) Symmetric encryption for privacy, and 3) Message authentication code (MAC) for data integrity and authenticity. TLS is able to secure both the connection-oriented Transmission Control Protocol (TCP), as well as the stateless User Datagram Protocol (UDP), through the use of the Datagram Transport Layer Security (DTLS) protocol [7] that was also standardized by the IETF.

As shown in Figure 2, TLS is a modular security protocol that is built around four main sub-protocols:

- The ***TLS-Handshake sub-protocol*** to mutually authenticate the client and server entities, to negotiate the cipher specifications and to exchange the session key.
- The ***Alert sub-protocol*** to handle and transmit alerts messages over the TLS-record sub-protocol.
- The ***Change Cipher Spec (CSS) sub-protocol*** to change the cipher specifications during the communication.
- The ***TLS-Record sub-protocol*** provides the encryption, authentication and integrity security features, and eventually data compression.

Due to its modular philosophy, TLS is completely transparent to high level layers and can therefore secure any application that runs over TCP or UDP. For example, TLS was already adopted to secure applications such as HTTP/HTML over TLS[8], FTP over TLS [9], SMTP over TLS [10], etc.

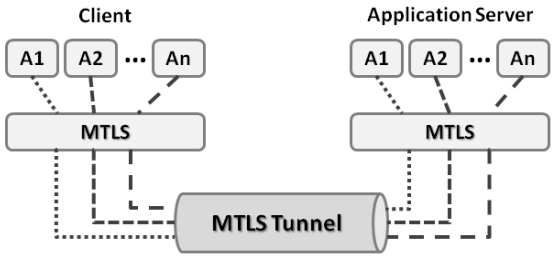


Figure 1. Data multiplexing over a single security session using MTLS.

However, the TLS protocol was not initially designed as a VPN technology, and classical TLS-VPN approaches are generally limited to HTTP-based applications and are the target of various security threats such as traffic analysis [4]. Recently, new TLS-VPN approaches were developed such as OpenVPN [11] which is a free and open source software application for creating OSI layers-2/3 tunnels, however OpenVPN is based on a custom security protocol that exploits SSL/TLS for key exchange.

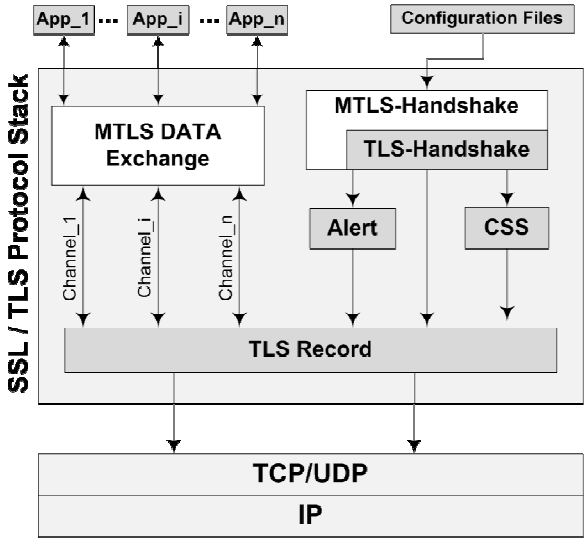


Figure 2. MTLS integration in the TLS/SSL stack.

2.2 MTLS Security Protocol Design Concepts

Due to the limitations of existing TLS-VPN technologies, the Multiplexing Transport Layer Security (MTLS) protocol was designed [4] with a particular focus on strong security features, boosted cryptographic performance, interoperability and backwards-compatibility with TLS/DTLS.

As shown in Figure 1, MTLS is a transport layer security protocol which aims at extending the classical TLS/DTLS protocol architecture by providing each client the capability of negotiating different types of applications (TCP, UDP) over a single security session (TLS). In comparison to TLS/DTLS, MTLS brings several new advantages: protection against traffic analysis and data eavesdropping, bandwidth saving, reduction of the cryptographic overhead, IPSec-VPN like solution at the transport layer, secure VPN solution over unreliable transport protocols (e.g. UDP), etc.

2.2.1 MTLS over (D)TLS

Currently, SSL/TLS represents the de-facto standard to provide end-to-end secure communications over Internet, and due to the increasing deployment of TLS-based security architectures, incompatible changes to the protocol are unlikely to be accepted.

Hence, the MTLS security protocol design focuses on the interoperability and the backwards-compatibility with existing TLS clients and servers.

Thanks to the modularity and versatility of the TLS security architecture, MTLS is integrated into the TLS protocol stack as a new sub-protocol to provide application data multiplexing over single TLS session, as shown in Figure 2. In this way, TLS servers that support MTLS can communicate with TLS clients that do not, and vice versa. The negotiation of MTLS security sessions is done during the classical TLS handshake between clients and servers that are compliant with MTLS.

The operations of the MTLS protocol are described in what follows. For complementary MTLS protocol specifications and message types, readers may refer to [4-5].

2.2.2 TLS and MTLS Handshakes

When a client is willing to use MTLS over a TLS session, it first connects to a MTLS server and sends a classical TLS *ClientHello* message to initiate a TLS Handshake, as shown in Figure 3.

Once the TLS Handshake is complete, a secure TLS session, aka. the MTLS tunnel, is established between the client and the server. The MTLS server sends a *MTLS_LOGIN_REQUEST* message to the client to start the MTLS Handshake procedure whose aims is to authenticate the client and to negotiate the list of applications to be secured over the MTLS tunnel.

To that end, the client transmits its login and password through the secured TLS session to the server. Once the client is successfully authenticated at the MTLS server, it sends an application list request to the server. This list contains all the applications that the client wishes to use over the MTLS tunnel. Then, the server performs an application-based access control, where each requested application is checked against local security policies (e.g. based on user login, permissions, time, bandwidth, etc.). Finally, the server returns the list of authorized applications to the client. If the MTLS client has no prior knowledge about the applications that are available at the MTLS server, it sends an empty application list request to the server which returns the complete list of applications that are granted/authorized to that specific client login.

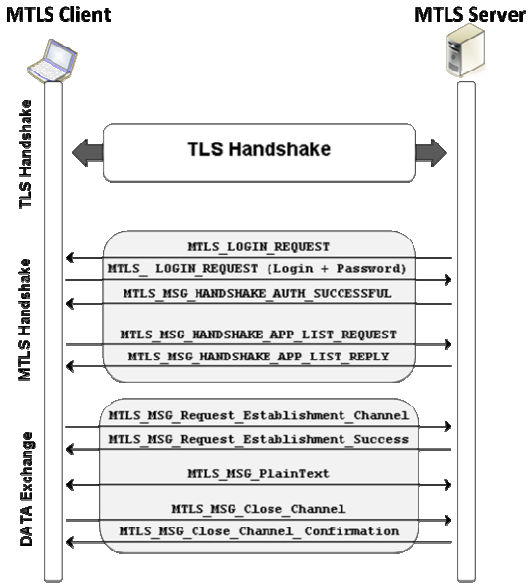


Figure 3. The TLS and MTLS Handshake Protocol.

2.2.3 Opening/Closing Virtual Channels

At the end of the MTLS Handshake procedure, and if at least one application was successfully negotiated with the MTLS server, the client can start communicating with the remote applications by opening/closing virtual channels through the MTLS tunnel. Virtual channels aim at multiplexing multiple application flows through a single MTLS tunnel, as shown in Figure 1. Each virtual channel is characterized by a unique channel identifier and is associated to a particular remote application.

The client can dynamically open many channels to communicate with its approved list of remote applications by sending *OpenChannel* requests. For each received *OpenChannel* request, the MTLS server allocates a send/receive buffer and returns the generated unique channel identifier to the MTLS client. When the client will not send any more data to a given virtual channel, it notifies the MTLS server by sending a Close Channel request.

It should be noted that the MTLS security architecture provides a perfect isolation between virtual channels, such that each MTLS client can only communicate with its list of approved applications through dedicated virtual channels.

2.2.4 MTLS Flow Control

Once virtual channels are opened inside the MTLS tunnel, the MTLS client can start communicating with the remote applications through local MTLS ports. Hence, for each opened virtual channel, the MTLS client open a local TCP or UDP port, aka. the local MTLS port, on the loopback network interface. This local port aims at receiving data from high level client applications (e.g. FTP, HTTP, DNS, IMAP, etc.) and delivering them to the client MTLS layer.

The received data from high level client applications are encapsulated by the MTLS layer (i.e. the MTLS DATA Exchange layer in Figure 2) into *MTLS_MSG_PlainText* messages with the corresponding virtual channels identifiers, and are sent to the client (D)TLS Record layer, as shown in Figure 2.

The (D)TLS Record layer receives data from the MTLS layer, considers it as uninterpreted data and applies the fragmentation and the cryptographic operations as defined in [1]. The corresponding *TLSPlainText* messages are then sent through the MTLS tunnel towards the MTLS server.

At the MTLS server side, received data at the (D)TLS Record layer are decrypted, verified, decompressed and reassembled, then delivered (i.e. the *MTLS_MSG_PlainText* messages) to the MTLS layer. Finally, the MTLS layer sends the data to the remote application servers using the corresponding channel identifiers. The reverse communication flow from the application servers (or MTLS server) to the client high level applications (or MTLS client) undergoes the same process.

It should be noted that all MTLS messages that are sent through the MTLS tunnel are first encapsulated into *TLSPlainText* messages [1] by the (D)TLS Record layer.

2.3 MTLS-VPN Architecture

The design of the MTLS protocol was mainly driven by the limitations of current TLS-VPN technologies. In this context, MTLS aims at extending the classical TLS/DTLS protocol architecture by providing new security features such as the protection against traffic analysis, application based access control and VPN functionalities. This section describes the two main VPN architectures that are currently supported by the MTLS security protocol.

The first VPN architecture that is supported by the MTLS protocol is the client-to-site (C2S) scheme. The C2S scheme aims at providing the users a secure access over a public network to the corporate network and resources (e.g. emails, applications, remote desktops, servers, files, etc.), as shown in Figure 4. Since MTLS is interoperable and backwards-compatible with TLS/SSL, existing TLS/SSL based clients can also connect to the MTLS server.

The second VPN architecture that is supported by the MTLS protocol is the site-to-site (S2S) scheme. As depicted in Figure 4, the S2S scheme aims at securely interconnecting multiple

company/office locations between each other over a public network, such as Internet. The main objective is to extend the company's network and to share the corporate resources with other locations/offices.

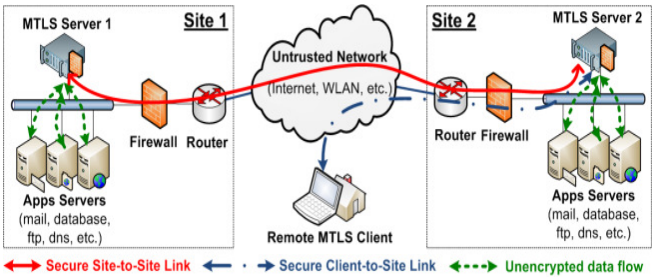


Figure 4. Client-to-Site and Site-to-Site VPN architectures using MTLS.

2.4 MTLS versus Existing VPN Solutions

This section briefly compare and discuss the newly designed MTLS-VPN security protocol against existing VPN technologies, i.e. SSL/TLS-VPN, IPSec-VPN, OpenVPN and SSH.

MTLS-VPN versus TLS/SSL-VPN. As already discussed above, MTLS brings several new advantages in comparison to the basic TLS/SSL-VPN. Among them are the protection against traffic analysis and data eavesdropping, bandwidth saving, reduction of the cryptographic overhead, integrated *client-to-site* and *site-to-site* VPN functionalities, granular application level access control, etc.

MTLS-VPN versus IPSec-VPN. Similarly to IPSec-VPN, MTLS provides *site-to-site* VPN connection and can secure both TCP and UDP traffic. However, MTLS's design philosophy brings several advantages in comparison to IPSec, such as the support of NAT traversal, implementation at the application level, integrated application access control and filtering, compliance with existing SSL/TLS and MTLS solutions, allowing thus easier VPN deployment, configuration and administration.

MTLS-VPN versus OpenVPN. As already discussed above, MTLS brings several new advantages in comparison to OpenVPN, in particular, the interoperability and the backwards compatibility with TLS/SSL solutions. Moreover, MTLS is based on standardized and well known/deployed protocols, whereas OpenVPN is partly based on a proprietary protocol.

MTLS-VPN versus SSH. Unlike SSH, MTLS is able to secure UDP traffic and provides a complete and transparent VPN solution for secure remote access to servers and applications. Moreover, configuring, administering and securing multiple applications through MTLS is much easier than with SSH.

Despites all these advantages, MTLS presents three main limitations.

First, since MTLS is a transport layer security protocol, it cannot handle layer-2/3 tunnels and protocols, such as the ICMP and IP protocols.

Second, depending on the deployment scenario, MTLS may require the presence of a Public Key Infrastructure (PKI) for the management of users and servers certificates.

Finally, MTLS in its current version supports only point-to-point (P2P) secured TCP or UDP connections.

3. MTLS SECURITY PROTOCOL IMPLEMENTATION

MTLS was implemented based on the popular *OpenSSL* library [12]. *OpenSSL* is an open-source implementation of the Secure Sockets Layer (SSL v2/v3) and the Transport Layer Security (TLS v1) protocols, and is currently used by many quality products (e.g. Apache) and companies.

MTLS was implemented as a new sub-protocol running on top of the TLS/DTLS architecture (cf. Figure 2) and required about 30500 lines of additional C code to the *OpenSSL* library. The current MTLS implementation supports: client-to-site and site-to-site VPN connections, advanced user authentication features (i.e. LDAP, Active Directory, etc.), application-level access control, application-oriented quality-of-service (QOS) through tx/rx buffer sizing, and the support of multiple operating systems for the MTLS Client, including Linux, Windows and MacOS. Regarding the MTLS Server side, the current implementation runs on the Linux 2.6 kernel and is based on the latest stable release of OpenSSL (version 1.0.0g).

Due to its modular design philosophy, MTLS highly depends on the *OpenSSL* security features, which were proven to be highly stable and are currently used in several production servers. Moreover, unlike IPSec, MTLS runs in the OS user space instead of the kernel space, and is based on well defined APIs, thus making it compliant with the majority of operating systems and interoperable with existing SSL/TLS implementations.

4. TESTBED SETUP AND EXPERIMENTAL RESULTS

For the sake of the performance evaluation of MTLS and its comparison with existing security protocols, we designed a real-world network test-bed using two computers (64bits architecture, 2Ghz processor and 4GB memory) running the Linux Debian operating system (kernel version 2.6.32) and connected to a 100Mbps wired network. The main objective of this test-bed is to provide a detailed comparative analysis between MTLS and TLS using a *Client-to-Site* VPN architecture case study.

An Apache web server is setup on the first computer as well as an MTLS-VPN server. The web server is configured to support both HTTP (TCP port 80) and HTTPS (TCP port 443) connections, while the MTLS server (TCP port 999) is configured to provide a secure access to the local web server (HTTP) using the MTLS protocol.

The performance evaluation is performed at the second computer using the Apache server benchmarking tool in order to analysis the performance of HTTP flows: 1) *HTTP over SSL/TLS* (HTTPS); and 2) *HTTP over MTLS*. To that end the Apache web server and the MTLS server were both configured with the following SSL/TLS parameters: *OpenSSL 1.0.0g*, *TLSv1 protocol*, *DHE-RSA-AES256-SHA* cipher suite and same server/CA certificates.

Three performance metrics are evaluated: the *requests rate*, the *transfer rates*, and the *time per request*, as a function of the number of concurrent HTTP requests (from 1 to 35) and the size of HTTP flows (from 2K to 100K). The results are averaged over 100 runs.

Figure 5 shows the obtained average HTTP requests rates over TLS and MTLS tunnels. We observe that for a low number of concurrent requests the two protocols provide quite similar

performance behavior (around 20 requests/sec). However, at higher requests frequency, MTLS outperforms TLS thanks to the data multiplexing technique which reduces the bandwidth and cryptographic overheads. Hence, using MTLS, HTTP requests processing at the server side is 2 to 10 times faster than with TLS, which provides a performance of 35 requests/sec regardless of the size of HTTP flows and the number of concurrent requests. The resulting average time per HTTP request is depicted in Figure 6. Again, we observe that MTLS alleviates the cost of HTTP requests processing at the server side, making it an interesting alternative to SSL/TLS to secure large-scale and distributed TCP/UDP applications.

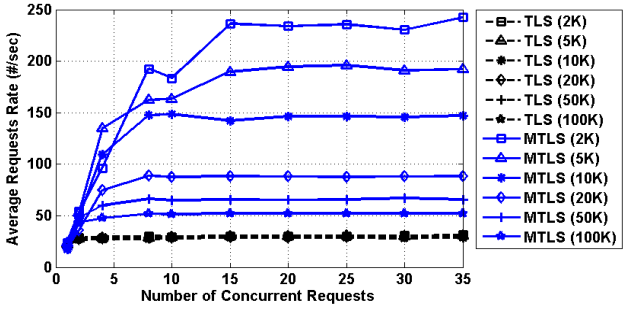


Figure 5. MTLS versus TLS: Average Requests Rate.

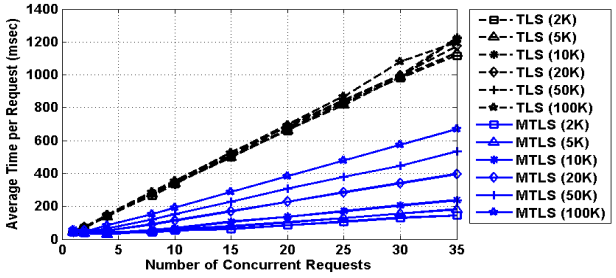


Figure 6. MTLS versus TLS: Average Time per Request.

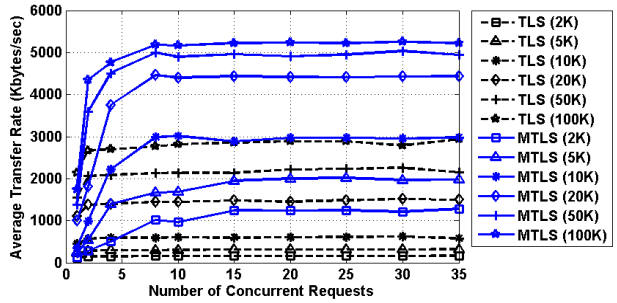


Figure 7. MTLS versus TLS: Average Transfer Rate.

Finally, the achieved average transfer rates are illustrated in Figure 7. As expected, at low number of concurrent requests, SSL/TLS slightly outperforms MTLS due the extra overhead introduced by the MTLS handshake and the management of virtual channels. However, when the number of concurrent requests increase, MTLS provides better transfer rates in comparison to SSL/TLS. As an example, considering a 10K HTTP flow and 35 concurrent requests, the achieved transfer rates are 581Kbytes/sec and 2980Kbytes/sec for SSL/TLS and MTLS, respectively. In this case, the resulting MTLS average transfer gain is around 512% in comparison to SSL/TLS. Indeed, the MTLS data multiplexing technique optimizes the bandwidth usage by reducing the overhead of cryptographic operations and handshaking, thus making it a worthy solution, for example, to secure low-cost and low-bandwidth wireless networks. It should be noted that all these results were found to hold for different types of cipher suites, data traffic and SSL/TLS protocol versions.

5. CONCLUSIONS

This paper described the MTLS security protocol which aims at providing a complete VPN solution at the transport layer. MTLS extends the classical SSL/TLS protocol architecture by enabling the multiplexing of multiple application flows over single established tunnel. The experimental results show that MTLS outperforms SSL/TLS in terms of achieved transfer rates and requests processing times, thus making it an interesting solution to secure large-scale and distributed applications. Future research directions will focus on extending MTLS to support multi-hop VPN connections and its integration in Cloud environments. Finally, experimental evaluation and comparison with IPSec and OpenVPN will also be investigated in more details.

6. ACKNOWLEDGMENTS

This work is partly supported by the Systematic Paris Region Systems and ICT Cluster under the **OnDemand** FEDER5 research project.

7. REFERENCES

[1] Dierks, T., Rescorla, E. Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246. Retrieved from http://tools.ietf.org/html/rfc5246. 2008.

[2] Kent, S., Atkinson, R. Security Architecture for the Internet Protocol. RFC 2401. Retrieved from http://www.ietf.org/rfc/rfc2401.txt. 1998.

[3] Bellovin, S. Guidelines for Mandating the Use of IPSec Version 2. IETF DRAFT. Retrieved from http://tools.ietf.org/html/draft-bellovin-useipsec-07. 2007.

[4] Badra, M., Hajjeh, I. Enabling VPN and Secure Remote Access using TLS Protocol. In WiMob'2006, the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. 2006.

[5] Badra, M., Hajjeh, I. MTLS: (D)TLS Multiplexing. IETF Draft. Retrieved from http://tools.ietf.org/html/draft-badra-hajjeh-mtls-06. 2011.

[6] Freier, A., Karlton, P., Kocher, P. The Secure Sockets Layer (SSL) Protocol Version 3.0. RFC 6101. Retrieved from http://tools.ietf.org/html/rfc6101. 2011.

[7] Rescorla, E., Modadugu, N. Datagram Transport Layer Security Version 1.2. RFC 6347. Retrieved from http://tools.ietf.org/html/rfc6347. 2012.

[8] Rescorla, E. HTTP over TLS. RFC 2818. Retrieved from http://www.ietf.org/rfc/rfc2818.txt. 2000.

[9] Ford-Hutchinson, P. Securing FTP with TLS. RFC 4217. Retrieved from http://www.ietf.org/rfc/rfc4217.txt. 2005.

[10] Hoffman, P. SMTP Service Extension for Secure SMTP over TLS. RFC 2487. Retrieved from http://www.ietf.org/rfc/rfc2487.txt. 1999.

[11] Feilner, M. OpenVPN: Building and Integrating Virtual Private Networks. Packt Publishing. 2006.

[12] The OpenSSL project, viewed March, Retrieved from http://www.openssl.org, 2012.

Optimization of Quality of Experience through File Duplication in Video Sharing Servers

Emad Abd-Elrahman, Tarek Rekik and Hossam Afifi
Wireless Networks and Multimedia Services Department
Institut Mines-Télécom, Télécom SudParis - CNRS Samovar UMR 5157, France.
9, rue Charles Fourier, 91011 Evry Cedex, France.
{emad.abd_elrahman, tarek.rekik, hossam.afifi}@it-sudparis.eu

ABSTRACT
Consumers of short videos on Internet can have a bad Quality of Experience (QoE) due to the long distance between the consumers and the servers that hosting the videos. We propose an optimization of the file allocation in telecommunication operators' content sharing servers to improve the QoE through files duplication, thus bringing the files closer to the consumers. This optimization allows the network operator to set the level of QoE and to have control over the users' access cost by setting a number of parameters. Two optimization methods are given and are followed by a comparison of their efficiency. Also, the hosting costs versus the gain of optimization are analytically discussed.

Categories and Subject Descriptors
C.2.1 [COMPUTER-COMMUNICATION NETWORKS]: Network Architecture and Design.

General Terms
Algorithms, Performance.

Keywords
Server Hits; File Allocation; File Duplication; Optimization; Access Cost.

1. INTRODUCTION
The exponential growth in number of users access the videos over Internet affects negatively on the quality of accessing. Especially, the consumers of short videos on Internet can perceive a bad quality of streaming due to the distance between the consumer and the server hosting the video. The shared content can be the property of web companies such as Google (YouTube) or telecommunication operators such as Orange (Orange Video Party). It can also be stored in a Content Delivery Network (CDN) owned by an operator (Orange) caching content from YouTube or DailyMotion. The first case is not interesting because the content provider does not have control over the network, while it does in the last two cases, allowing the network operator to set a level of QoE while controlling the network operational costs.

Quality of Experience (QoE) is a subjective measure of a customer's experiences with his operator. It is related to but differs from Quality of Service (QoS), which attempts to objectively measure the service delivered by the service provider. Although QoE is perceived as subjective, it is the only measure that counts for customers of a service. Being able to measure it in a controlled manner helps operators understand what may be wrong with their services to change.

There are several elements in the video preparation and delivery chain, some of them may introduce distortion. This causes the degradation of the content and several elements in this chain can be considered as "QoE relevant" for video services. These are the encoding systems, transport networks, access networks, home networks and end devices. We will focus on the transport networks behavior in this paper.

In this work, we propose two ways of optimization in file duplication either *caching* or *fetching* the video files. Those file duplication functions were tested by feeding some YouTube files that pre-allocated by the optimization algorithm proposed in the previous work [1].

By caching, we mean duplicate a copy of the file in different or another place than its original one. While, in fetching, we mean retrieve the video to another place or zone in order to satisfy instant needs either relevant to many requests or cost issues from the operators point of view.

The file distribution problem has been discussed in many works. Especially the multimedia networks, the work in [2] handled the multimedia file allocation problem in terms of cost affected by network delay. A good study and traffic analysis for inter-domain between providers or through social networks like YouTube and Google access has been conducted in [3]. In that work, there is high indication about the inter-domain traffic that comes from the CDN which implies us to think in optimizing files allocations and files duplications.

Akamai [4] is one of the more famous media delivery and caching solution in CDN. They proposed many solutions for media streaming delivery and enhancing the bandwidth optimization for video sharing systems.

The rest of this paper is organized as follows: Section 2 presents the stat-of-the-art relevant to video caching. Section 3 highlights allocation of files based on the number of hits or requests to any file. In section 4, we propose the optimization by two file duplication mechanisms either caching or fetching. Numerical results and threshold propositions are introduced in Section 5. The hosting issues and gain from this optimization are handled by Section 6. Finally, this work is concluded in Section 7.

2. STAT-OF-THE-ART
Caching algorithms were mainly used to solve the problems of performance issues in computing systems. Their main objective was to enhance computing speed. After that, with the new era of multimedia access, the term caching used to ameliorate the accessing way by caching the more hits videos near to the clients. With this aspect, the Content Delivery Network CDN appeared to manage such types of video accessing and enhance the overall performance of service delivery by offering the content close to the consumer.

For the VOD caching, the performance is very important as a direct reflection to bandwidth optimization. In [5], they considered the caching of titles belonging to different video services in IPTV network. Each service was characterized by the number of titles, size of title and distribution of titles by popularity (within service) and average traffic generated by subscribers' requests for this service. The main goal of caching

was to reduce network cost by serving maximum (in terms of bandwidth) amount of subscribers' requests from the cache. Moreover, they introduced the concept of "cacheability" that allows measuring the relative benefits of caching titles from different video services. Based on this aspect, they proposed a fast method of partition a cache optimally between objects of different services in terms of video files length. In the implementation phase of this algorithm, different levels of caching were proposed so as to optimize the minimum cost of caching.

The work presented in [6] had considered different locations of doing caching in IPTV networks. They classified the possible locations for caching to three places (STB, aggregated network (DSLAMs) or Service Routers SRs) according to their levels in the end-to-end IPTV delivery. Their caching algorithm takes the decisions based on the users' requests or the number of hits per time period. Caching considers different levels of caching scenarios. But, this work considered only the scenarios where caches are installed in only a single level in the distribution tree. In the first scenario each STB hosts a cache and the network does not have any caching facilities while, in the second and the third scenarios the STBs do not have any cache, but in the former all caching space resides in the DSLAMs, while in the latter the SRs have the only caching facility. The scenario where each node (SR, DSLAM and STB) hosts a cache and where these caches cooperate was not studied in that work. Actually, the last scenario could be complicated in the overall management of the topology. Finally, their caching algorithm was considered based on the relation between Hit Ratio (HR) and flows Update Ratio (UR) in a specific period of time.

The same authors presented a performance study about the caching strategies in on-demand IPTV services [7]. They proposed an intelligent caching algorithm based on the object popularity in context manner. In their proposal, a generic user demand model was introduced to describe volatility of objects. They also tuned the parameters of their model for a video-on-demand service and a catch-up TV service based on publicly available data. To do so, they introduced a caching algorithm that tracks the popularity of the objects (files) based on the observed user requests. Also, the optimal parameters for this caching algorithm in terms of capacity required for certain hit rate were derived by heuristic way to meet the overall demands.

Another VOD caching and placement algorithm was introduced in [8]. They used heuristic ways based on the file history for a certain period of time (like one week) to guide the future history and usage of this file. Also, the decisions were made based on the estimations of requests of specific video to a specific number as a threshold value. The new placements are considered based on the frequency of demands that could update the system rates or estimated requests. Finally, their files distributions depend on the time chart of users' activities during a period of time (for example one week) and its affections on files caching according to their habits. To test this algorithm, they used traces from operational VOD system and they had good results over Least Recently Used (LRU) or Least Frequently Used (LFU) algorithms in terms of link bandwidth for caching emplacements policies. So, this approach is considered as a rapid placement technique for content replications in VOD systems.

An analytical model for studying problems in VOD caching systems was proposed in [9]. They presented hierarchical cache optimization (i.e. the different levels of IPTV network). This model depends on several basic parameters like: traffic volume, cache hit rate as a function of memory size, topology structure like DSLAMs and SRs and the cost parameters. Moreover, the optimal solution was proposed based on some assumptions about the hit rates and network infrastructure costs. For example, the hit rate is a function of memory used in cache and there is a threshold cost per unit of aggregation networks points like DSLAMs. Also, they demonstrated different scenarios for optimal cache configuration to decide at which layer or level of topology.

A different analysis introduced in [10] about data popularity and affections in videos distributions and caching. Through that study, they presented an extensive data-driven analysis on the popularity distribution, popularity evolution, and content duplication of user-generated (UG) video contents. Under this popularity evolution, they proposed three types of caching:

- *Static caching:* at starting point of cache, handling only long-term popularity
- *Dynamic caching:* at starting point of cache, handling the previous popularities and the requests coming after the starting point in the period of trace
- *Hybrid caching:* same as static cache but with adding the most popular videos in a day

By simulation, the hybrid cache improved the cache efficiency by 10% over static one. Finally, this work gave complete study about popularity distribution and its correlations to files allocations or caching.

We will now start by analyzing the optimization of the file allocation introduced in [1] in order to reduce the total access cost, taking into account the number of hits on the files. Then, we will analyze the file duplication algorithms.

3. OAPTIMIZATION OF FILES BEST LOCATION

In order to reduce the total access cost, we first move files so as to have every one of them located in the node from where the demand is the highest, i.e. where it has the greatest number of hits. We can follow the steps of the algorithm in Table 1. The main objective from this algorithm is to define the best allocation zone i servers of file f uploaded from any geographical zone i where (i=1:N) zones.

Table 1. Best location algorithm

```
Input N          //number of zones
Input M          //number of servers
Input h_ij       //number of hits from zone i to server j for a specific file
Input d_ij       //the assumed cost between zone i and server j
     Total cost C_i = ∑ h_ij * d_ij        // where i from 1 to N and j from 1 to M
                If C_i < C_i+1,..,C_N
                        Then allocate this file in zone i
                        Else move this to min (C_i+1,...,C_N) value location
                End If
Return the best location for this file      //best j
End
```

We applied that algorithm on six YouTube files [1] chosen from different zones as shown in (Table 2). The numbers are rounded for better clarity. Those files were preselected as the most hit files from different zones analyzed briefly in [1].

Table 2. Examples of files chosen from YouTube and related hits

	Hits					
	File 1	**File 2**	**File 3**	**File 4**	**File 5**	**File 6**
Zone 1	135000	16000	55000	8700	3800	42000
Zone 2	40000	5000	175000	2700	1000	15000
Zone 3	132000	51000	55000	8800	3900	43000
Zone 4	100000	12000	40000	6400	2800	138000
Zone 5	12000	1000	5000	800	12400	4000
Zone 6	17000	1000	5000	900	400	5000

We reach to the following results of files' distribution or best allocation:

The best location of File 1 is the servers in zone 1 where the algorithm gave the minimum cost; File 2 moves from zone 2 to zone 3; File 3 moves from zone 3 to zone 2; File 4 moves from zone 4 to zone 3: File 5 stays in zone 5; File 6 moves to zone 4.

The new files distribution is shown in Figure 1.B and will be the distribution that we'll depend on for the next optimization (Duplication).

Since geography plays an important role in delivering video to customers, we propose the following network representation shown in Figure 1. (A). We divide the network into six zones each zone represented by a node and we select a file uploaded from each zone to be studied by our algorithms.

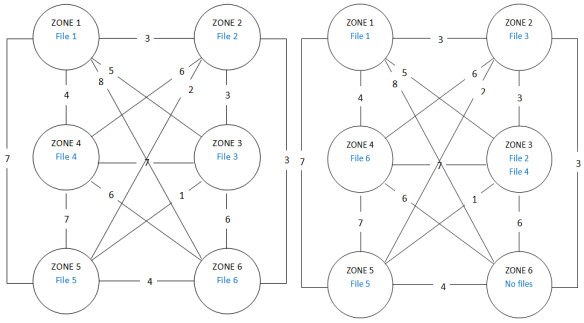


Figure 1. (A) Network Representation (B) File distribution after the application of the best location algorithm

The nodes in this figure represent the servers of a geographical zone; it can be a city for a national operator. A node can also be assimilated to the edge router connecting the customers who live in that geographical zone to the operator's core network or (CDN). The arcs connecting the nodes are physical links, and the numbers on them represent an access cost, i.e. the cost of delivering a video from zone A to a consumer from zone B. That access cost (is assumed to be symmetric) can be a combination of many parameters such as the path's length, the hop count (from edge router to edge router), the cost of using the link (optical fiber, copper, regenerators and repeaters, link rental cost) and the available bandwidth. The files of a given zone are hosted by the servers of that zone.

4. OPTIMIZATION THROUGH FILE DUPLICATIONS

To grant a given QoE, we choose to give access to popular file (located in zone **j** servers and having a number of hits higher than a given value **Y**) for a given consumer (from zone **i**) only if this consumer can access that file with a cost $\mathbf{a_{ij}}$ lower than a given value **A** ($\mathbf{a_{ij}} =< \mathbf{A}$). If it is the case, we check the demand on that same file from the zone **i** of the customer (which is $\mathbf{y_{ij}}^{\mathbf{f}}$). If the demand is higher than a threshold **Y**, we duplicate file **f**. Then, any consumer can access any content with good quality and minimum cost. The steps of file duplication algorithm are shown in Table 3.

Thus, we make a compromise between access cost and storage cost. Actually, the higher the value of **Y**, the higher the access cost and the lower the need for storage capacity. Also, the higher the value of **A**, the higher the access cost and the lower the need for storage capacity. The values of **A** and **Y** are thresholds being adjusted by the operators (Content Delivery).

We are going to experiment two different duplication functions:

- ***duplicate1***(f,i,j) duplicates file ***f*** (located in zone j) in zone i which means ***(caching)*** Table 4.
- ***dupliacate2***(f,i,j) duplicates file ***f*** in the zone **k** that has the second highest number of hits on that file (located in zone j) and that grants an access cost lower than **A** at the same time, zone j being the zone with the first highest number of hits (the best location). This means ***(fetching)*** Table 5.

We have to make sure that the duplicated files are not taken into consideration while running the algorithm in Table 3, i.e. in line "For f from 1 to $\mathbf{p_j}$" ***f*** can't be a duplicated file.

We apply the duplication algorithm (Table 3) to the 6 files, starting from the file distribution on Figure 1.B where every file is located in its best location.

We are going to look at different values of Y (0, 5000, 10000 and 20000) and compare the efficiency of the 2 duplicating methods for each value of Y. Then, we will also see the effect of changing Y value on the total gain associated with the optimization. For all cases of Y, we suppose A=5 unit cost.

Table 3. Duplication algorithm

```
Input N        // number of zones
Input A        // maximum access cost, set by the network operator
Input Y        // minimum number of hits allowing file duplication,
               //set by the network operator

Input a_ij     // access cost to a file located in zone j from a customer located in zone i
Input p_j      // number of files in zone j
Input y_ij^f   // number of hits on file f (located in zone j) from zone i users

For i from 1 to N                                   // user's zone
    For j from 1 to N                               // file's zone
        If a_ij > A                                 // we suppose that a_ii=0
            Then For f from 1 to p_j
                If y_ij^f > Y
                    Then duplicate(f,i,j)           // duplicates file f
                                                    //in the best location
                End If
        End If
End
```

Table 4. Function'duplicate1'

```
Function duplicate1(f,i,j)
    Duplicate file f in zone i
End
```

Table 5. Function 'duplicate2'

```
Function duplicate2(f,i,j)

    Input k=j                   // zone k is the best location to duplicate file f
    While a_ik > A
        k:=retrieve(f,i)
    Duplicate file f in zone k  // we suppose that there's at least one zone k
                                // that ensures a_ik =< A
End
```

5. METHODS AND NUMERICAL RESULTS

This section focuses on performance investigation of the proposed methods for files duplication by applying these two functions on different files from YouTube. Moreover, we will test different use cases by changing the threshold value Y that supposes to be adjusted by the operator as follows:

5.1 First Case Y=0

We get the following file distribution with ***'duplicate1'***. Assume **'o'** indicates the file's best location, while **'x'** indicates the file has been duplicated as in Table 6.

Let's now compute the gain achieved through the first duplication method, i.e. the difference between the access costs before and after the application of the Duplication algorithm, $\boldsymbol{a_{ij}^f}$

being the access cost to File f located in zone j from a consumer located in zone i.

Table 6. File distribution with Y=0 & caching

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
File 1	o				x	x
File 2			o	x		x
File 3		o		x		
File 4			o	x		x
File 5	x			x	o	
File 6		x	x	o	x	x

Example: Gain from the duplication of File 5 in zone 1

- File 5 is no longer delivered to zone 1 consumers from zone 5 but from zone 1.

Gain from zone 1 = $(a_{15} - a_{11}) * y_{15} f_5 = (7\text{-}0)*3800=26600$

- Consumers from zone 2 still access File 5 that is located in zone 5 because it is the zone with the cheapest access cost (a_{25}=2) compared with the other zones that host File 5 (zones 1 and 4 for which the access costs from zone 2 are respectively 3 and 6). Gain from zone 2 = 0
- Consumers from zone 3 still access File 5 that is located in zone 5 because it is the zone with the cheapest access cost (a_{35}=1) compared with the other zones that host File 5 (zones 1 and 4 for which the access costs from zone 3 are respectively 5 and 7). Gain from zone 3 = 0
- Consumers from zone 4 access File 5 that is located in zone 4 because it is the zone with the cheapest access cost (a_{44}=0). Gain from zone 4 = 0
- Consumers from zone 5 access File 5 that is located in zone 5 because it is the zone with the cheapest access cost (a_{55}=0). Gain from zone 5 = 0
- Consumers from zone 6 still access File 5 that is located in zone 5 because it is the zone with the cheapest access cost (a_{65}=4) compared with the other zones that host File 5 (zones 1 and 4 for which the access costs from zone 6 are respectively 8 and 6). Gain from zone 6 = 0

We repeat the same operation for the other duplicated files and we get the following file distribution with the fetching method ***(duplicat2)***. This distribution is shown in Table 7 with **'o'** indicates the file's best location, while **'x'** indicates the file has been duplicated.

Table 7. File distribution with Y=0 & fetching

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
File 1	o	x	x			
File 2	x	x	o			
File 3	x	o				
File 4	x	x	o			
File 5	x		x		o	
File 6		x	x	o		

We notice that with caching, the majority of the duplicated files are distributed among zones 4, 5 and 6, while with fetching zones 1, 2 and 3 contain all the duplicated files.

After computing the new access costs, we compare the gain achieved through the duplicating methods in the figures below (Figure 3).

If we look at File 1, we notice that the sum of the access costs from all the zones after caching (the sum of the red columns) is higher than that after fetching (the sum of the green columns). Therefore, there's a greater gain with fetching than with caching to access File 1 from all over the network.

After computing the new access costs, we compare the gain achieved through the duplicating methods in the figures below (see Figure 2).

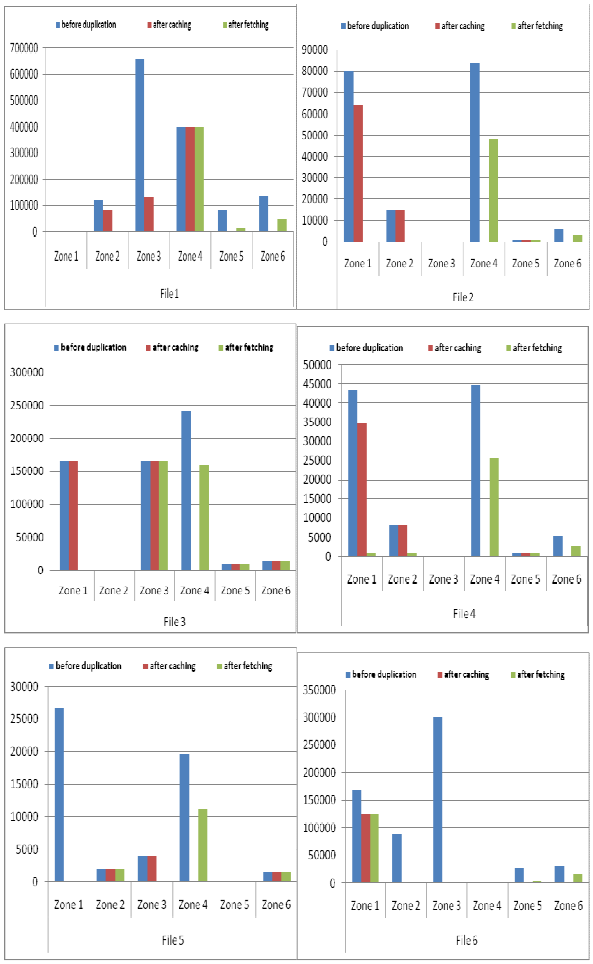


Figure 2. Access cost before and after *duplication1* and *duplication2* for Y=0

For the delivery of File 1 to zone 4, no gain was achieved. That means zone 4 customers still access File 1 from the same zone (zone 1 here). In other cases, there may be a duplication of that file on another zone that has the same access cost to zone 4 as zone 1 (the best location of File 1).

We notice that fetching is more efficient than caching for Files 1, 2, 3 and 4 and zones 1, 2 and 3, while caching is better for Files 5 and 6 and zones 4, 5 and 6.

5.2 Second Case Y=5000

For this threshold, we notice that with caching algorithm, the majority of the duplicated files are distributed among zones 4, 5 and 6, while with fetching algorithm, zones 1, 2 and 3 contain all the duplicated files, like for Y=0.

In Figure 3 below, we only show the files that were duplicated by any of the 2 duplication methods.

We notice that fetching is more efficient than caching for Files 1, 2, 3 and 4 and zones 1, 2 and 3 (except for File 6), while caching is better for File 6 and for zones 4, 5, 6, even if the two methods are of equal efficiency on certain zones.

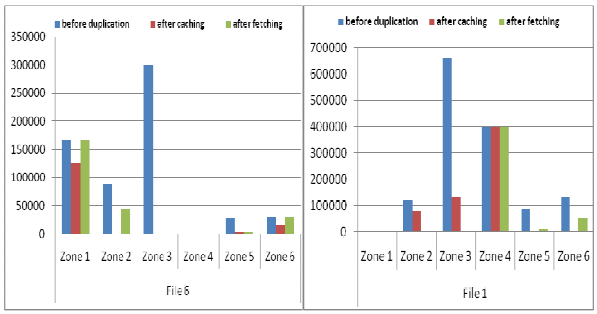


Figure 3. Access cost before and after caching and fetching for Y=5000

We also notice that File 5 was not duplicated because there is no enough demand for it. The only zone from which the demand exceeds the demand threshold Y (5000) is zone 5 (which is the best location of File 5) with 12400 hits (see Table 2).

5.3 Third Case Y=10000

After applying the two methods, we get the following files distribution either for caching or for fetching:

Like for Y=0 and Y=5000, we notice that with caching, the majority of the duplicated files are distributed among zones 4, 5 and 6, while with fetching zones 1, 2 and 3 contain all the duplicated files.

This is understandable for fetching as the highest demands are those of Files 1, 2 and 3, and these files are located in zones 1, 3 and 2 respectively. It is also understandable for caching as the access cost from zones 4, 5 and 6 to Files 1, 2 and 3 is high compared to access costs between zones 1, 2 and 3:

- Zone 4 customers access File 3 with an access cost of 6, and File 2 with 7;
- Zone 5 customers access File 1 with an access cost of 7;
- Zone 6 customers access File 1 with an access cost of 8 and File 2 with 6.

In Figure 4, we only show some files that were duplicated by any of the 2 duplication methods which are files 3 and 6.

We notice that fetching is more efficient than caching for Files 1, 2 and 3 and zones 1, 2 and 3 (except for File 6), while caching is better for File 6 and for zones 4, 5, 6, even if the two methods are of equal efficiency on certain zones.

We also notice neither File 4 nor File 5 was duplicated because there's not enough demand for them. The highest demand for File 4 is 8800 (as shown in Table 2) and doesn't exceed Y (10000), and the only zone from which the demand exceeds Y for File 5 is zone 5 (which is the best location of File 5) with 12400 hits (see Table 2).

5.4 Forth case Y=20000

We get the following file distribution for both file duplication methods as shown below.

In the Figure 5, we only show sample of files that were duplicated by any of the two duplication methods which are files 3 and 6.

We notice that fetching is more efficient than caching for File 3 and zones 1 and 2, while it has the same efficiency as caching on File 6 and zones 3, 5 and 6. Caching is more efficient than fetching only on zone 4.

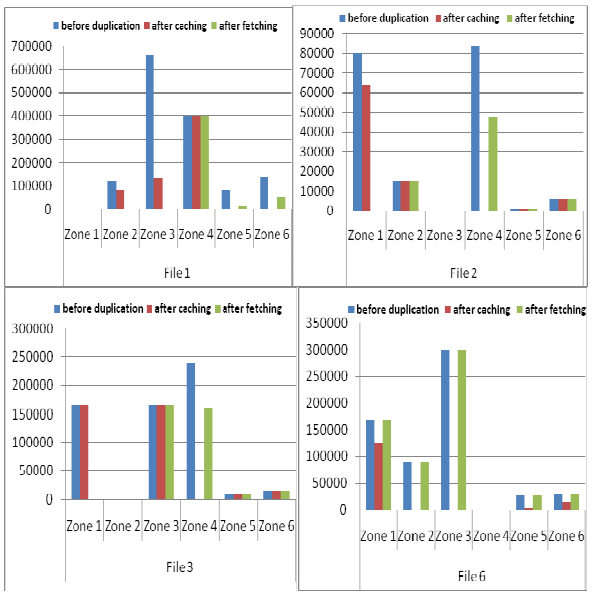


Figure 4. Access cost before and after caching and fetching for Y=10000 to files 1,2,3 and 6

Here, despite the fact that there is a demand higher than Y (20000) on File 1 from zones other than its best location (zone 1), such as zone 3 (132000 hits), zone 4 (100000 hits) or zone 2 (40000 hits), File 1 was not duplicated. This is due to the access cost threshold A for zones 2, 3 and 4, and due to the demand threshold Y for zones 5 and 6 as there is not enough demand on File 1 from these 2 zones (12000 and 17000 respectively).

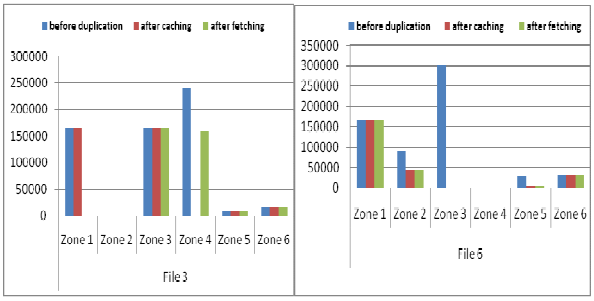


Figure 5. Access cost before and after caching and fetching for Y=20000 to files 3 and 6

6. HOSTING COST AND NET GAIN

In this part, we take into account the file duplication cost. This cost is mainly due to file hosting. We assume that:

- A hit cost is 0.01 m.u. (money unit). Then, we multiply the access costs by 100 to have all the costs expressed in m.u. (just to scale the values) ;
- 1 TB (unit size in Bytes) hosting cost is 20 m.u. ;
- The files are sets of files, each set is 100 TB; in fact, the hosting servers contain a great number of videos, and the network operator may duplicate sets of files instead of running the Duplication algorithm for every single file. These sets may encompass files that have almost the same viewers (all the episodes of a TV show for example).

The hosting cost = number of duplicated files * file size in TB * 1 TB hosting cost.

Table 8. Number of duplicated files with both duplication methods

Y	0	5000	10000	20000
Number of duplicated files through caching	13	7	6	2
Number of duplicated files through fetching	11	5	4	2

We notice that the number of duplicated files decreases if Y increases. This is due to the fact that only the most popular files are duplicated with a high value of Y as shown in Table 8.

Now if we compare the gain in access cost and the hosting cost (a shown Figure 6), we notice that there is a financial loss for Y=0 with both duplication methods. This is understandable as duplicating files that are not popular enough require important hosting resources and benefit to a small number of customers. Hence the need to compute the Net gain, which is the difference between the gain in access cost and the hosting cost, in order to properly assess the efficiency of both duplication methods.

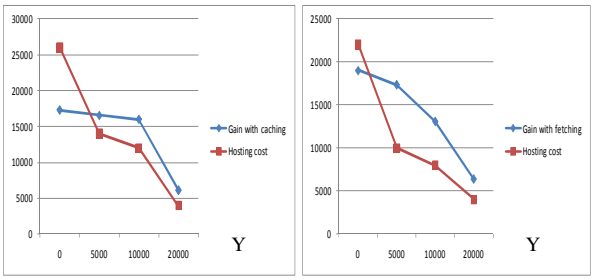


Figure 6. Gain vs. hosting cost

With caching, there is a gain starting from Y≈4000 and the biggest Net gain is achieved for Y=10000 according to Figure 8. While with fetching, there is a gain starting from Y≈2000 and the biggest net gain is achieved for Y=5000 according to the same Figure 7.

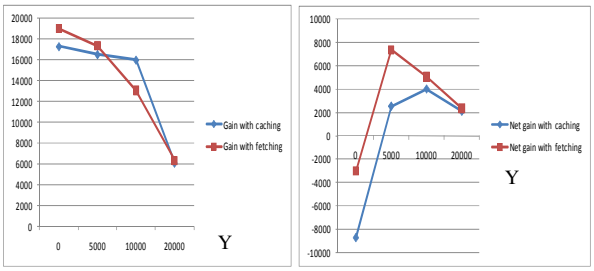


Figure 7. Gain and Net gain comparison

We notice that fetching is more efficient in terms of Net gain, despite the fact that caching is better for Y between 6000 and 19000 in terms of Access gain. This is due to the fact that the duplicated files are smartly allocated with fetching, allowing more customers to access them with a minimum cost, and because there's a lower need for duplicating files than with caching.

We also notice that beyond a certain value of Y, the Net gain decreases, and can even become negative beyond a certain limit (> 20000 in our example). Moreover, as Y increases beyond 10000, the advantage of fetching over caching diminishes.

7. CONCLUSION

This paper provided a study for video files distribution and allocation. We started by highlighting some techniques used in video files caching and distributions. Then, we proposed two mechanisms of files duplications based on files hits and some access cost assumptions set by the operators.

Despite the fact that fetching is better in terms of Net gain, both duplication methods can be more or less efficient, as we have seen that caching is more efficient for the delivery of File 6 for Y=0, 5000 or 1000, and has equal efficiency with fetching for Y=20000. We have also seen that, setting the right demand threshold Y is a key step to achieve cost optimization, along with the access cost threshold A (A=5 in this paper) that we didn't discuss.

We can also combine duplication methods or use new ones for better results. We can, for example, duplicate a file in the zone that has the second and the third highest number of hits on that file and that grants an access cost lower than A at the same time.

However, the results found in the treated example may vary depending on the network configuration (the access costs, the number of files and their distribution) and a duplication method that proves to be efficient for a given network may not work for another one.

Finally, if we compare our proposed techniques with the previous stated works we can find that our algorithms are heuristic ones. This means that, the correlation made between user requests and the network conditions (set by the operator) play an important role in caching decisions.

8. REFERENCES

[1] E.Abd-Elrahman and H.Afifi, 'Optimization of File Allocation for Video Sharing Servers', *NTMS 3rd IEEE International Conference on New Technologies, Mobility and Security NTMS2009*, (20-23 Dec 2009), 1-5.

[2] Nakaniwa, A.; Ohnishi, M.; Ebara, H.; Okada, H.;"File allocation in distributed multimedia information networks," *Global Telecommunications Conference, 1998. GLOBECOM 98*. The Bridge to Global Integration. IEEE, vol.2, (1998), 740-745.

[3] C.Labovitz, S.Iekel-Johnson, D.McPherson, J.Oberheide and F.Jahanian; 'Internet Inter-Domain Traffic', SIGCOMM'10, (2010), 75-86.

[4] Akamai: http://www.akamai.com/

[5] L.B. Sofman, B. Krogfoss and A. Agrawal; "Optimal Cache Partitioning in IPTV Network", *CNS* (2008), 79-84.

[6] D. De Vleeschauwer, Z. Avramova, S. Wittevrongel, H. Bruneel ; "Transport Capacity for a Catch-up Television Service"; *EuroITV'09*, (June 3–5, 2009), 161-169.

[7] D. De Vleeschauwer and K. Laevens; "Performance of caching algorithms for IPTV on-demand services Transactions on Broadcasting", *IEEE TRANSACTIONS ON BROADCASTING, VOL. 55, NO. 2*, (JUNE 2009), 491-501.

[8] D. Applegate, A. Archer and V. Gopalakrishnan; "Optimal Content Placement for a Large-Scale VoD System", *ACM CoNEXT 2010*, (Nov 30 – Dec 3 /2010), 1-12.

[9] L.B. Sofman, and B. Krogfoss; "Analytical Model for Hierarchical Cache Optimization in IPTV Network," Broadcasting, *IEEE Transactions on, vol.55, no.1*, (March 2009), 62-70.

[10] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon; "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System", *In Proceedings of ACM-IMC'07*, (October 24-26, 2007).

Personalized TV Service through Employing Context-Awareness in IPTV NGN Architecture

Songbo SONG
Altran Technologies
Paris, France
Songbo.Song@altran.com

Hassnaa MOUSTAFA
Telecom R&D (Orange Labs), Issy Les Moulineaux, France
Hassnaa.Moustafa@orange.com

Hossam AFIFI
Telecom & Management SudParis ,Evry, France
Hossam.Afifi@it-sudparis.eu

Jacky FORESTIER
Telecom R&D (Orange Labs), Issy Les Moulineaux, France
jacky.forestier@orange.com

ABSTRACT
The advances in Internet Protocol TV (IPTV) technology enable a new model for service provisioning, moving from traditional broadcaster-centric TV model to a new user-centric and interactive TV model. In this new TV model, context-awareness is promising in monitoring user's environment (including networks and terminals), interpreting user's requirements and making the user's interaction with the TV dynamic and transparent. Our research interest in this paper is how to achieve TV services personalization using technologies like context-awareness and presence service on top of NGN IPTV architecture. We propose to extend the existing IPTV architecture and presence service, together with its related protocols through the integration of a context-awareness system. This new architecture allows the operator to provide a personalized TV service in an advanced manner, adapting the content according to the context of the user and his environment.

Categories and Subject Descriptors
D.5.2 [Information Interfaces and Presentation]: user Interfaces – User-centered design; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems H.1.2 [Models and Principles]: User/Machine Systems – Human information processing; H.3.4 [Information Storage and Retrieval]: Systems and software – Current awareness systems, User profiles and alert services Classification Scheme: http://www.acm.org/class/1998/

General Terms
Design, Human Factors.

Keywords
NGN, IPTV, Personalized services, content adaptation and user-centric IPTV, presence service, context awareness.

1. INTRODUCTION
IPTV (Internet Protocol TV) presents a revolution in digital TV in which digital television services are delivered to users using Internet Protocol (IP) over a broadband connection. The ETSI/TISPAN [1] provides the Next Generation Network (NGN) architecture for IPTV as well as an interactive platform for IPTV services. With the expansion of TV content, digital networks and broadband, hundreds of TV programs are broadcast at any time of day. Users need to take a long time to find a content in which he is interested. IPTV services personalization is still in its infancy, where the consideration of the context of the user and his environment (devices and network) and distinguishing of each user in a unique manner still presents a challenge.

IPTV services personalization is beneficial for users and for service and network providers. For users, more adaptive content could be provided as for example targeted advertisement, movie recommendation, provision of a personalized Electronic Program Guide (EPG) saving time in zapping between several programs, and providing content adapted to users' devices considering the supported resolution by each device in a real-time manner and the network conditions which in turn allows for better Quality of Experience (QoE). IPTV services personalization is promising for services providers in promoting new services and opening new business and for network operators to make better utilization of network resources through adapting the delivered content according to the available bandwidth.

Context-awareness is promising in services personalization through considering in real-time and transparent mean the context of the user and his environment (devices and network) as well as the context of the service itself [2]. There are several solutions to realize context awareness. Presence Service allows users to be notified of changes in user presence by subscribing to notifications for changes in state. Context aware solution can be implemented in the presence server to benefit the presence service architecture to convey the more general context information. In this paper we present a solution for IPTV services personalization that considers NGN IPTV architecture while employing presence service to realize context-awareness, allowing access personalization for each user and triggering service adaptation.

The remainder of this paper is organized as follows: Section 2 gives an overview on the related work. Section 3 describes our proposed architecture. In Section 4, we present the related protocols' extension and the communication between the entities in the proposed architecture. Finally, we conclude the paper in Section 5 and highlight some points for future work.

2. RELATED WORK
2.1 Presence Server for Context-Aware
Presence Service enables dynamic presence and status information for associated contacts across networks and applications, it allows users to be notified of changes in user presence by subscribing to notifications for changes in state. The IETF defines the model and protocols for presence service. As presented in Figure 1, the model defines a presentity (an abbreviation for "presence entity"),

as a software entity that sends messages to the presence server for publishing information. The model also defines an entity known as a watcher, which requests information from the presentity through the presence server. The presence service or presence server is in charge of distributing presence information concerning the presentity to the watchers, via a message, called a Notification. In this model the presence protocol is used for the exchange of presence information in close to real time, between the different entities defined by the model.

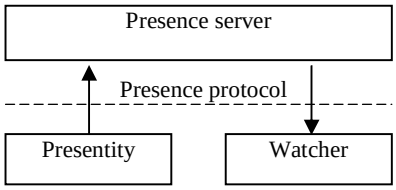


Figure 1 Presence service model

In presence service, the presence information is a status indicator that conveys ability and willingness of a user. However presence information can contains more general information, not only the user's status, for example the information about the environment, devices, etc. Thus presence server can be extended to a general information server and provide more intelligent services: context-awareness services. The work in [3] uses presence server as a context server to create a context-aware middleware for different types of context-aware applications. The proposed context server: (i) obtains the updated context information, (ii) reads, processes, and stores this information in the local database, and (iii) notifies interested watchers about context information. The context delivery mechanisms are based on SIP-SIMPLE [4] for supporting asynchronous distribution of information. The context information is distributed using the Presence Information Data Format. A context-aware IMS-emergency service employing the presence service is proposed in [5]. In the proposed solution, the context information is defined as user's physiological information (e.g., body temperature and blood pressure) and environment information (sound level, light intensity, etc.) and is acquired by Wireless Sensor Networks (WSNs) in the user's surroundings. This context information is then relayed to sensor gateways acting as interworking units between the WSN and the IMS core network. The gateway is considered as the Presence External Agent, where it collects the information, processes it (e.g. through aggregation and filtering), and then forwards it to the presence server that plays the role of a context information base responsible for context information storage, management and dissemination. The Emergency Server retrieves the context information about the user and relays it to the Public Safety Answering Point to enhance the emergency service.

2.2 Context-aware IPTV

Several solutions for context-aware TV systems have been proposed for services personalization. [6] proposes a context-aware based personalized content recommendation solution that provides a personalized Electronic Program Guide (EPG) applying a client-server approach, where the client part is responsible for acquiring and managing the context information, and forwarding it to the server (residing in the network operator/service provider side) which collects the TV programs and determines the most appropriate content for the user. A recommendation manager in the client side notifies the user about the content recommended for him according to the acquired context, including the user context information (user identity, user preference, time) and the content context information (content description). This work does not consider the network context and does not describe the integration with the whole IPTV architecture, however focuses on the Set-Top-Box (STB) as the client and a server in the network operator/service provider side. In addition, services personalization is limited to the content recommendation without considering any other content adaptation means.

A personalization application for context-aware real-time selection and insertion of advertisements into live broadcast digital TV stream is proposed in [7]. This work is based on the aggregation of advertisement information (i.e. advertised products type, advertised products information) and its association with the current user context (identity, activity, agenda, past view) in order to determine the most appropriate advertisement to be delivered. This solution is implemented in STBs and includes modules for context acquisition, context management and storage, advertisement acquisition and storage and advertisement insertion. This solution does not consider the devices and network contexts and does not describe the integration with the whole IPTV architecture, however focusing on the STB side in the user domain.

In [8], we proposed integrating a context-awareness system on top of IPTV/IMS (IP Multimedia Subsystem) architecture aiming to offer personalized IPTV services. This solution relies on the core IMS architecture for transferring the different context information to a context-aware server. The main limitation of this solution is its dependency on IMS which necessitates employing the SIP (Session Initiation Protocol) protocol and using SIP-based equipments, which in turn limits the interoperability of different IPTV systems belonging to different operators and which requires also a complete renewal of the existing IPTV architecture (currently commercialized) which does not employ IMS. Furthermore, the dependency on the SIP protocol limits the possible integration of IPTV services with other rich internet applications (which is an important NGN trend) and hence presents a shortcoming. Consequently, we aim by the solution presented in this paper to increase the integration possibilities with web services in the future and to ease the interoperability with the current IPTV systems. So we advocate the use of HTTP protocol, where the presented solution introduces a presence service based context-awareness on top of NGN IPTV non-IMS architecture. In addition, a mechanism for personalized identification allowing distinguishing each user in a unique manner and a mechanism for content adaptation allowing the customization of the EPG (Electronic Program Guide) and the personalized recommendation are proposed.

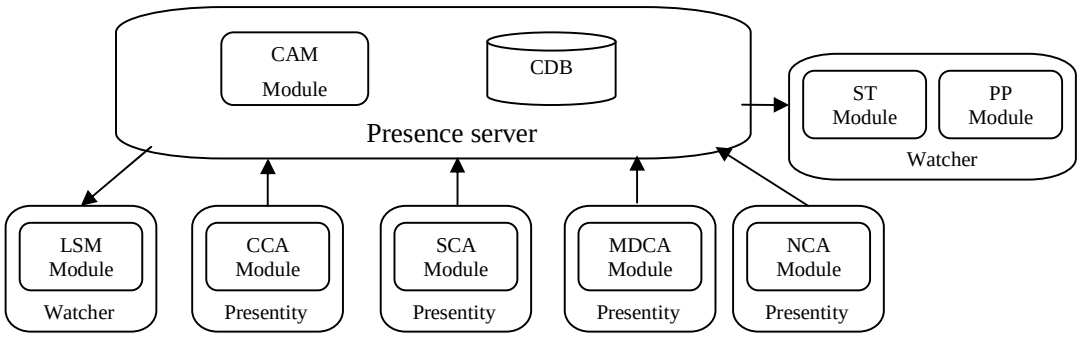


Figure 2 Presence service based Context-Aware System

3. CONTEXT-AWARE IPTV SOLUTION

3.1 Overview of the solution

The presented solution in this paper extends a Context-Aware System which we proposed in [8] to integrate in the presence server on top of NGN IPTV non-IMS architecture. The necessary communication between the Context-Aware System and the other architectural entities in the core network and IPTV service platform is achieved through Extensible Messaging and Presence Protocol (XMPP) [9] and DIAMETER protocols while extending them to allow for the transmission of the acquired context information in real-time. We considered the following personalization means: the customization of the EPG to match the users' preferences, recommending the content that best matches the users' preferences, and adapting it according to the device used by each user and the network characteristics.

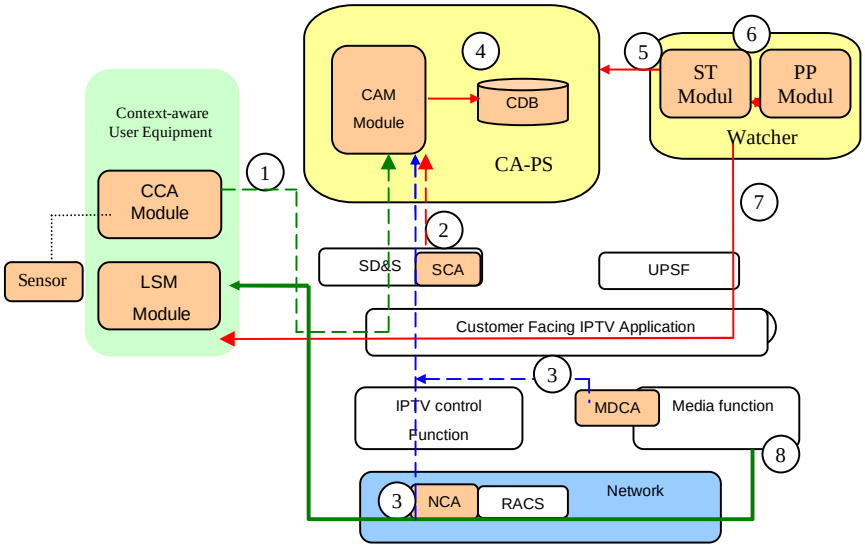
3.2 Presence service based Context-Aware system

We follow a hybrid approach in the design of the Context-Aware System including a centralized server (Context-Aware Server "CAS") and some distributed entities to allow the acquisition of context information from the different domains along the IPTV chain (user, network, service and content domains) while keeping the context information centralized and updated in real-time in the CAS enabling its sharing (and the sharing of users' profiles) by different access networks belonging to the same or different operator(s).

The CAS is a central entity residing in the operator network and includes four modules: i) A Context-Aware Management (CAM) module, gathering the context information from the user, the application server and the network and deriving higher level context information through context inference. ii) A Context Database (CDB) module, storing the gathered and inferred context information. iii) A Service Trigger (ST) module, triggering the personalized service for the user according to the different contexts stored in the CDB. iv) A Privacy Protection (PP) module, verifying the different privacy levels for the user context information before triggering the personalized service.

Other distributed modules gather different context information in real-time: i) In the user domain, the User Equipment (UE) includes a Client Context Acquisition (CCA) module and a Local Service Management (LSM) module. The CCA module discovers the context sources in the user sphere and collects the raw context information (related to the user and his devices) for transmission to the CAM module located in the CAS. On the other hand, the LSM module controls and manages the services personalization within the user sphere according to the context information (example, volume change, session transfer from terminal to terminal, etc). ii) In the service domain, the Service Context Acquisition (SCA) module collects the service context information and transmits it to the CAM, and the Media Delivery Context Acquisition (MDCA) module dynamically acquires the network context information during a media session and transmits it to the CAM. iii) In the network domain, the Network Context Acquisition (NCA) module acquires the network context (mainly the bandwidth information) during the session initiation and transmits it to the CAM.

We integrate Context-Aware Management (CAM) module and Context Database (CDB) module into the presence server to make the latter to treat and store the context information and to realize a Context-Aware Presence Server (CA-PS). The Service Trigger (ST) module and the Privacy Protection (PP) module are integrated into a Watcher (called Context-Aware Watcher) to monitor the change of the context information and actives the appropriate services. The Client Context Acquisition (CCA) module, Service Context Acquisition (SCA) module, Media Delivery Context Acquisition (MDCA) module and Network Context Acquisition (NCA) module can be considered as Context-Aware Presentities which collect the context information and send them to the Context-Aware Presence Server. The Local Service Management (LSM) module can be considered as a Context-Aware Watcher which requests context information about a Context-Aware Presentity from a context-aware presence server then actives the personalized service in user sphere. The Figure 2 presents the proposed Presence service based Context-Aware system.



1.User and device context information and transition
2. Service context information and transition
3. Network context information and transition
4. Context information is stored into the CDB.
5. The ST communicates with the CDB to monitor the context information and discovers the services.
6. The ST communicates with the PP to verify if there is a privacy constraint.
7. The ST sends a request to set up the services or to personalize the services.
8. Adapted content is sent to the UE

Figure 3 Architecture of the Proposed Context-aware IPTV Solution

3.3 Presence service based context awareness on top of NGN IPTV-non IMS architecture

Our solution extends the NGN IPTV architecture defined by the ETSI/TISPAN [1] through integrating the context-aware system described in the previous sub-section and defining the different communication procedures between the different architectural entities in the new system for the transmission and treatment of the different context information related to the users, devices, networks, and service itself. Consequently, TV services personalization could take place through content recommendation and selection, and content adaptation. Figure 3 presents the architecture of this context-aware IPTV solution.

The NGN IPTV architecture includes the following functions: the Service Discovery and Selection (SD&S), for providing the service attachment and service selection; the Customer Facing IPTV Application (CFIA), for IPTV service authorization and provisioning to the user; the User Profile Server Function (UPSF), for storing user's related information mainly for authentication and access control; the IPTV Control Function (IPTV-C), for the selection and management of the media function; and the Media Function (MF), for controlling and delivering the media flows to the User Equipment (UE).

To extend this NGN IPTV architecture, in the service plane, the SCA (Service Context Acquisition) module is integrated in the SD&S IPTV functional to dynamically acquire the service context information making use of the EPG received by the SD&S from the content provider which includes content and media description. And the MDCA (Media Delivery Context Acquisition) module is integrated in the MF to dynamically acquire the network context information during a media session through gathering the network information statistics (mainly on packet loss, jitter, round-trip delay) delivered by the Real Time Transport Control Protocol (RTCP) [10] during the media session. In the network plane, the NCA (Network Context Acquisition) module is integrated in the classical Resource and Admission Control Sub-System (RACS) [11] extending the resource reservation procedure during the session initiation to collect the initial network context information (mainly bandwidth).

In the user plane, we benefit from the UPSF to store the static user context information including user's personal information ("age, gender …"), subscribed services and preferences. In addition, the CCA (Client Context Acquisition) and LSM (Local Service Management) modules extend the UE (User Equipment) to acquire the dynamic context information of the user and his surrounding devices.

After each acquisition of the different context information (related to the user, devices, network and service), it is stored in the CDB (Context Data Base) in the CAM (Context-Aware Management) module, and the ST (Service Trigger) module monitors the context information in the CDB and prepares the personalized service. Before triggering the service, the PP (Privacy Protection) module verifies the different privacy levels on the user's context information to be used. Finally, services personalization is activated by two means matching the different contexts stored in CDB: i) the ST module sends to the UE a personalized EPG, where an HTTP POST message contains the personalized EPG and is sent by ST module to the user through the CFIA. ii) The ST module sends to the IPTV-C (IPTV Control Function) an HTTP message encapsulating the necessary information for triggering the personalized service (for example mean for content adaptation or types of channels/movies to be diffused). The IPTV-C then selects the proper MF (Media Function) and forwards the adaptation information to it to complete the content adaptation process.

4. COMMUNICATION PROCEDURES BETWEEN THE ENTITIES

In this subsection, we present the communication procedures and messages exchange for contextual service initiation, and the context information transmission from the end-user/network/application servers to the CA-PS. The architecture benefits from the existing interfaces (interface between UE and CFIA, interface between SSF and CFIA, interface between CFIA and IPTV-C, interfaces between Presence server and CFIA) to transfer the context information. Extensible messaging and presence protocol (XMPP) is chosen as the context transmission protocol as the XMPP protocol is used in the presence service to stream XML elements for messaging, presence, and request-response services and it can be easily integrated in the HTTP protocol [12] which is used in IPTV non-IMS architecture in the communication process. Another reason is that the XMPP has been designed to send all messages in real-time using a very efficient push mechanism. This characteristic is very important to the context-aware services especially for the IPTV services which are time sensitive services. Furthermore XMPP is widely used in the WEB service (like WEB Social Services) which is promising for interoperation of different services.

Contextual Service initiation

This procedure extends the classical IPTV service initiation and authentication procedure to include user's static context information acquisition. The user's static context information is stored in the UPSF. We extend the Diameter Server-Assignment-Answer message which is sent by the UPSF to the CFIA to include the user static context information by adding a User-Static-Context Attribute Value Pair (AVP). We use the HTTP XMPP message to transmit the user static context information, stored in the UPSF, to the CA-PS. This procedure extends the classical IPTV service initiation and authentication procedure to transfer user's static context information as illustrated in Figure 4.

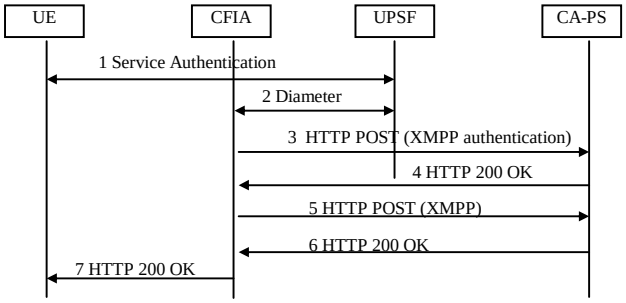


Figure 4 Contextual Service initiation

UE sends HTTP message to the CFIA to initialize the IPTV service. In this step, CFIA communicates with UPSF to perform authentication of UE as in classical IPTV service scenario (message 1). After authentication, the CFIA downloads the user's profile including the user static context information using diameter protocol [13] (message 2), extending the diameter message to include the user static context information by adding a User-Static-Context Attribute. Then the CFIA performs authentication with CA-PS on behalf of the UE using XMPP authentication solution (message 3-4) and sends the user static context information to the CA-PS using HTTP POST message encapsulating an XMPP message which is extended to include more context information attributes (mainly concerning the user's preference, user's subscribed services and user's age) (message 5). The CA-PS accepts the context information, and sends 200 OK response to the UE through CFIA (message 6-7). Figures 5-7 respective illustrates the message 3-5 which presents the usage of XMPP in the authentication phase and the message exchanges phase.

```
POST /webclient HTTP/1.1
Host: httpcm.example.com
Accept-Encoding: gzip, deflate
Content-Type: text/xml; charset=utf-8
Content-Length: 231
<body rid='1573741821'
      sid='SomeSID'
      xmlns='http://jabber.org/protocol/httpbind'>
  <auth xmlns='urn:ietf:params:xml:ns:xmpp-sasl'
        mechanism='SCRAM-SHA-1'>
biwsbj1qdWxpZXQscj1vTXNUQUF3QUFBQU1BQUFBTlAwVEFBQ
UFBQUJQVTBBQQ==
  </auth>
</body>
```

Figure 5 XMPP authentication request message

```
HTTP/1.1 200 OK
Content-Type: text/xml; charset=utf-8
Content-Length: 149
<body xmlns='http://jabber.org/protocol/httpbind'>
  <success xmlns='urn:ietf:params:xml:ns:xmpp-sasl'>
    dj1wTk5ERlZFUXh1WHhDb1NFaVc4R0VaKzFSU289
  </success>
</body>
```

Figure 6 XMPP authentication response message

```
POST /client HTTP/1.1
Host: httpcm.example.com
Accept-Encoding: gzip, deflate
Content-Type: text/xml; charset=utf-8
Content-Length: 188
<body rid='1249243562'
        sid='SomeSID'
       xmpp:version='1.0'
         xmlns='http://jabber.org/protocol/httpbind'
         xmlns:xmpp='urn:xmpp:xbosh'
         xmlns:stream='http://etherx.jabber.org/streams'>
  <stream>
    <message from='user@client.com'
          to='user@cas.com'
          xmlns='jabber:client'>
        <preference></preference>
        <subscription></subscription>
    </message>
  </stream>
</body>
```

Figure 7 Contextual Service initiation

User/device context information acquisition

This procedure is proposed to allow the CCA module of the UE to update in the CA-PS the user/device context information that it dynamically acquires. We use the HTTP POST message encapsulating an XMPP to transmit the context information (the user and device context mainly concerning user's location (indoor location), devices location, supported network type, supported media format, screen size, and the network type) as illustrated in the Figure 8.

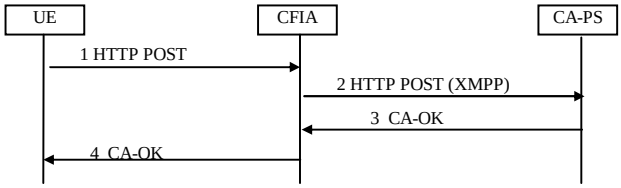


Figure 8 User/device context information acquisition

Figure 9 illustrates the User/Device Dynamic Context Information Transmission. Once the UE is successfully registered, it collects the context information from the environment (dynamic user context information like position, activity; device context information) and updates the context information to the context-aware server using HTTP POST message encapsulating an XMPP message through CFIA (message 1-2). Once the CAM within the CA-PS receives the current user context information, it will infer it and store it in the Context-aware DB and send an CA-OK message to the UE which is similar to the 200 OK message (message 3-4).

```
POST /client HTTP/1.1
Host: httpcm.example.com
Accept-Encoding: gzip, deflate
Content-Type: text/xml; charset=utf-8
Content-Length: 188
<body rid='1249243562'
        sid='SomeSID'
       xmlns='http://jabber.org/protocol/httpbind'>
     <message from='user@client.com'
       to='user@cas.com'
       xmlns='jabber:client'>
      <place-type><home/></place-type>
      < location>< salon/></location>
    </message>
    <message from='user@client.com'
       to='user_device@cas.com' xmlns='jabber:client'>
       <suppported_network_type> < fix> < suppported_network_type >
       <supported_media_format>< mpeg2/>< supported_media_format>
       <screen_size><>< screen_size>
    </message>
</body>
```

Figure 9 user/device context information

Service information acquisition

This procedure is similar to the procedure of the user context information dynamic transmission to the CA-PS and the dynamic update for service information, where the HTTP POST and XMPP messages are also used as illustrated in Figure 10. The service information is acquired by the SCA module through extracting the service context information from the Electronic Program Guide (EPG) received by the SSF, and transferred to the CA-PS via CFIA. The context information is encapsulated in the XMPP message followed the predefined XML format, while the context information attributes representing the service context (mainly, the service start-time, end-time, content-type and codec). Figure 11 illustrates the HTTP and XMPP message which includes the attributes representing the service context.

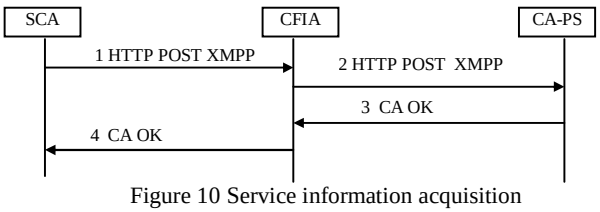


Figure 10 Service information acquisition

```
POST /client HTTP/1.1
Host: httpcm.example.com
Accept-Encoding: gzip, deflate
Content-Type: text/xml; charset=utf-8
Content-Length: 188
<body rid='1249243562'
        sid='SomeSID'
       xmlns='http://jabber.org/protocol/httpbind'>
     <message from='SCA@IPTV.com'
       to='service@IPTV.com'
       xmlns='jabber:client'>
      < content-type></ content-type>
     < starttime></ starttime>
     < endtime></ endtime>
     < codec></ codec>
   </message>
</body>
```

Figure 11 Service context information

Network Context Information Transmission during the Session Initiation

This procedure concerns the network context information transmission during the session initiation through extending the classical resource reservation process. In this latter, the IPTV-C receiving the service request sends a Diameter protocol AA-Request message to the Resource and Admission Control Sub-System (RACS) for the resource reservation. Based on the available resources, the RACS will decide whether to do or not a resource reservation for the service. An AA-answer message is sent by the RACS to the IPTV-C for informing the latter the results of the resource reservation (successful resource reservation or not). We extend this process in order to send the bandwidth information to the IPTV-C, where the NCA that we proposed to integrate in the RACS generates a Context AA-Answer (CAA-Answer) message extending the AA-Answer message through adding a Network-Information Attribute Value Pair to include the bandwidth information. As illustrated in Figure 12, when the user wishes to begin the service, he sends an RTSP message to the IPTV control function (IPTV-C) as in the classic IPTV service scenario. After receiving the RTSP message, the IPTV-C contacts the NCA-RACS module using the classic Diameter AA-Request message. Then the NCA-RACS collects the resource information (bandwidth) and sends it to IPTV-C through CAA-Answer message. After receiving the response from the NCA-RACS, the IPTV-C sends the bandwidth information to the CA-PS using HTTP POST message which encapsulates an XMPP message. Figure 13 illustrates the HTTP POST message which includes the attributes representing the network context.

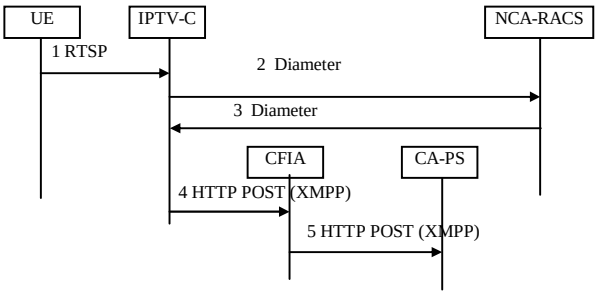


Figure 12 Network Context Information Transmission during the Session Initiation

```
POST /client HTTP/1.1
Host: httpcm.example.com
Accept-Encoding: gzip, deflate
Content-Type: text/xml; charset=utf-8
Content-Length: 188
<body rid='1249243562'
        sid='SomeSID'
       xmlns='http://jabber.org/protocol/httpbind'>
     <message from='NCA@IPTV.com'
       to='network@IPTV.com'
       xmlns='jabber:client'>
            <bandwidth> </bandwidth>
 </message>
</body>
```

Figure 13 Network context information

Network Context Information Dynamic Transmission

This procedure allows the MDCA to dynamically transmit the network context information related to the media session to the CA-PS as illustrated in Figure 14. The HTTP POST and XMPP message is used to transfer the context information, where the representation of the network context information follows the XML format including context information attributes representing the network context (mainly, jitter, packet loss and delay). As illustrated in Figure 15, the message is used to transfer the context information by the MDCA module to the CA-PS via the CFIA.

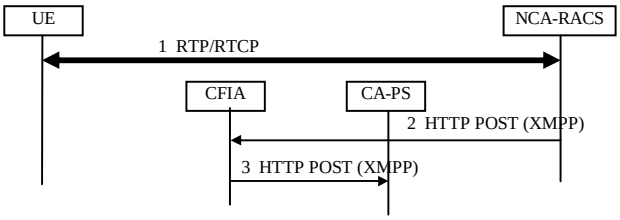


Figure 14 Network context information acquisition

```
POST /client HTTP/1.1
Host: httpcm.example.com
Accept-Encoding: gzip, deflate
Content-Type: text/xml; charset=utf-8
Content-Length: 188
<body rid='1249243562'
        sid='SomeSID'
       xmlns='http://jabber.org/protocol/httpbind'>
     <message from='MDCA@IPTV.com'
       to='network@IPTV.com'
       xmlns='jabber:client'>
            <jitter> </jitter>
             <packet_loss> </packet_loss>
             <delay> </delay>
  </message>
</body>
```

Figure 15 Dynamic network context information

5. CONCLUSION

In this paper, we propose the integration of a context-aware system into the NGN IPTV-non IMS architecture to allow IPTV service personalization in a standard manner. The proposed context-aware IPTV architecture follows a hybrid approach making the system more competent to support the personalized services. The CAS (Context-Aware Server) is integrated into the presence server to benefit the presence service architecture and protocol for the context acquisition. Our proposed solution is easy to be deployed since it extends the existing IPTV architecture (standardized at the ETSI/TISPAN) and existing protocols (standardized within the IETF). The proposed solution could assure personalized IPTV service access with mobility within the domestic sphere as well as nomadic access of personalized IPTV service, since the user identity is not attached to the used device and the proposed context-aware system is not fully centralized. Furthermore, service acceptability could be assured by users thanks to the privacy consideration. Our next step is to test the performance of the proposed context-aware IPTV personalization services.

6. REFERENCES

[1] ETSI TS 182 028. 2008. "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN integrated IPTV subsystem Architecture".

[2] S. Song, H. Moustafa, H. Afifi, "Context-Aware IPTV," IFIP/IEEE Management of Multimedia & Mobile Network Service (MMNS 2009)

[3] M. Zarifi, A Presence Server for Context-aware Applications, Master thesis, School of Information and Communication Technology, KTH, December 17, 2007.

[4] SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE) IETF Working Group. http://www.ietf.org/html.charters/simple-charter.html

[5] El Barachi, M., Kadiwal, A., Glitho, R., Khendek, F., Dssouli, R.: An Architecture for the Provision of Context-Aware Emergency Services in the IP Multimedia Subsystem. In: Vehicular Technology Conference, Singapore, pp. 2784–2788 (2008).

[6] Santos Da Silva, F., Alves, LGP., Bressan, G. Personal TVware: A Proposal of Architecture to Support the Context-aware Personalized Recommendation of TV Programs. 7th European Conference on Interactive TV and Video (2009).

[7] Thawni, A., Gopalan, S.. Sridhar, V.: Context Aware Personalized Ad Insertion in an Interactive TV Environment. 4th Workshop on Personalization in Future TV (2004)

[8] Songbo Song, Hassnaa Moustafa, Hossam Afifi, "Personalized TV Service through Employing Context-Awareness in IPTV/IMS Architecture," FMN (2010)

[9] Extensible Messaging and Presence Protocol (XMPP) IETF Working Group. http://www.ietf.org/html.charters/xmpp-charter.html

[10] H. Schulzrinne, S. Casner, R.Frederick, V.Jacobson.: RTP: A Transport Protocol for Real-Time Applications. In: IETF RFC 3550 (2003)

[11] ETSI ES 282 003: Resource and Admission Control Sub-System (RACS): Functional Architecture

[12] XEP-0124: Bidirectional-streams Over Synchronous HTTP (BOSH) http://xmpp.org/extensions/xep-0124.html

[13] 3GPP TS 29.229: "Cx Interface based on Diameter – Protocol details.

Quality of Experience for Audio-Visual Services

Mamadou Tourad DIALLO
Orange Labs, issy les moulineaux, France
mamadoutourad.diallo@orange.com

Hassnaa MOUSTAFA
Orange Labs, issy les moulineaux, France
hassnaa.moustafa@orange.com

Hossam AFIFI
Institut telecom Telecom SudParis
hossam.Afifi@it-sudparis.eu

Khalil Laghari, INRS-EMT, University of Quebec, Montreal, QC, Canada
Khalil.laghari@gmail.com

ABSTRACT

Multimedia and audio-visual services are observing a huge explosion during the last years. In this context, users' satisfaction is the aim of each service provider to reduce the churn, promote new services and improve the ARPU (Average Revenue per User). Consequently, Quality of experience (QoE) consideration becomes crucial for evaluating the users' perceived quality for the provided services and hence reflects the users' satisfaction. QoE is a multidimensional concept consisting of both objective and subjective aspects. For adequate QoE measure, the context information on the network, devices, the user and his localization have to be considered for successful estimation of the user's experience. Up till now, QoE is a young subject with few consideration of context information. This paper presents a new notion of user's experience through introducing context-awareness to the QoE notion. A detailed overview and comparison of the classical QoE measurement techniques are given. The literature contributions for QoE measurement are also presented and analyzed according to the new notion of user's experience given in this paper.

General Terms:

Measurement, Human Factors,

Keywords:

QoE, objective, subjective, hybrid, MOS, Human behavior, human physiological, cognitive factors

Introduction

With the network heterogeneity and increasing demand of multimedia and audio-visual services and applications, **Quality of Experience (QoE)** has become a crucial determinant of the success or failure of these applications and services. As there is burgeoning need to understand human hedonic and aesthetic quality requirements, QoE appears as a measure of the users' satisfaction from a service through providing an assessment of human expectations, feelings, perceptions, cognition and acceptance with respect to a particular service or application [1]. This helps network operators and service providers to know how users perceive the quality of video, audio, and image which is a prime criterion for the quality of multimedia and audio-visual applications and services [2]. QoE is a multidimensional concept consisting of both objective (e.g., human physiological and cognitive factors) and subjective (e.g., human psychological factors) aspects. As described in [3], context is fundamental part of communication ecosystem and it influences human behavior. Sometimes as human behavior is subjective and random in nature, it varies over the time. So, an important challenge is to find means and methods to measure and analyze QoE data with accuracy. QoE is measured through three main means: i) objective means based on network parameters (e.g. packet loss rate and congestion notification from routers), ii) subjective means based on the quality assessment by the users giving the exact user perception of the service, and iii) hybrid means which consider both objective and subjective methodologies. Some research contributions also present methods to evaluate the QoE based on users' behavior, technical parameters, some statistical learning, and measuring the users' feedbacks aiming to maximize user satisfaction and optimize network resources uses. Furthermore, there is a growing demand for methods to measure and predict QoE for multimedia and audio-visual services through some tools as (conviva, skytide, mediamelon etc)[4,5,6]. These tools use metrics such as startup time, playback delay, video freeze, number of rebuffering caused by network congestion to evaluate the perceived quality by the users and measure the QoE. According to [7], the parameters that affect QoE can be generally classified into three types: The quality of video/audio content at the source, the delivery of content over the network and the human perception (including acceptation, ambiance, preferences, needs, etc).

However, QoE as defined and measured today is not sufficient to adapt content or delivery for improving the users' satisfaction. Consequently, we define in this paper a new notion for QoE introducing more contextual parameters on the user and his environment to accurately predict the QoE. To maximize the end-user satisfaction, it is important to do some adaptation at both the application layer (e.g. choice of the compression parameters, change bitrate, choice of layers which will be send to client)and the network layer (e.g. delivery means, unicast, multicast, choice of access network, source server choice, delivery from a cloud...)

The remainder of this paper is organized as follows, Section 2 defines a context-aware QoE notion, Section 3 gives a review on the QoE measuring techniques with a comparison between them, Section 4 describes the related research contributions on QoE evaluation mean comparing them according to the user experience notion. Finally, Section 5 concludes the paper and highlights some perspectives for the work.

2. Context-Aware QoE

For improving the QoE and maximizing the user's satisfaction from the service, we extend the QoE notion through coupling it with different context concerning the user (preferences, content consumption style, level of interest, location, …..) and his environment including the used terminal (capacity, screen size, ..) and network (available bandwidth, delay, jitter, packet loss rate..).

Human moods, expectations, feelings and behavior could change with variation in his/her context [8]. The context-aware QoE notion presents an exact assessment of QoE with respect to contextual information of a user in a communication ecosystem. To measure user experience**,** context information needs to be gathered as a first step in a dynamic and real-time manner during services access. This information includes: i): Devices Context (Capacity, Screen size, Availability of Network Connectivities …), ii) Network Context (jitter, packet loss, available bandwidth …), iii) User Context (Who is the user, his preferences, his consumption style, gender, age…), and iv) User Localization. User localization considers both the physical location of the user (indoor or outdoor) and the user location within the network that we call the Geographical Position Within the Network (GPWN) as illustrated in Figure 1. For the physical localization, user's can be localized indoors (within their domestic sphere for examples) through the Wi-Fi or Bluetooth techniques [9]. RFID can be used also to localize users with respect to RFID readers that are supposed to have known places. For the outdoor localization of users, the following methods can be employed: GPS (Global Positioning System), Radio Signal Strength Indicator (RSSI), and Cell-ID (based on the knowledge of base stations locations and coverage). On the other hand, the GPWN localization of users aims to indicate the users' proximity to base stations, DSLAM, access points, content source servers…etc. This latter type of localization allows to choose in an adaptive manner the most suitable content delivery mean (for example, multicast can be used if many users are proximate to the same DSLAM, the content source server can be chosen according to its proximity to users, users who are near the base stations can receive the high video/audio quality for example, …).

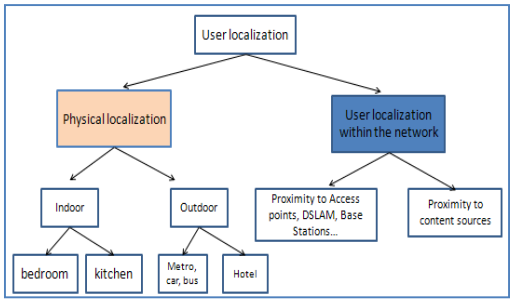


Figure 1: Localization principle

After context information gathering, a second step is to personalize and adapt the content and the content delivery means according to the gathered context information for improving user's satisfaction of services and for better resources consumption. Figure 2 illustrates our vision of content adaptation.

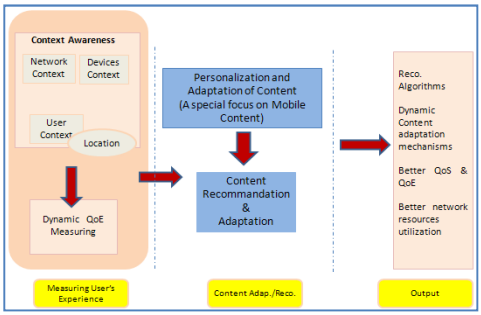


Figure 2: New vision of content adaptation

This paper focuses on the QoE measuring means presenting a comparison between them showing their suitability and/or limitation to the new notion of user's experience described in this section. The content personalization and adaptation are out of the scope of this paper and are a next step of our work.

3. QoE Measuring Techniques

Internet Service Providers (ISPs) use Quality of Service (QoS) parameters such as bandwidth, delay or jitter to guarantee good service quality. QoS is achieved if a good QoE is also achieved for the end users in addition to the classical networking configuration parameters [10]. The challenging question is how to quantify the QoE measure. In general there are three main techniques for measuring the QoE as discussed in the following sub-sections.

3.1 Objective QoE Measuring Techniques

Objective QoE measuring is based on network related parameters that need to be gathered to predict the users' satisfaction. Objective QoE measuring follows either an intrusive approach that requires a reference image/video/audio to predict the quality of the perceived content or a non intrusive approach that does not require reference information on the original content.

3.1.1 Intrusive Techniques

A lot of objectives QoE measurement solutions follow an intrusive approach. They need both the original and degraded signal (audio, video, and image) to measure QoE. Although intrusive methods are very accurate and give good results, they are not so feasible in real-time applications as it's not easy to have the original signal. The following subsections present some objective intrusive techniques.

- **PSNR (Peak Signal-to-Noise Ratio)** is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is defined via the Mean Squared Error (MSE) between an original frame *o* and the distorted frame *d* as follows [11].

$$MSE = \frac{1}{M.N} \sum_{m=1}^{M} \sum_{n=1}^{N} |o(m,n) - d(m,n)|^2$$

Where each frame has $M \times N$ pixels, and $o(m, n)$ and $d(m, n)$ are the luminance pixels in position (m, n) in the frame. Then, PSNR is the logarithmic ratio between the maximum value of a signal and the background noise (MSE). If the maximum luminance value in the frame is L (when the pixels are represented using 8 bits per sample, $L = 255$) then

$$PSNR = 10 \log \frac{255^2}{MSE}$$

- **Perceptual Evaluation of Video Quality (PEVQ):** PEVQ is an accurate, reliable and fast video quality measure. It provides the Mean Opinion Score (MOS) estimates of the video quality degradation occurring through a network, e.g. in mobile and IP-based networks. PEVQ can be ideally applied to test video telephony, video conferencing, video streaming, and IPTV applications. The degraded video signal output from a network is analyzed by comparison to the undistorted reference video signal on a perceptual basis. The idea is to consider the difference between the luminance and the chrominance domains and calculates quality indicators from them. Furthermore the activity of the motion in the reference signals provide another indicator representing the temporal information. This indicator is important as it takes into account that in frames series with low activity the perception of details is much higher than in frames series with quick motions. After detecting the types of distortions, the distorted detected information is aggregated to form the Mean Opinion Score (MOS) [12].
- **Video Quality Metric (VQM):** VQM is a software tool developed by the Institute for Telecommunication Science (ITS) to objectively measure the perceived video quality. It measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block [13].
- **Structural Similarity Index (SSIM):** SSIM uses a structural distortion based measurement approach. Structure and similarity in this context refer to samples of the signals having strong dependencies between each other, especially when they are close in space. The rational is that the human vision system is highly specialized in extracting structural information from the viewing field and it is not specialized in extracting the errors. The difference with respect to other techniques mentioned previously such as PEVQ or PSNR, is that these approaches estimate perceived errors on the other hand SSIM considers image degradation as perceived change in structural information . The resultant SSIM index is a decimal value between -1 and 1, where the value 1 indicates a good score and the value -1 indicates a bad score [14].

3.1.2 Non Intrusive Techniques

It is difficult to estimate the quality in the absence of the reference image or video which is not usually available all time, as in

streaming video and mobile TV applications. The objective non intrusive approach presents methods that can predict the quality of the viewed content based on the received frames without requiring the reference signal but using information that exist in the receiver side. The following are some methods that predict the user perception based on the received signals.

The method presented in [15] is based on the blur metric. This metric is based on the analysis of the spread of the edges in an image which is an estimated value to predict the QoE. The idea is to measure the blur along the vertical edges by applying edge detector (e.g. vertical Sobel filter which is an operator used in image processing for edge detection.). Another method is presented in [16] based on analyzing the received signal from the bit stream by calculating the number of intra blocks, number of inter blocks, and number of skipped blocks. The idea proposed in this work is to predict the video quality using these parameters. The predictor is built by setting up a model and adapts its coefficients using a number of training sequences. The parameters used are available at the decoder (client side).

The E-model proposed in [17] uses the packet loss and delay jitter to quantify the user perception of service. The E-model is a transmission rating factor R:

$$R = Ro - Is - Id - Ie + A,$$

Where Ro represents the basic signal-to-noise ratio, Is represents the impairments occurring simultaneously with the voice signal, Id represents the impairments caused by delay, and Ie represents the impairments caused by low bit rate codecs. The advantage factor A is used for compensation when there are other advantages of access to the user.

3.2 Subjective QoE Measuring Techniques

Subjective QoE measurement is the most fundamental methodology for evaluating QoE. The subjective measuring techniques are based on surveys, interviews and statistical sampling of users and customers to analyze their perceptions and needs with respect to the service and network quality. Several subjective assessment methods suitable for video application have been recommended by ITU-T and ITU-R. The subjective measures present the exact user perception of the viewed content (audio, video, image…) which is considered as a better indicator of video quality as it is given by humans.

The most famous metric used in subjective measurement is the MOS (Mean Opinion Score), where subjects are required to give a rating using the rating scheme indicated in Table 1.

In order to analyse subjective data, quantitative techniques (e.g., statistics, data mining etc) and qualitative techniques (e.g., grounding theory and CCA framework) could also be used [18]. Once subjective user study is complete, data are to be analyzed using some statistical or data mining approaches. Conventionally, non-parametric statistics is used for ordinal and nominal data, while parametric statistic or descriptive statistics is used for interval or ratio data.

Table 1: MOS Rating (source ITU-T)

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

3.3 Hybrid QoE Measuring Techniques

Hybrid QoE measurement merges both objective and subjective means. The objective measuring part consists of identifying the parameters which have an impact on the perceived quality for a sample video database. Then the subjective measurement takes place through asking a panel of humans to subjectively evaluate the QoE while varying the objective parameters values. After statistical processing of the answers each video sequence receives a QoE value (often, this is a Mean Opinion Score, or MOS) corresponding to certain values for the objective parameters.To automate the process, some of the objective parameters values associated with their equivalent MOS are used for training an RNN (Random Neural Network) and other values of these parameters and their associated MOS are used for the RNN validation. To validate the RNN, a comparison is done between the MOS values given by the trained RNN and their actual values. If these values are close enough (having low mean square error), the training is validated. Otherwise, the validation fails and a review of the chosen architecture and its configurations is needed [19]. PSQA (Pseudo-subjective Quality Assessment) is a hybrid technique for QoE measurement and is illustrated in Figure 3. In PSQA, training the RNN system is done by subjective scores in

real-time usage. The system maps the objective values to obtain the Mean Opinion Score (MOS). The advantages of this method are minimizing the drawbacks of both approaches as it is not time consuming and it does not require manpower (except in the subjective quality assessment preliminary step).

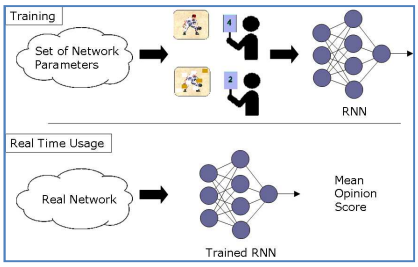


Figure 3: PSQA principle

3.4 Comparison of QoE Measuring Techniques

We compare the previously discussed QoE measurement techniques through considering the required resources, feasibility, accuracy and application type. Table 2 illustrates this comparison.

Table 2: Comparison of QoE measuring means

parameters / methods		REQUIRED RESOURCES	FEASIBILITY	ACCURACY	APPLICATION TYPE
OBJECTIVES	INTRUSIVE	Gathering network parameters requiring more signaling, monitoring sensors, monitoring algorithms	It's easy to collect network parameters and to have reference image/video for no real time traffic and then measure the QoE	Does not consider users' opinions. Not very accurate	Not suitable for Real time traffic (image/VoD)
	NON INTRUSIVE	High receiver (user terminal capacity)	Less useful in real applications as in mobile equipment where the computational power frequently is limited.	Does not consider users' opinions. Not very accurate	Suitable for Live traffic (Streaming videos, Mobile TV)
SUBJECTIVE		Manpower Engaged users	Not realistic to ask all viewers to give scores	Very accurate (based on user's perception)	Any multimedia traffic
HYBRID		Engaged users. Random Neural Network (RNN)	Easy to have subjective measures (only once for learning). Once the tool has been trained, non linear function that can map any possible value of parameters into MOS.	Very accurate (subjective+objective)	Any multimedia traffic

4. Research Contributions on QoE Measurement

Several methods to measure QoE exist in the literature which can be grouped into three main types: i) User behavior and technical parameters, ii) Learning process and iii) Measuring QoE based on subjective measure, objective and localization.

4.1 User behavior and technical parameters

The solution proposed in [20] predicts the QoE based on user engagement and technical parameters. This method quantifies the QoE of mobile video consumption in a real-life setting based on user behavior and technical parameters to indicate the network and video quality. The parameters are: the transport protocol, quality of video source, types of network that was used to transmit the video, number of handovers during transmission and the user engagement (percentage of the video that was actually watched by the user). Then, a decision tree is defined to give the Mean Opinion Score (MOS) based on these parameters.

4.2 Learning process: The learning process is based on collecting parameters and studying their behavior. It is based on linear regression, non linear regression, statistical methods, neural network tools etc… to derive weights for different parameters. Regression is the process of determining the relationships between the different variables. The method in [21] uses statistical learning to predict the QoE. After gathering the network parameters and considering linear regression, this technique calculates the QoE. The network parameters considered in this method are: packet loss rate, frame rate, round trip time, bandwidth, and jitter and the response variable is the perceived quality of experience (QoE) represented by (MOS). In [22], another method based on learning use combination of both application and physical layer parameters for all content types. The application layer parameters considered are the content type (CT), sender bitrate (SBR) and frame rate (FR) and the physical layer parameters are the block error (BLER) and a mean burst length (MBL). After gathering these parameters (in application and physical layer), non-linear regression is used to learn the relation between these parameters and MOS (Mean Opinion Score). An analytical formula is then used to predict the MOS.

Regression

4.3 Measuring QoE based on subjective measure, objective and localization

The method proposed in [23] considers the multi-dimensional concept of QoE by considering the objective, subjective and user localization to estimate the quality of viewed content. For gathering the subjective measure, after watching a video, the user answers a short questionnaire to evaluate various aspects of the watched video considering: the actual content of video, the picture quality (PQ), the sound quality (SQ), the emotional satisfaction. In addition, users are asked to which extend these aspects influenced his general QoE. During the video watching, objectives parameters (packet loss, jitter) are gathered by using RTCP Receiver Records (RR) these reports are sent each 5 seconds to the streaming server. To localize the user, the RSSI is recovered by the tool Myexperience. After calculating the correlations between parameters and between parameters and general QoE using the Pearson coefficient (most commonly used method of computing a correlation coefficient between variables that are linearly related), a general QoE measure can be modeled by using the multi-linear regression.

4.4 Comparison of the Literature Contributions on QoE Measurement

We compare in Table 3 the previously discussed QoE measurement proposed in the literature through considering the network type, the localization, and the context consideration.

Table 3 Comparison of Literature Contributions

Techniques / Parameters	Learning		User behavior and technical's parameters	Subjectives/ psychological and cognitives aspects	Objectives		Subjectives objectives and localization
	Repression based on applications and physicals parameters	Linear regression based on network parameters			Non intrusive	Intrusive	
Oriented	Mobile	Mobile	Mobile/fixed	Mobile/fixed	Mobile/fixed	Mobile/fixed	Mobile
User localization	NO	NO	NO	NO	NO	NO	YES
User preferences	NO	NO	NO	YES	NO	NO	YES
Dynamic consideration of network context	YES	YES	YES	YES	NO	YES	YES
Terminal context	NO	NO	NO	NO	NO	NO	NO
Considering the content context	NO	NO	NO	NO	NO	NO	NO

As we can see, there are lot of methods in the research field for measuring QoE. The learning methods include only network context information without considering other context information as (device context, user localization, his preferences, content format…).

The method presented in [20], which is based on user behavior and technical parameters focuses only on the transport protocol, network parameters and user engagement. In this method there is no consideration of other context information like the user localization, and the terminal context. The presented technique in [23] uses much context information in QoE model. The user localization, the users' preferences, the network context are considered but the terminal context and the content context are omitted. Based on this comparison, our proposition is to build a QoE model which considers more context information to perform the QoE measure. Figure 4 illustrates our vision.

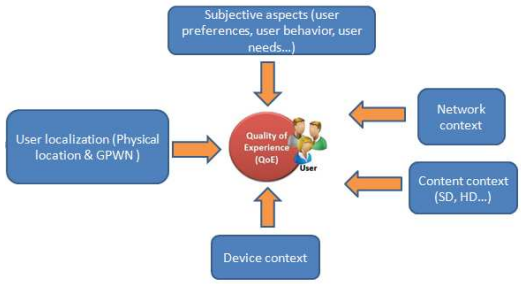


Figure 4: QoE Model

5. Conclusion and Perspectives

With the explosion of multimedia and audio-visual services and the high competition between service providers offers, Quality of Experience (QoE) becomes a crucial aspect for service providers and network operators to continue gaining users' satisfaction. In this context, there is burgeoning need to understand human hedonic and aesthetic quality requirements. This paper presents a study on the QoE measuring means considering both the classical methods and the research contributions. Classical methods consist of objective, subjective and hybrid techniques. A comparison is given between these methods based on the required resources, the feasibility, the accuracy and the application type suitability. The existing research contributions are also presented and classified into three types while presenting a comparison between them based on our presented vision of user experience. A new notion of QoE is presented in this paper considering context information (network context, device context, user context, content context…)

aiming to better improve the user satisfaction from the consumed service.

The perspective of this work is to associate the different context information with QoE measurement in a dynamic manner to satisfy the new notion of the user experience for content and delivery adaptation.

References

[1] Laghari, Khalil Ur Rehman, Crespi, N.; Molina, B.; Palau, C.E.; , "QoE Aware Service Delivery in Distributed Environment," Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of AINA Conference on , vol., no., pp.837-842, 22-25 March 2011.

[2]T.Tominaga, T.Hayashi, J.Okamoto, A.Takahashi, "Performance comparison of subjective quality assessment methods for mobile video",June 2010

[3]Laghari, Khalil ur rehman., Connelly.K, Crespi, N.: K. Laghari, " Towards Total Quality of Experience: A QoE Model in Communication Ecosystem". IEEE Communication Magazine April 2012.

[4] www.conviva.com

[5] www.skytide.com

[6] www.mediamelon.com

[7] F. Kuipers, R. Kooij,D.D. Vleeschauwer, K. Brunnström, "Techniques for Measuring Quality of Experience", 2010

[8] V. George Mathew. (2001) ENVIRONMENTAL PSYCHOLOGY. http://www.psychology4all.com/environmentalpsychology.htm.

[9] Deliverable D3 1 1 Context-Awareness- Version15 Dec 2011.pdf, Task 3.1 (T3.1): Context Definition, Gathering and Monitoring

[10] I.Martinez-Yelmo, I.Seoane, C.Guerrero, "Fair Quality of Experience (QoE) Measurements Related with Networking Technologies",2010

[11] A. N. Netravali and B. G. Haskell, "Digital Pictures: Representation, Compression, and Standards (2nd Ed)," 1995.

[12] http://www.pevq.org/video-quality-measurement.html

[13] M. Pinson, S. Wolf, "A New Standardized Method for Objectively Measuring", 2004 IEEE. Transactions on Broadcasting,

[14] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human", Journal of the Optical Society of America, 2007

[15] P.Marziliano, F.Dufaux, S.Winkler ,T. Ebrahimi, "A no-reference perceptual bluc metric",2002

[16] A.Rossholm, B.Lovstrom, "A New Low Complex Reference Free Video Quality Predictor", Multimedia Signal Processing, IEEE 10th Workshop ,2008

[17] L.Ding and R A. Goubran, "Speech Quality Prediction in VoIP Using the Extended E-Model", in IEEE GLOBECOM,2003

[18] Laghari, Khalil ur rehman., Khan, I, Crespi, N.K.Laghari, " Qualitative and Quantitative Assessment of Multimedia Services in Wireless environment" *MoVid, ACM Multimedia system* USA 22-24 Feb 2012.

[19]K.D.Singh,A.ksentini,B.Marienval,"Quality of Experience measurement tool for SVC video coding", ICC 2011

[20] De Moor, K. ; Juan, A. ; Joseph, W. ; De Marez, L. ; Martens, L. "Quantifying QoE of Mobile Video Consumption in a Real-Life Setting Drawing on Objective and Subjective Parameters" , , Broadband Multimedia Systems and Broadcasting (BMSB), 2011 IEEE International Symposium on IEEE, June 2011

[21] Muslim Elkotob ; Daniel Granlund ; Karl Andersson ; Christer Åhlund, "Multimedia QoE Optimized Management Using Prediction and Statistical Learning", on Local Computer Networks, IEEE,2010

[22] A.Khan, L.Sun,E.Ifeachor, J.O.Fajardo, F Liberal Michal Ries,O.Nemethova,M.Rupp, "Video Quality Prediction Model for H.264 Video over UMTS Networks and their application on mobile video streaming,,IEEE International Conference on Communications (2010)

[23] István Ketykó, Katrien De Moor, Wout Joseph, "Performing QoE-measurements in an actual 3G network", 2010 IEEE International Symposium on Broadband Multimedia

A Triplex-layer Based P2P Service Exposure Model in Convergent Environment

Cuiting HUANG*, Xiao HAN*, Xiaodi HUANG**, Noël CRESPI*
*Institut Mines-Telecom – Telecom Sudparis
9, Rue Charles Fourier, 91000, Evry, France
** Charles Sturt University, Albury, NSW 2640, Australia
*{cuiting.huang, han.xiao, noel.crespi}@it-sudparis.eu, **xhuang@csu.edu.au

ABSTRACT
Service exposure, including service publication and discovery, plays an important role in next generation service delivery. Various solutions to service exposure have been proposed in the last decade, by both academia and industry. Most of these solutions are targeting specific developer groups, and generally use centralized solutions, such as UDDI. However, the reality is that the number of services is increasing drastically with their diverse users. How to facilitate service discovery and publication processes, improve service discovery, selection efficiency and quality, and enhance system scalability and interoperability are some challenges faced by the centralized solutions. In this paper, we propose an alternative model of scalable P2P-based service publication and discovery. This model enables converged service information sharing among disparate service platforms and entities, while respecting their intrinsic heterogeneities. Moreover, the efficiency and quality of service discovery are improved by introducing a triplex-layer based architecture for the organizations of nodes and resources, as well as for the message routing. The performance of the model and architecture is evident by theoretical analysis and simulation results.

Categories and Subject Descriptors
C.2.4 [**Computer-Communication Networks**]: Distributed Systems – distributed applications, distributed databases

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – retrieval models, search process, selection process

General Terms
Management, Performance, Design

Keywords
Service Exposure, Service Publication, Service Discovery, P2P, Convergence, Service Composition

1. INTRODUCTION
Service composition, which enables the creation of innovative services by integrating existing network resources and services, has received considerable attention in these years. This is partly because of its promising advantages such as cost-effective, reducing time-to-market, and improving user experience. As a prerequisite for service composition, service exposure, including service publication and discovery, plays a critically important role in this novel service ecosystem. That is because enabling a service to be reusable implies that this service needs to be known and accessible by users and/or developers. Various solutions to service exposure have been proposed in the last decade. However, these solutions suffer from several limitations, such as convergence, user-centricity, scalability, controllability, and cross-platform service information sharing. Taking these challenges into account, in this article, we propose a distributed and collaborative service exposure model that enables users to reuse existing services and resources for creating new services regardless of their underlying heterogeneities and complexities. This new method could lead to a new paradigm of service composition by providing the agile service discovery and publication support.

The rest of this paper is organized as follows. Section 2 introduces some backgrounds on service exposure, and the relevant mechanisms of service publication and discovery. Section 3 provides an overview of the distributed service exposure model. A mechanism of two-phases based service exposure, a triplex-layer based P2P overlay for service information sharing, the generation processes of different layers, and the process of service discovery relying on this triplex-layer based architecture, are described in Sections 4 to 7, respectively. The performance analysis is provided in Section 8. Finally, Section 9 concludes this paper by discussing the advantages of the proposed solution and mentioning our future work.

2. BACKGROUND AND RELATED WORK
Service exposure plays a significant role in the evolution of service composition, since enabling a service to be reusable means that this particular service needs to be known and accessible by users and/or developers. Various service exposure solutions have been proposed. For instances, using Simple Object Access Protocol (SOAP) or Representational State Transfer (REST) technologies for facilitating the invocation of services, exposing the services through open Application Programming Interfaces (APIs), and adopting service description and publication mechanisms such as Web Service Description Language (WSDL), Universal Description, Discovery and Integration (UDDI) as well as the more recent semantic annotation mechanisms [1][2][3], Web services has gained an immense popularity in a short time. Meanwhile, Telecom operators, menaced by IT competitors, are being forced to open up to both professional and non-professional users in order to retain or expand their service market sharing. Parlay/OSA Gateway, OMA Service Environment (OSE), Next Generation Service Interfaces (NGSI), and OneAPI are all specified by standardization bodies to allow access to Telecom services/capabilities by 'outside' applications through unified interfaces. Industrial solutions, such as British Telecom's Web 21c SDK, France Telecom's Orange Partner Program, Deutsche Telecom's Developer Garden, and Telefonica's Open Movilforum, are proposed by different operators. These operators aim to expose their network functionalities to 3rd party service developers and users by using Web based technologies. In

addition to these solutions, some industry alliances, such as the Wholesale Application Community (WAC), are formed by operators, vendors, manufacturers, and integrators in order to provide unified interfaces to access device functionalities, network resources, and services.

Most of the above current solutions (both in the Telecom and Web worlds) focus on using centralized systems for storing service information and managing service access. Such centralized systems are easy to implement and to monitor their resources. However, they are also the easy targets for malicious attacks, introduce a single failure point, require a high maintenance cost, and cannot be scalable and extensible enough to deal with a dynamic service environment. Moreover, centralized systems limit the interoperation among different platforms. This leads to many isolated service islands, which inevitably incurs resource redundancy.

Overcoming above-mentioned limitations of centralized solutions, decentralized P2P appears to be an obvious choice to support distributed service exposure on a large scale. Recently, some P2P-based solutions have been proposed for enabling distributed service discovery and publication on a large scale. For example, one solution is to group peers into clusters according to their similar interests. HyperCup [4] is an early example. According to predefined concept ontology, peers with similar interests or services are grouped as concept clusters. These concept clusters are then organized into a hypercube topology in a way that enables to route messages efficiently for service discovery. METEOR-S [5] attempts to optimize registry organization by using a classification system based on registry ontology. In this approach, each registry maintains only the services that pertain to certain domain(s). Acting as peers in an unstructured P2P network, the different registries are further grouped into clusters, according to mappings between registries and the domain ontology. Similar solutions classify either registries or service providers, which act as peers in an unstructured P2P network. These peers further form a federation that represents similar interests in a decentralized fashion [6 – 9]. As we can observe from these examples, federation-based solutions are generally related to unstructured P2P architectures. Unstructured P2P networks still have some common issues such as high network traffic, long delay, and a low hit rate, even if the available solutions have addressed these issues to a certain extent. Other alternative solutions based on a structured P2P system have also been proposed. SPiDer [10] employs ontology in a Distributed Hash Table (DHT) based P2P infrastructure. Service providers with good (better) resources in SPiDer, selected as super nodes, are organized into a Chord ring to take on the role of indexing and query routing. Chord4S [11] tries to avoid the single failure point in Chord-based P2P systems by distributing functionality-equivalent services over several different successor nodes. As a hybrid service discovery approach, PDUS [12] integrates the Chord technology with the UDDI-based service publication and discovery technology. Authors of [13] and [14] introduce solutions based on alternative structured P2P topologies such as Kautz Graph-based DHT or Skip Graph. These solutions generally use peers to form a Chord ring that stores all the service information about functionally similar services. As such, these peers are required to have good resources, such as high availability and high computing capacity.

From the literature, we find that most of P2P based solutions are based on unstructured P2P architecture due to its simplicity, robustness, and wide application in the current Internet domain. This is also because unstructured P2P can easily support complex requests, which is an essential requirement for service discovery. Unstructured P2P architectures, however, face challenges such as high network traffic, long delay, and a low hit rate. Because of these limitations in unstructured P2P, structured P2P is naturally selected to address certain issues. Nevertheless, one problem of service discovery based on structured P2P architectures, especially for the most widely used DHT-based solutions, is that the search is deterministic. This implies users must provide at least the exact name information about the resource they want. Such a requirement conflicts with a real-life situation, in which users often have only a partial or rough description of the service they want.

Taking the limitations of both central solutions and the current distributed solutions into account, we propose an enhanced model of P2P based service publication and discovery to improve the efficiency of service discovery on a large scale. This solution reuses the concept of Semantic Overlay Network (SON), which was originally proposed in [15] for the content sharing among nodes. For improving the efficiency of service discovery further, it is extended as a triplex-layer based architecture. This new solution thus enables the service publication and discovery in a purely distributed and collaborative environment.

3. AN UNSTRUCTURED P2P BASED SERVICE EXPOSURE MODEL: OVERVIEW

In our proposed solution, not only traditional Telecom and Web services are exposed, but device-offered services, and user-generated services are also accommodated. When different kinds of services expose themselves to a network, they generally use different technologies, or go through the different service exposure gateways. In order to reuse these existing service exposure technologies and limit the modifications to existing service publication and discovery platforms, we propose an unstructured P2P based architecture. The architecture uses a P2P overlay to share diverse services information, while respecting and maintaining the underlying heterogeneities, as shown in Figure 1.

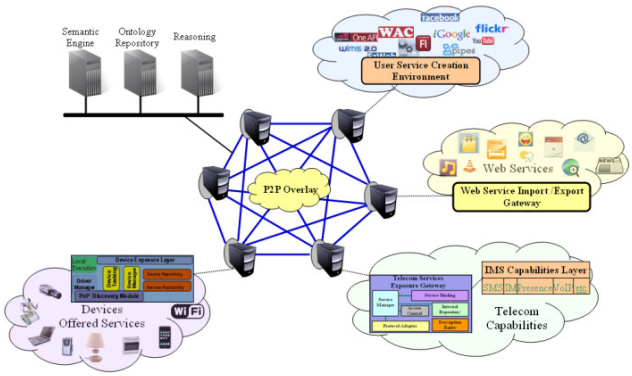


Figure 1. Overall architecture for the P2P based service information sharing system

As shown in Figure 1, this system is composed of several global servers and a number of nodes. The global servers include Semantic Engine, Ontology Repository, and Reasoning Engine, which are in charge of coordinating the representation forms of

the different descriptions of services during a service publication process. The nodes are responsible for the service information storage, service retrieval, service binding, service accessing, and service discovery. Behind these nodes, there are various service exposure gateways for catering to the different kinds of service exposure requirements. Examples are Telecom Service Exposure Gateway, Web Services Import/Export Gateway, Local Device Exposure and Management Gateway, which we introduced in [16], and even some service creation platforms which embed the functionality of enabling personal service publication by end users. These service exposure gateways can be the existing ones that are used in traditional UDDI-based centralized solutions. Within our proposed architecture, however, they are more flexible than what they are used alone in a centralized environment. The reason for this is that the gateways expose their services in a local domain to target a certain group of users or developers using their own technologies; moreover, they also share their service information with users or developers outside their domain by adding some overlay functions into their platforms.

4. LOCAL SERVICE EXPOSURE AND GLOBAL SERVICE EXPOSURE

In our solution, we divide the process of service exposure into two phases: Local Service Exposure and Global Service Exposure.

4.1 Local Service Exposure

After creating a service, service providers or users should make it to be accessible by other users through either the user-centric interface or APIs. They need to generate a service description file of this service in order to be discovered and understood by others. The service description file can be published in a global UDDI registry as a centralized solution does, or published in a local registry residing in the service platform. The latter case is related to Local Service Exposure.

Each service platform contains an internal service repository which contains the service description files for the services held on it. Once a new service is introduced, a service description file is generated. This service description file is then stored in the internal service repository. When a user wants to discover a service or create a personal service through this service platform, she/he can use the local facilities, and search in the local service repository by using the specific mechanisms of service discovery proposed by this service platform. In this context, a great number of local repositories are distributed over the network. They are managed by the respective service providers or service platform providers using the different technologies they prefer. The real situation is that these local registries are generally independent of each other and designed for a specific group of developers and/or a specific network. When a developer or a user wants to create a converged service that involves Telecom, Web, or device-based services, she/he has to search for services from several different service platforms, understand the different service description patterns, as well as their underlying heterogeneities. The need to manipulate different platforms increases the complexity of integrating the heterogeneous services into a unified composite service.

To improve user experience and enhance the reuse of existing services, a mechanism on efficient collaboration among the independent service registries is considered as necessary. Addressing this requirement, we add a complement process to service exposure, which enables seamless collaborations among different service registries. That is the Global Service Exposure.

4.2 Global Service Exposure

Global Service Exposure relies mainly on the Service Exposure Gateway for reporting the internal service information to the large scale P2P based network. That is to say, we assume that a Service Exposure Gateway is added to each service platform. It is used to connect the internal services with outer world application, expose them to external users or applications, and enable sharing its stored service information with other service platforms through the P2P method. An example of such a Service Exposure Gateway is shown in Figure 2.

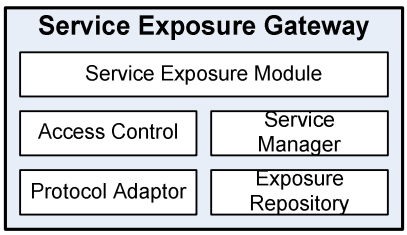


Figure 2. An example of Service Exposure Gateway

When a new service is introduced to a service platform, its service description file is added into the internal repository as we mentioned in Section 4.1. If the provider of this service platform wants to share it with other service platforms, the service description file is sent to the local Service Exposure Gateway. Service Exposure Gateway then contacts the Global Semantic Engine for translating the service description file to a globally-comprehensible format, and for adding the necessary semantic annotations into the description file by consulting with the global Ontology Repository. After that, this unified and semantic enriched service description file is added into the Exposure Repository. Meanwhile, some access control rules and protocol adaptation rules associated with this particular service are also added into the relevant components by the service provider.

For a service publication process, we further introduce two kinds of publications: abstract service publication and concrete service publication. An abstract service is a virtual service that only contains the generalization information about one kind of services (e.g., SMS service). One abstract service can be mapped into several concrete services (e.g., Orange SMS service, Telefonica SMS service and BT SMS service). The abstract service publication is handled by the members of the system administration group. They send a service description file that contains only the generalization information of one type of services to Semantic Engine. Semantic Engine then contacts Ontology Repository for supplementing the necessary semantic annotations. The transformed service description file is then sent to a global Semantic Abstract Service Repository. The concrete service information is published in the local Service Exposure Gateway, as we introduced in the generation process of service description file in the local Service Exposure Gateway.

In order to make concrete services discoverable by other service platforms, each gateway needs to map its concrete services into the corresponding abstract services, create a Local Abstract Service Table, and store this table in the Service Exposure Module. After this, the number of gateways are then interconnected each other through an underlying P2P network in

which they act as peers. This process is considered as Global Service Exposure.

5. A TRIPLEX-LAYER BASED SOLUTION FOR SERVICE INFORMATION SHARING IN P2P ENVIRONMENT

As introduced before, diverse service description files are published in the respective registries residing in the Service Exposure Gateways. The registries share their service information through the P2P pattern for avoiding the single failure point, improving the service discovery efficiency, and reducing the maintenance cost.

In order to improve the query performance and maintain a high degree of node autonomy, a set of Semantic Overlay Networks (SONs) are created over distributed gateways. Essentially, each gateway is connected to a global network as a peer in a P2P system. These gateways share the service information stored in their Exposure Repository through Service Exposure Module. As the description files of services are semantically enriched, the peers that hold these semantic description files can be regarded as semantic-enriched as well. Nodes with semantically similar service description files are then "clustered" together. Consequently, the system can select the SON(s) that is (are) in the better position to respond when a user wants to discover a service. The query is then sent to one of the nodes in the selected SON(s), being further forwarded to the other members of the SON(s). In this way, the query performance is greatly improved, since queries are only routed in the appropriate, selected SONs. This would increase the chance of matching the files quickly with a limited cost.

To implement the above-mentioned SON based semantic service discovery, we propose a triplex-overlay based P2P system as shown in Figure 3.

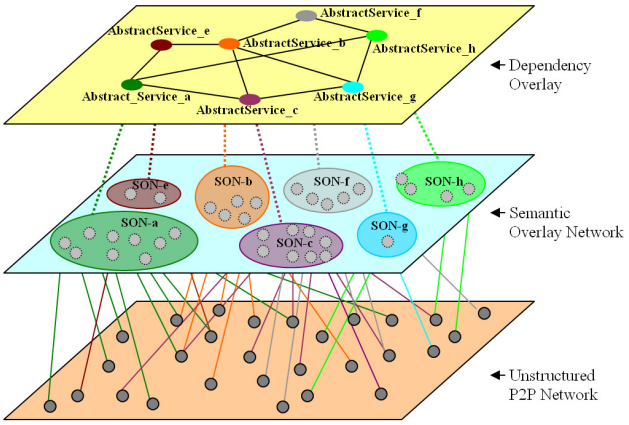


Figure 3. Triplex overlay for P2P based service discovery

In Figure 3, the diverse network repositories, service discovery and publication platforms, service creation environments, and device gateways join the P2P based distributed network. Acting as nodes in an Unstructured P2P Network layer, they interact with each other using blind search solutions (e.g., Flooding based solutions or RandomWalk based solution).

In the above unstructured P2P network, the nodes providing similar resources are clustered together. We call this kind of clusters as SON (Semantic Overlay Network). The benefits of this strategy are as follows. On the one hand, a request is sent to the nodes with a high probability of offering target services, so that the request can be answered quickly. On the other hand, the nodes with a low probability will not receive the request. A waste of resources can therefore be avoided in transferring the request, so that other relevant requests are allowed to be processed in a quick way.

For further improving the efficiency of service discovery, an additional Dependency Overlay is introduced upon the SON layer. This is because different kinds of services may be able to interoperate with each other. For example, the output of one service can be the input of another. Such cooperation among services allows service providers/developers to provide some more meaningful services to end-users. Thus we can infer that the services stored in the same gateway have a very high probability of having certain interdependency amongst each other. The service dependency relationship can also be specified according to some social network information, such as service usage frequencies, users' service usage habits, or some network statistic data. Consequently, defining the dependency among the abstract services, then providing recommendations for the message routing during a service discovery process, will improve the success rate. This definition of service dependency is not limited to one kind of services, but rather includes the dependency among the different kinds of service, such as the devices offered services, Telecom services, and Web services, e.g., the dependency of "camera -> MMS". As the publication of abstract services is performed by the members of the system administration group, the dependency relationship is created simultaneously, once a new abstract device or abstract service is published.

6. SON LAYER GENERATION

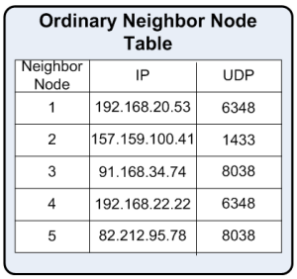
We assume that a set of bootstrap nodes can guarantee the minimal running requirements for the proposed system. These bootstrap nodes form a small scale SON overlay before running the system. That is to say, when an abstract service is introduced to a network by a system administration member, this new abstract service is assigned to a bootstrap node randomly.

Each gateway contains a table called Local Abstract Service Table. This table is created by mapping concrete service description files stored in a local exposure repository, into abstract service profiles stored in the global Abstract Service Profile Repository. In this table, each abstract service entry contains the basic information about the relevant concrete services (e.g. concrete service's name), as well as the links (e.g., a URL) to the corresponding concrete service description files. Based on Local Abstract Service Table, a local gateway can join the relevant SONs automatically. Since several abstract services are contained in one registry, one gateway can join several SONs according to the different abstract services.

To clarify the SON generation process and the update process, we consider two cases in the following: (1) a new gateway is added into the network with a list of abstract services to be exposed; and (2) an existing gateway updates its list of abstract services, which means a new type of service has been introduced to its local network.

When a gateway is introduced to the system, it first joins in the global network through some bootstrap nodes as the ordinary P2P networks do. That is, once receiving the *Join* message from a gateway, the selected bootstrap node broadcasts the *Join* message

to the global network. According to certain neighborhood selection rules (e.g., the solutions introduced in [17] and [18]), some nodes are selected as the logic neighbors for this newly introduced registry and added into the Ordinary Neighbor Node Table (ONNT) as shown in Figure 4. In this example, we assume that each node contains information about 5 neighbor nodes (e.g. IP and UDP port). This process provides another possible way to search a service: if both the SON Overlay based search and the Service Dependency Overlay based search have not found out the relevant services, the system can use the basic Random Walk or Flooding solution according to the information of neighbor nodes stored in the Ordinary Neighbor Node Table.

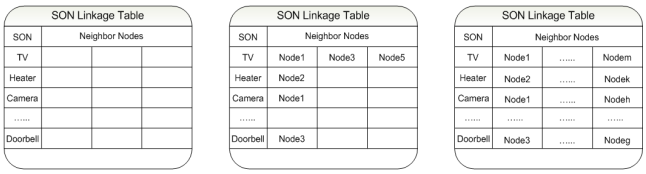


Ordinary Neighbor Node Table

Neighbor Node	IP	UDP
1	192.168.20.53	6348
2	157.159.100.41	1433
3	91.168.34.74	8038
4	192.168.22.22	6348
5	82.212.95.78	8038

Figure 4. An example of Ordinary Neighbor Node Table

We assume that each gateway contains another table called SON Linkage Table (SLT) as shown in the Figure 5, which is used to form the SONs. When a gateway joins the network for the first time, this SLT is empty as shown on the left side of Figure 5. This means the gateway has not joined any SON yet. After joining the unstructured P2P network, this new gateway extracts the names of the abstract devices from its Local Abstract Service Table, and encapsulates them as an *Update* message. This *Update* message is injected to the network by the blind search method according to the neighbor nodes' information stored in the Ordinary Neighbor Node Table. In particular, we use the Random Walk as the basic message routing approach, in which only one neighbor node is selected from the Ordinary Neighbor Node Table for routing the *Update* message. The names of the extracted abstract services in the *Update* message are put into a list, with each entry assigned a time-to-live (TTL) called Time-to-Live for Node (TTL_node). TTL_node is a pre-assigned number *n,* which indicates how many neighbors in the relevant SON have to search for. A global time-to-live called Time-to-Live for Hop (TTL-hop) is also attached to the *Update* message for setting the maximum number of the logic hops for the message.



SON Linkage Table

SON	Neighbor Nodes		
TV			
Heater			
Camera			
.....			
Doorbell			

SON Linkage Table

SON	Neighbor Nodes		
TV	Node1	Node3	Node5
Heater	Node2		
Camera	Node1		
.....			
Doorbell	Node3		

SON Linkage Table

SON	Neighbor Nodes		
TV	Node1	.....	Nodem
Heater	Node2	.....	Nodek
Camera	Node1	.....	Nodeh
.....	.....	.....	.....
Doorbell	Node3	.....	Nodeg

Figure 5. An example of SON Linkage Table in a device gateway

The *Update* message is first forwarded to one of this gateway's neighbors, denoted by *Ni*, and the node that receives this message checks its own SLT. If this table has no entries containing the same name of the abstract service contained in the *Update* message, this request is directly forwarded to another neighbor of *Ni*. As such, only the TTL-hop decreases by 1. Otherwise, if one or several entries in the table contain the same name(s) as the abstract services listed in the *Update* message, both the TTL-node(s) for the relevant abstract service(s) and the TTL-hop decrease by 1. The node information (e.g. IP and Port number) is sent back to the original node. The original node then stores the node information in relevant entries of SLT as a neighbor, as shown in the middle table of Figure 5. After this, the *Update* message is forwarded to one ordinary neighbor of *Ni* node. The above process repeats until either all the TTL-nodes or the TTL_hop have reached 0. If all the entries in the SLT have been populated with information of SON neighbors, the join process for SON is regarded as a successful one; otherwise, the original node selects one of its other neighbors to send the same *Update* message. If all the neighbors of this original node have been selected for forwarding this *Update* request, and some entries in the SLT are still not completely filled, then the original node sends the *Update* message to its bootstrap. This connected bootstrap searches in the bootstrap nodes for finding out which bootstrap node(s) is responsible for the corresponding abstract service(s), and sends information of the relevant bootstrap node(s) back to the original node. The original node adds the information into the relevant entries in the SLT and marks it as bootstrap node's information. This process guarantees that if such an abstract service has been predefined in the network, even the relevant SON has not been created yet; the original node can create a new SON, or join in an existing SON by connecting itself to the bootstrap node directly. Once another neighbor in the same SON is found, the information of the relevant bootstrap node is replaced by the newly found neighbor. This process is repeated on a regular basis to make sure that all the information stored in the SLT is up-to-date. In other words, once either the information of a neighbor node becomes obsolete, or a new abstract service is added into a gateway, a new *Update* message with the names of relevant abstract devices is sent out again to the network and a new join SON process begins.

7. SERVICE DISCOVERY BASED ON TRIPLEX-LAYER BASED P2P SYSTEM

Before introducing service discovery based on triplex overlay, we classify the neighbor nodes of a node into two kinds: the ordinary neighbor nodes and the SON neighbor nodes.

When a user wants to discover a service for the purpose of executing it directly or integrating it into her/his own created composite service, she/he can issue a service discovery request through a user-friendly service discovery front-end or a user-centric service creation environment. We assume that these service discovery and creation frameworks have connected themselves to the global network in advance.

Once a user issues a service discovery request either by entering keywords or natural language based text, or even by selecting the abstract service building blocks, the relevant frameworks formalize a request that contains the target service information (e.g., a *get(target_service, TTL, auxiliary_info)* message). By contacting with the global Semantic Engine, the global Ontology Repository, as well as the global Abstract Device Repository, this newly created request is interpreted into a system recognizable format. The request is injected into the unstructured P2P network through the node that initiates this request. To facilitate the request interpretation process, the relevant service discovery or

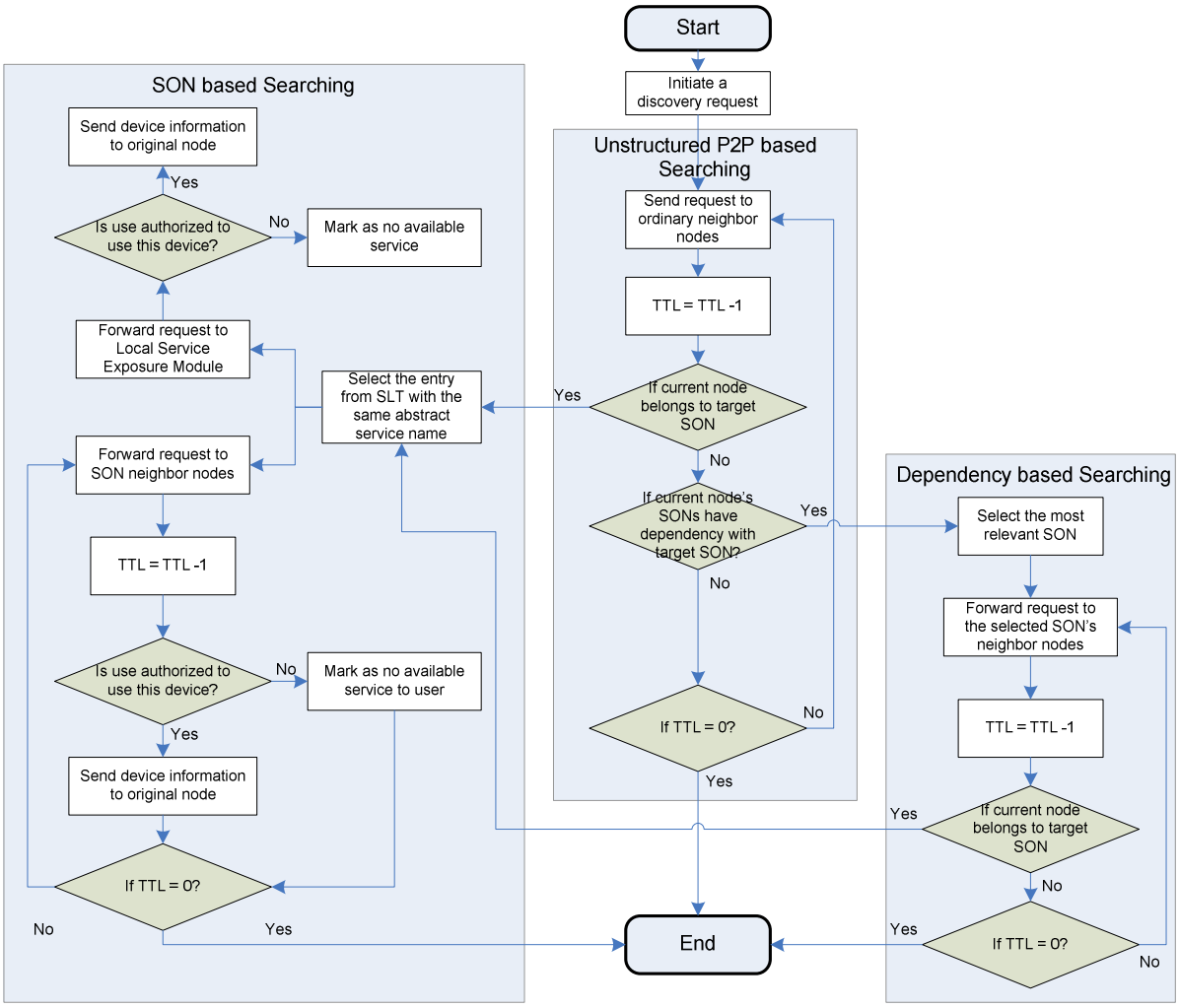


Figure 6. Flowchart for service discovery process in triplex-layer based service exposure model

service creation environment can install a local Semantic Engine and Ontology Repository, and store the abstract service description files in its local repository. In this case the request can be formalized automatically inside its local framework before being injected into the P2P network.

The *get()* request is injected into the network through the blind search method. That is to say, the node, which initiates this request, forward the request to either all its ordinary neighbor nodes or one of its ordinary neighbor nodes, which mainly depend on which blind search strategy (e.g., Flooding or Random Walk) the system adopts. To improve the service discovery efficiency, the system first needs to discover the corresponding SON for the message routing. As shown in Figure 6, the first nodes that receive the service search request will verify if the target abstract service is contained, by comparing the entries' name stored in their Local Abstract Service Tables with the target service's name indicated in the incoming request. If matched, the entry with the same abstract name in the SLT is selected. The received request is then forwarded to the SON neighbor nodes whose IP and Port information is stored in this selected entry. Meanwhile, this node extracts the context information (e.g. user identifier, location, preference, etc.) encapsulated in the received message, and forwards this context information, together with the name of the target abstract service, to its own local Service Exposure Module. Service Exposure Module contacts with Exposure Repository for checking if any services stored in its local gateway can be used by the user, who issues this service discovery request. It then makes a primary filtering for the discovered result according to the auxiliary context information. For example, if a service is set to be public (e.g. a map service), a user can discover this service once the request reaches this gateway. If a service is set to be private, even it matches the functional requirement for the target device, this gateway, however, still needs to verify if the user identifier in the request is included in the identifiers that have been granted by the owner of this gateway for the use of this service (e.g. the user who has subscribed to this service platform). If the original user identifier is matched, the relevant service description file is sent back to the service discovery framework. Otherwise, no service description file is sent back, and this node is marked as no resource matched to user's request.

When the SON neighbor nodes of this node receive the forwarded request, as these nodes are already in the same SON, and this SON is the target SON in which all the nodes contain the service associated with the target abstract service, they thus invoke local Service Exposure Module for local service discovery directly without the need to verify if any SON they belong to match the

target abstract service. At the same time, they forward the request to their SON neighbor nodes, which are extracted from the entry in their SLT. This process continues until the TTL for this request is reduced to 0, which means the service search is terminated.

If the first few nodes that receive the service search request do not belong to the SON for the target service (e.g. Camera), this means that no entry name in Local Abstract Service Table is the same as the abstract name of the target service. In this case, this node needs to discover which neighbor nodes are the most possible ones that in turn find out the relevant SON to which the target service belongs. It is assumed that each gateway contains a copy of such an abstract service dependency relationship file. In order to select the most possible neighbor to match the target SON, the node that does not belong to the target SON extracts all names of the abstract services from its Local Abstract Service Table and the name of the target abstract service from the incoming message. Relying on the dependency file of abstract services, the node then analyzes services dependency relationships among these extracted abstract services, and selects the abstract service whose dependency intensity with the target abstract service is the highest one.

According to the name of the selected abstract service, the node selects the SON neighbor nodes from an entry in SLT whose entry's name is the same as the name of the selected abstract service. The service search request is then forwarded to these selected SON neighbor nodes. When these SON neighbor nodes receive the service search message, they first check their Local Abstract Service Table to see whether they contain the target abstract service in their local gateways or not. If a node finds that it contains the target abstract service, it searches its SLT, and selects the SON neighbor nodes from the entry whose name is identical to the target abstract service. The node then forwards the service search request to the corresponding SON. Finally the service search is limited to the SON that is responsible for the target abstract service. If a node cannot find the target service, it forwards the request to their neighbor nodes that are in the same SON as the first contacted node selected.

If the first contacted node(s) does not belong to the target SON, and it also has no dependency relationship with the target abstract service, the request in the first contacted node(s) is forwarded to its ordinary neighbor node(s) following the blind search method adopted by the underlying unstructured P2P network. When its ordinary neighbor node(s) receives this service discovery request, it repeats the service discovery process we introduced above. This process repeats until either the target SON or a dependent SON has been found out, or the TTL is time out.

8. PERFORMANCE ANALYSIS

In order to evaluate the performance of the proposed service exposure and information sharing model, the theoretical analysis on the success rate and the network traffic impact have been performed. We also compare it with the traditional blind search strategy for the unstructured P2P solutions (which are based on Flooding or Random Walk), as well as the pure SON based solution.

In the following, we use the average messages required for discovering a service as a metric for evaluating our proposed system. This metric can be regarded as an indicator of the invoked network traffic for a service discovery request. In ideal cases, in which only one message exchanges between two neighbor nodes, then the value of this metric is equal to the average number of nodes that are involved in the search process.

To simplify the analysis, we denote the success rate as S(T), and the average message as E(S). We assume that services are distributed in the network evenly and their probabilities being discovered p are the same if the blind search solution is used. The SONs are also assumed to be evenly distributed in the network, and the difference in their sizes is ignored. When a request arrives at a node, the probability that the connected node belongs to the target SON is R. If this node does not belong to the target SON, Dependency Layer makes a recommendation that has the higher probability of finding out the target SON in the next hop. This increased probability is denoted as K. In general, we have K>R. If the request arrives at the target SON, its probability of discovering the target service (q) is much higher than that of searching in the underlying unstructured layer by using the blind search strategy (p). That is because, when a request arrives at a SON, the search space is reduced. But the total number of the user accessible services within this SON remains unchanged, thus the local discovery probability is much higher. Moreover, it can guarantee that all the target resources are clustered into this SON, thereby the nodes outside this SON can be out of consideration.

In the following, the relationships between the success rates and the numbers of the visited nodes for the three solutions of Triplex, SON, and Blind, and their comparisons are illustrated by simulation results with varied values of p, R, q, and K.

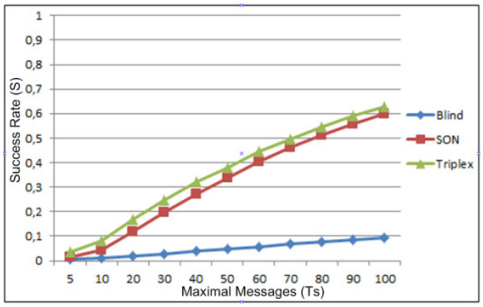


Figure 7. Comparison for the success rates of three types of P2P systems (p=0.1%, R=10%, q=10%, K=50%)

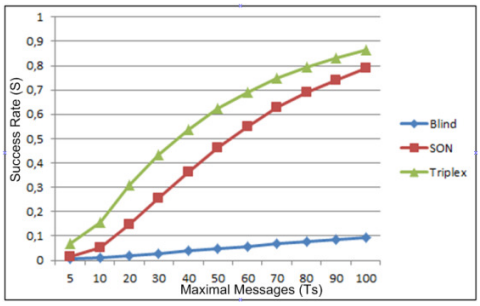


Figure 8. Comparison for the success rates of three types of P2P systems (p=0.1%, R=5%, q=20%, K=50%)

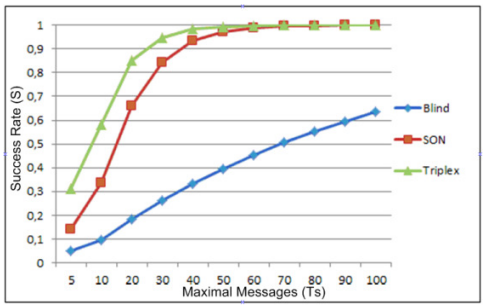


Figure 9. Comparison for the success rates of three types of P2P systems (p=1%, R=10%, q=10%, K=50%)

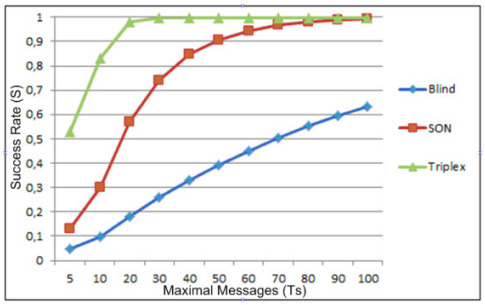


Figure 10. Comparison for the success rates of three types of P2P systems (p=1%, R=5%, q=20%, K=50%)

The simulation (Figure 7 to Figure 10) results show that our proposed solution can greatly improve the success rate with the limited number of messages transferred among the nodes. This improvement is more evident when the popularity of services is low in a network.

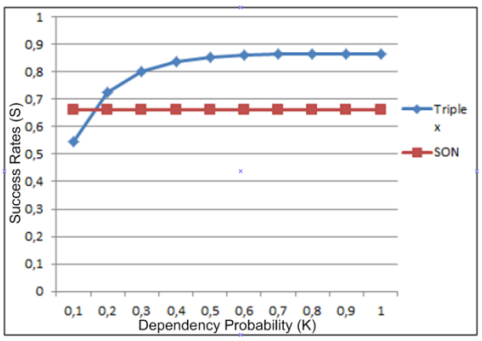


Figure 11. Comparison of the success rates (S) by modifying the dependency probability (K)

Figure 11 illustrates the degree of the influence of the dependency probability *K* on the success rates. The red line with rectangles is the success rate of ordinary SON based system without dependency recommendation, while the blue one is for the SON based system with dependency recommendation.

From the results shown in Figure 11, we can observe that as the dependency probability increases, the success rate increases accordingly. However, after K is more than 0.5, this increase rate becomes very slow. Moreover, only when K is bigger than a certain level (in our case, when R=10%, K needs to be bigger than 15.3%), our solution can improve the service discovery performance of the system; otherwise it may weaken the performance.

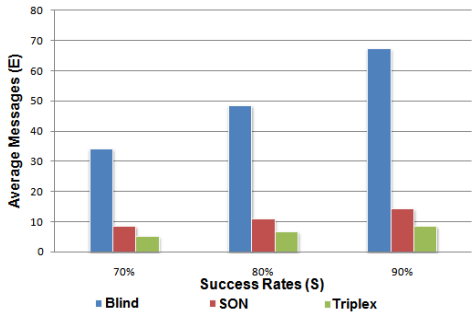


Figure 12. Average messages needed for achieving the success rates of 70%, 80%, and 90%

Figure 12 illustrates the average number of messages needed for finding out the first target service. It shows that our proposed solution can reduce the average messages, with a high success rate.

The above simulation results have demonstrated that the use of SON layer and dependency layer for routing the service discovery request are able to greatly increase the success rate and to reduce network traffic at the same time. From the simulation results, we also note that the classification of SONs should be appropriate. If too many types of SONs are in a network, on the one hand, the probability of locating the target SON will be reduced. On the other hand, once a request reaches the target SON, the probability of discovering a service within this SON will be higher than that in the network with few types of SONs. The reason for this is that more SONs in the network, the search space within each SON will be smaller. Thus we need to find a solution that balances these two parameters. The use of the dependency analysis for the SON selection process is one of the possible solutions. It can keep the high service discovery probability within a SON, meanwhile it makes efficient recommendations for the SONs selection. This aims at resolving the problems incurred by the large number of SONs. How to define such dependency rules is challenging in our proposed triplex layer based model. It is also one of our future research topics.

9. CONCLUSION

This paper has presented a model of P2P based distributed service exposure. In this model, the service exposure process is mainly divided into two main phases: local service exposure, which exposes local services to a gateway for local monitoring, and global service exposure, which enable the services to be discovered by external parties. Considering the factors of a great number of disperse gateways in a network, the single failure point for the centralized UDDI likes solution, and the possible bottleneck for the certain level number of services, we use a P2P based distributed method to enable the interoperation among the gateways. For further improving the efficiency and accuracy of service discovery, a triplex overlay based solution has been proposed. This triplex overlay architecture includes an unstructured P2P layer, a SON based overlay, and a service dependency overlay. The performance analysis and simulation results have shown that our proposed triplex overlay based solution can greatly improve the system performance by increasing the success rate of service discovery and reducing the number of forwarded messages needed for achieving certain level of success rate.

The proposed model improves a service exposure system in that it respects the diversity, autonomy, and interoperability of service exposure platforms. Each platform of service exposure can use its preferred techniques for exposing its services to users. By introducing a P2P based service information sharing system, the model enables the service information to be shared among different operators, service providers, and users regardless of their underlying heterogeneities. Telecom or Web service exposure platforms, Telecom or Web service discovery and publication platforms, or even some service creation environment can join this system as a peer, for example. And a user can discover a service provided by her/his service discovery platform provider through its specific technologies, or by other service discovery platform through the P2P based service information sharing system. The model also makes it possible to apply enhanced controls for service exposure in accordance with the requirements of different operators or service providers via gateways. The triplex-layer based P2P architecture can enhance the system scalability, and improve the service discovery and publication efficiency in the unstructured P2P based system. These improvements for the unstructured P2P based system also avoid the irrelevant nodes, which are responsible for other services, to be disturbed during a service discovery process. These nodes can thus reserve their resource for other relevant tasks.

Further work will focus on defining a widely-usable service description data model, and developing the more efficient strategies for service selection. The use of semantic, ontology, and social network information for providing service recommendations is also left for further study.

10. REFERENCES

[1] Semantic Annotations for WSDL and XML Schema, DOI=http://www.w3.org/TR/sawsdl/

[2] Web Service Semantics - WSDL-S, DOI=http://w3.org/2005/04/FSWS/Submissions/17/WSDL-S.htm

[3] OWL-S: Semantic Markup for Web Services, DOI=http://www.w3.org/Submission/OWL-S/

[4] Schlosser, M., Sintek, M., Decker, S., and Nejdl, W. 2002. HyperCuP—Hypercubes, Ontologies and Efficient Search on P2P Networks, In *Proceedings of 1st International Conference on Agents and Peer-to-Peer Computing (AP2PC '02)*, pp. 112-124, Bologna, Italy, July 2002

[5] Verma, K., Sivashanmugam, K., Sheth, A., Patil, A., Oundhakar, S., and Miller, J. 2005. METEOT-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services, *Information Technology And Management*, Vol. 6, pp. 17-39, Jan. 2005

[6] Banaei, K.F., Chen, C.C., and Shahabi, C. 2004. Wspds: Web services peer-to-peer discovery service. In *Proceedings of 5th International Conference on Internet Computing (IC '04)*, pp. 733-743, Las Vegas, USA, Jun. 21-24, 2004.

[7] Modica, G., Tomarchio, O., and Vita, L. 2011. Resource and Service Discovery in SOAs: A P2P Oriented Semantic Approach, *International Journal of Applied Mathematics and Computer Science*, Vol. 21, pp. 285-294, Jun. 2011

[8] Li, R., Zhang, Z., Wang, Z., Song, W., and Lu, Z. 2005. WebPeer: A P2P-based System for Publishing and Discovering Web Services, In *Proceedings of IEEE International Conference on Services Computing (SCC '05)*, pp. 149-156, Orlando, USA, Jul. 11-15, 2005

[9] Papazoglou, M.P., Kramer, B.J., and Yang, J. 2003 Leveraging Web Services and Peer-to-Peer Network, in *Proceedings of 15th International Conference on Advanced Information Systems Engineering (CAiSE '03)*, pp. 485-501, Klagenfurt/Velden, Austria, Jun. 16-20, 2003

[10] Sahin, O.D., Gerede, C.E., Agrawal, D., Abbadi, A.E., Ibarra, O., and Su, J. 2005. SPiDeR: P2P-based Web Service Discovery", in *Proceedings of 3rd International Conference on Service Oriented Computing (ICSOC '05)*, pp. 157-169, Amsterdam, Holland, Dec. 12-15, 2005

[11] He, Q., Yan, J., Yang, Y., and Kowalczyk, R. 2008. Chord4S: A P2P-based Decentralised Service Discovery Approach, *In Proceedings of IEEE International Conference on Service Computing (SCC '08)*, pp. 221-228, Honolulu, USA, July 7-11, 2008

[12] Ni, Y., Si, H., Li, W., and Chen, Z. 2010. PDUS: P2P-base Distributed UDDI Service Discovery Approach, In *Proceedings of International Conference on Service Sciences (ICSS '10)*, pp. 3-8, Hangzhou China, May 13-14. 2010

[13] Zhang, Y., Liu, L., Li, D., Liu, F., and Lu, X. 2009. DHT-based Range Query Processing for Web Service Discovery", in *Proceedings of IEEE International Conference on Web Services (ICWS '09)*, pp. 477-484, Los Angeles, USA, July 2009

[14] Zhou, G., Yu, J., Chen, R., and Zhang, H. 2007. Scalable Web Service Discovery on P2P Overlay Network, in *Proceedings of IEEE International Conference on Services Computing (SCC '07)*, pp. 122-129, Salt Lake, USA, July 2007

[15] Crespo, A., and Garcia-Molina, H. 2005. Semantic Overlay Networks for P2P Systems, *Agents And Peer-to-Peer Computing*, Vol. 3601/2005, pp. 1-13, 2005

[16] Huang, C., Lee, G.M., and Crespi, N. 2012. A Semantic Enhanced Service Exposure Model for Converged Service Environment, *IEEE Communications Magazine*, pp. 32-40, vol. 50, Mar., 2012

[17] Beverly, R., and Afergan, M. 2007. Machine Learning for Efficient Neighbor Selection in Unstructured P2P Networks, in *Proceedings of Second Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML'07)*, Cambridge, MA, Apr. 10, 2007

[18] Liu, H., Abraham, A., and Badr, Y. 2010. Neighbor Selection in Peer-to-Peer Overlay Network: A Swarm Intelligence Approach, in *Pervasive Computing, Computer Communications and Networks*, pp 405-431, 2010

TV widgets: Interactive applications to personalize TV's

Rodrigo Illera and Claudia Villalonga
LOGICA Spain and LATAM
Avenida de Manoteras 32
Madrid, Spain
+34 91 304 80 94
{rodrigo.illera, claudia.villalonga}@logica.com

ABSTRACT
Interactive TV is a new trend that has become a reality with the development of IPTV and WebTV. Moreover, a huge number of applications that process data stored on the web are available nowadays. Interactive TV can benefit from content and user preferences stored on the Web in order to provide value added services. The rendering of personalized news on the TV is one of these new Interactive TV applications. TV Widgets can present personalized news to the user by combining available online resources. This paper presents the characteristics of TV Widgets and describes how mashups can be used to develop advanced TV Widgets. MyCocktail, a mashup editor tool used to develop TV Widgets, is described. The personalization of the TV Widgets using a recommendation system is introduced. Finally, the implementation of the components that allow the development of TV widgets is presented and an example widget is shown.

Categories and Subject Descriptors
I.7.1 WEB

General Terms
Documentation, Design.

Keywords
TV Widgets, Interactive TV, Mashups, Mashup editor, IPTV.

1. INTRODUCTION
TV information flow has been traditionally unidirectional, that is, information is broadcast from the TV station to terminals. No return channel is established. Hence, feedback from viewers had to be collected through other mechanisms like phone surveys or polls.

Currently, TVs have become more interactive. Content provision can be achieved by streaming video signal over IP networks, and interactivity is possible by establishing a return channel from users IP addresses. Viewers tended to be passive content consumers, but now users can interact actively with TV service providers. Interaction enables a wide range of applications like Social Media or Electronic Program Guide (EPG) just to name a few. Several value-added services can be provided to the user therefore new business models can be developed to exploit Interactive TV. All these services can be personalized according to users preferences. In order to do so, users are required to feed systems via IP-based return channels.

On the other hand, the last trends in Web 2.0 have derived in an overwhelming number of users storing preferences, networking information and personal data in the Cloud. Social networks are the most known example of users feeding a system in such way. One of the most interesting aspects of these systems is that they offer APIs to third parties, so they can access personal information to create value-added services (only when user's privacy settings allow it). A considerable number of applications have emerged based on massive processing of this data. This is done, for example, to find synergies and predict preferences for a determined user. This is the basis for valuable services like Personalized Advertising and Recommender Systems.

There is a huge potential in making interactive TV and content provision over IP networks benefit from users' personal information available on the Web. News is an example of relevant content that can be presented on TVs. It seems reasonable to use this kind of content to demonstrate the potential of reusing available resources (cloud services and sources of information) to create value-added interactive TV applications. A solution using this strategy is presented along this paper.

2. STATE-OF-THE-ART: IPTV AND WEBTV
Two different trends have emerged and flourished in the field of interactive TV: IPTV and WebTV. Both technologies rely on standard Internet Protocols to stream video through an ISP network. The main difference between them is that IPTV is a managed service running on a managed network, while WebTV is an unmanaged service running on an unmanaged network (best-effort delivery system). Non-operator-based IP protocol television is also known as "Over the top" (OTT).

Regarding output terminals, according to John Allen, CEO of Digisoft.tv, difference between IPTV and WebTV is that in IPTV video information is streamed over IP to a STB (Set-Top Box) while in Web TV the signal is sent exclusively to a PC [1].

Depending on the technology we are using to stream and consume video over IP networks (IPTV or Web TV) different considerations have to be taken into account [1]. The way users interact with terminals determines the way applications have to be designed in order to make interactivity a reality ensuring high-level quality of experience (see Table 1).

Table 1: Comparative of Interactive TV trends

	IPTV	**Web TV**
Terminal	STB + TV	PC
Type of interaction	Passive	Active

Information output	Continuous	Burst
Interaction required	Rarely	Often
Physical distance to terminals	Far (~ 2m)	Close (< 70cm)
Input mechanism	Remote control (D-Pad) Keyboard (sometimes)	Mouse + Keyboard

Browsing along a TV screen is usually done by tabbing with a D-Pad on a remote control. Thus, less freedom of movement is perceived when we compare this with interaction via mouse. In order to ease user's leap from passive to active, sometimes is necessary to simplify applications by removing advanced functionality for IPTV users (this has been referred not tactfully as "dumbing-down" TV applications).

Besides, there is also an issue concerning difference in output quality between computer monitors and TV's. Applications are developed in the first place on computers, but displaying high resolution computer graphics on a TV can be tricky. According to [3] the three main problems are interlaced flicker, illegal colors and reduced resolution. Designers have to take special care on these aspects in order to provide quality graphics on TV applications.

Concerning the input mechanisms, in the last years, consoles like Wii offer an alternative to the mouse: a remote with features that enables users interacting with items on the screen via gesture recognition and pointing through the use of accelerometers and optical sensor technologies [4]. However, this solution presents some shortcomings in usability (e.g. sometimes is not easy to aim on a determined item on the screen). Instead, it seems that interaction will be mainly done with remote controls. Hence, IPTV users won't be wandering the screen with their mouse pointers looking for links. Regarding text input mechanism, some set-top-boxes come with a remote device with a QWERTY keyboard included, which eases the way users insert text into TV applications.

Developing IPTV interactive applications is more challenging than Web TV applications as described above. Therefore, this paper focuses exclusively on IPTV technologies. However, all methodologies described can be applied to Web TV solutions as well.

3. TV WIGDGETS

Widgets are small-scale software applications providing access to technical functionalities, like for example a web service. Widgets are client-side reusable pieces of code that can be embedded into third parties (a web page, PC desktop, etc). By inserting widgets, users can personalize any site where such code can be installed. For example, a weather report widget could be embedded on a website about tourist attractions in Germany, providing the weather forecast for all regions. Despite of looking like traditional stand-alone applications, widgets are developed using web technologies like HTML, CSS, Javascript and Adobe Flash [5]. Furthermore, widgets can be created following different standards like W3C [6] or HTML5 [7]. Widgets are a technology that has become very popular because of its simplicity. That is so, that important companies like Google or Yahoo release their own official widgets so everyone can personalize whatever application they want by inserting them in between.

When it comes to the subject of Interactive TV, widgets present an interesting feature: they can be rendered on one side of the screen while the remaining area keeps displaying the video streaming signal. The main advantage is that users can interact with a widget without stopping watching TV. Ideally, widgets embedded on the screen won't disturb the viewer, who will always have the option to hide them. It's obvious that widgets provide added value to the TV respecting the main purpose of the terminal, which is displaying video. Widgets can be a solution for TV interactivity.

We define TV widgets as those widgets that are presented on the screen using IPTV only. That is, the terminal is a traditional TV combined with a STB, not a computer monitor (that would be the case of Web TV technologies). As described in Section 2, usability shortcomings compel designers to take special considerations. When it comes to the subject of TV widget development the most important considerations are:

- TV widgets area has to be limited. Furthermore one of their sides needs to be especially narrow in order to make room for TV programs still be running on TVs.
- TV widgets have to be located besides one border of the screen. Ideally, the viewer should not be disturbed by widgets while watching TV.
- Because of the distance between viewers and TVs, text and links rendered on the TV widget have to be big enough, even if the dedicated area is limited.
- Due to browsing shortcomings, TV widgets have to be easy and fast to navigate through. Ideally a few tabs with a D-Pad should be enough to explore and interact with TV widgets.
- Because of screen rendering, space and browsing limitations, TV widget functionalities have to be "dumbed-down" (not advanced or low-level features should be provided). The user has to be able to interact with the application only with a few steps.

In the case of study presented in this article, TV widgets are generated to access news content and provide rich applications complementing that content. The logic behind the TV widget will combine different resources available online to offer added value using the concept of mashups [8].

4. MASHUPS FOR TV WIDGET DEVELOPMENT

Mashups are small to mid-scale web application combining several technical functionalities or sources of information to create a value-added service. The main idea behind this concept is that external data available in the Web can be imported and used as input parameters of a service. In comparison to big applications using heavy frameworks, mashups can be developed easier and faster [9].

A mashup architecture contains three main elements:

- Sources of information: personal data or content that is available online and can be accessed through APIs. Many technologies can be considered e.g. REST [10], SOAP [11], RSS [12], ATOM [13]. Another kind of sources of information could be input parameters inserted by the user through some application field.
- Logic: many services and operators can be used to process the data retrieved from the sources of information. The logic can reside on the client, on the server, or be distributed among both sides.
- Visualization: the outcome of the mashup logic has to be rendered on a screen (generally a PC, but in our case on a TV). Rendering relies strongly on the technologies used on the client side. In our case of study, visualization will be done on TV widgets whose language will depend on the underlying platform (e.g., Yahoo, Google).

Mashups became a revolution due to many factors: the availability of content resources, the increase of the number of APIs and services, the creation easiness, etc. One of the most important aspects of mashups is the reuse of the code, so different combinations of data and services results in different value-added applications. Reusing ensures that software resources are used efficiently. Mashup applicability covers a wide range of purposes, from mapping to shopping. Perhaps the most popular example of mashup is Panoramio [14]. Panoramio is a geo-location oriented photo sharing website where images can be accessed as a layer in Google Maps or Google Earth.

Mashups are generally created using some kind of toolkit or environment called Mashup Editor. A mashup editor is an easy-to-use tool that allows rapid composition ("mashing up") of smaller components to generate a bigger value added application (mashups). Computing is distributed in set of reusable modules. Each module has a graphical representation were inputs and outputs are clearly stated. The outcome of a module can be used as the input parameter of another one (see Figure 1). Potential composing building blocks are pure web services, simple widgets, but also other mashups. Mashups usually incorporate a scripting or modeling language. In our case, the final outcome of the mashup will be rendered as a TV widget.

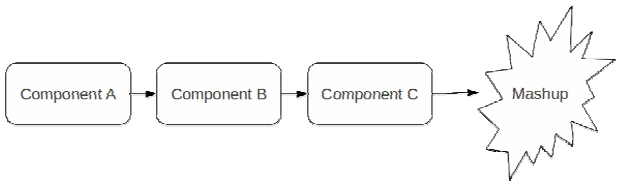


Figure 1: Mashup creation process

Mashup Editors usually have a graphical user interface to help the user build mashups easily and rapidly. The aim is to present an attractive and intuitive development environment. No technical education or background should be required to use these tools. There exist many different mashup editors, for example Yahoo Pipes [15], Impure [16] or ServFace Mashup Builder [17]. In our case we will use MyCocktail [18], a mashup tool developed in the FP7 project ROMULUS [19] and improved within the FP7 project OMELETTE [20].

MyCocktail is a web application providing a Graphical User Interface to agile mashup development. It allows information retrieval via REST services, processing with proprietary operators and display by a wide range of renderer applications. MyCocktail is based on Afrous, an Open Source AJAX-powered online platform (under MIT license).

MyCocktail presents a panel in the middle of the screen that serves as a dashboard to compose the mashups. Reusable modules are called "components". Another panel at the left side of the screen contains all available components represented by little icons. Different components can be dragged and drop by the user to the central panel in order to start combining them.

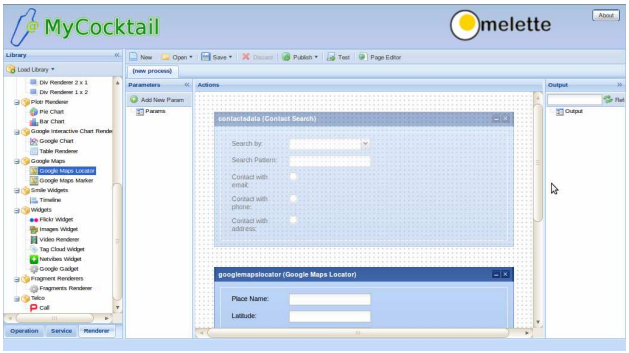


Figure 2: Screenshot of MyCocktail

There are three types of components defined in MyCocktail, one for each mashup architecture element:

- Services: Are invokable REST services like del.icio.us, Yahoo Web Search, Google AJAX Search, Flickr Public Photo Feed, Twitter, Amazon, etc.
- Operators: Retrieved data can be processed using different operators. Examples of operations that could be performed are: sorting, joining, filtering, concatenation, etc.
- Renderers: The outcome of the processed information can be presented using services available online. Depending on the whole purpose of the mashup a different renderer will be used. Possible purposes are HTML rendering, statistics presentation or map visualization, etc. Renderers are based on services like Google Maps, Google Charts, Simile Widgets, etc.

Each component dropped to the dashboard is represented by a box. Component input parameters can be passed through a form. The user can fill the form by dragging and dropping the outputs of other components to application fields, or by writing on application fields directly. The output panel is located on the lower part of the component box. Possible outcomes of a box are data or pieces of graphical user interface. This output panel allows the user to control at all times the outcome of every component.

Once the mashup presents the desired result, the outcome of the last component has to be dragged and dropped to the output

panel (located at the right side of the screen). This action defines the end of composition process. As a final step, the user has to export the result of the mashup in the desired format. Depending on the application, MyCocktail allows exporting to plain HTML, W3C widget, NetVibes widget, etc. For our scenario, which focuses on creating light-weight applications to be presented on TV screens, the required output format is TV widget. Several annotations and metadata may be included to help TVs display TV widgets.

5. WIDGET PERSONALIZATION USING RECOMMENDER SYSTEMS

In the case of study, TV widgets will be created to add value to content provision. The main purpose is to demonstrate applicability of TV widgets as the optimal solution to offer interactivity over IPTV networks. This is done by increasing the number of access to news content through widgets. Many standards are defined to improve navigation and interoperability on news content (NewsML [21], SportsML [22], etc). These standards define specific common metadata for every single new: location, date, time, genre, author, etc. Such metadata enable several semantic features. That is the reason why news contents have been selected among other types of video content.

Besides content retrieval using semantic queries, metadata opens the door to further personalization functionalities like recommendation. In a world where available content increases exponentially, a recommender system generates a strong value-added service. Recommender systems help users filter information, keeping only the content they really want to view. They also enable analyzing users to predict their preferences and prepare an attractive offer to them (personalized content, personalized advertising, etc).

In our scenario a separated server hosting a recommender system will be deployed. This server will be permanently available online. It will generate responses under different request received from client-side, that is, interactive TV widgets. Both client and server are aware of the exchange format. The response object is compound by a set of references to recommended content. Useful metadata like the title and the headings are attached to every reference. In order to calculate predictions the server gathers information from different social networks and content repositories.

The server establishes an interface with the client-side parties (TV widgets) to offer them different recommendation services:

- Content Recommendation based on similitude between users.
- Social Recommendation.
- Popular Content.

6. IMPLEMENTATION

Along this paper, several technologies have been described. All of them focus on reusing resources to create interactive TV applications in a cheap, fast and effective way. It is possible to build a solution taking advantage of all features offered by each of these elements. The solution will be an environment to create and consume personalized and interactive TV widgets reusing as much resources as possible. In order to keep the demonstrator as simple as possible, the resulting client side will only interact with the content provider and the recommendation server. Further interactions are not considered in this paper.

The different elements of this architecture are presented in Figure 3.

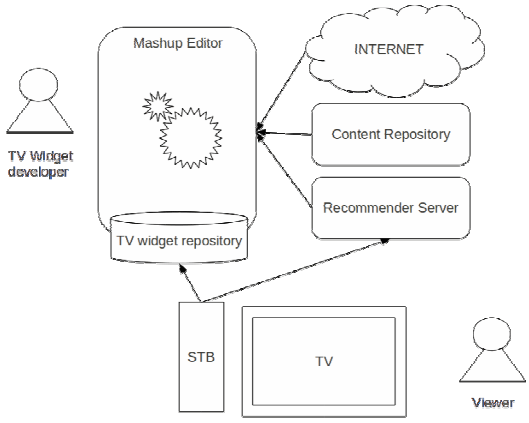


Figure 3: Architecture of the proposed solution

- Content repository: News content and related metadata will be stored in an Alfresco repository [23]. Alfresco offer standard interfaces to find and access information. This API has been exported to MyCocktail. It has also been imported to client-side since it's the one that will be used to offer video on-demand. Further functionalities and extensions can be easily developed as Alfresco webscripts (e.g. a web service to insert a new value in a content metadata).
- Recommender server: This server provides two different interfaces. The first interface will be used to accept new fixed preferences from client-side TV widgets. The other interface will be imported from MyCocktail, so the mashup editor can access a list of recommender items for a determined user. Many technologies support the recommender server. Apache Mahout provides collaborative filtering based recommender systems [24]. JGAP stands for Java Genetic Algorithms Package [25]. Both suites enable content recommendation.
- MyCocktail Mashup Editor: MyCocktail will reuse sources of information and functionalities by importing REST APIs. News metadata, recommendations and services will be combined to generate personalized interactive TV widgets according to users' preferences. Widgets will be stored in a dedicated widget repository.
- TV and Set-Top-Box: The user domain consists on a TV device and a Set-Top-Box which hosts the whole intelligence of the client side. The STB is in charge of invoking TV widgets stored in the widget repository. Once the code is received, the personalized interactive TV widget will be rendered on the screen (see Figure 4). From that

moment on, viewers can interact with the content provider and the server recommender. There are many platforms supporting TV widgets embedding on screens. The most popular ones are Google TV [26], Yahoo TV [27], and Microsoft Mediaroom [28].

A prototype of widget TV has been implemented using Yahoo TV as delivery platform. Contents are news stored in an Alfresco repository. While users are watching the news, they can open a widget to rate them. The output of the screen is shown in Figure 4. News-Rating TV widget is rendered on the left side of the screen while video is still running on the background. It allows viewers to feed the recommender system using a solution based on explicit ratings. TV widgets are a non-intrusive way to interact with the TV without giving up watching video content.



Figure 4: Example of interactive TV widget: A widget is presented to the user to rate news he has watched.

We have also created other examples of TV using this solution:

- Recommended News TV widget: it displays a list of news sorted by predicted preference
- News Search TV widget: it presents a list of news matching a determined query
- Additional information widget: while a determined new is playing on the TV, the widget presents value-added content (maps, timelines, tag clouds, etc)

7. CONCLUSION AND OUTLOOK

In this paper we have shown the benefits of TV Widgets that make use of Web information in order to provide Interactive TV services. We have described the mashup editor that has been used to create such TV Widgets. We have shown how the TV Widgets could be personalized using a recommendation system. Finally, we have described the implementation setup for a TV Widget that embeds news into the TV screen.

As future work, we envision the collection of TV Widget utilization data in order to personalize and recommend further services or news. Moreover, the TV Widgets could be extended to other scenarios and not only to the News case.

8. ACKNOWLEDGMENTS

This paper describes work undertaken in the context of Noticias TVi and OMELETTE.

Noticias TVi (http://innovation.logica.com.es/web/noticiastvi) is part of the research programme Plan Avanza I+D 2009 of the Ministry of Industry, Tourism and Trade of the Spanish Government; contract TSI-020110-2009-230.

OMELETTE, Open Mashup Enterprise service platform for LinkEd data in The TElco domain (www.ict-omelette.eu), is a STREP (Specific Targeted Research Project) funded by the European Commission Seventh Framework Programme FP7, contract no. 257635.

9. REFERENCES

[1] Scribemedia.org, 2007 *"Web TV vs IPTV: Are they really so different?"*. (Nov 2007). http://www.scribemedia.org/2007/11/21/webtv-v-iptv/

[2] Opcode.co.uk, 2001 *"Usability and interactive TV"*. http://www.opcode.co.uk/articles/usability.htm

[3] Gamasutra.com, 2001 *"What happened to my colors"*. (June 2001) http://www.gamasutra.com/features/20010622/dawson_01.htm

[4] Nintendo.co.uk , 2009 "Wii Technical details". http://www.nintendo.co.uk/NOE/en_GB/systems/technical_details_1072.html

[5] Rajesh Lal; *"Developing Web Widget with HTML, CSS, JSON and AJAX"* (ISBN 9781450502283)

[6] W3C widgets, 2007 *"Widget Packaging and XML configuration"* W3C Recommendation, 27 September 2007. http://www.w3.org/TR/widgets/

[7] Ian Hickson y David Hyatt, 2009 "HTML5 A vocabulary and associated APIs for HTML and XHTML" W3C, 6 de octubre de 2009 http://dev.w3.org/html5/spec/Overview.html#html-vs-xhtml

[8] Soylu, A., Wild, F., Mödritscher, F., Desmet, P., Verlinde, S., De Causmaecker, P. (2011) *"Mashups and widgets orchestration"* The International Conference on Management of Emergent Digital EcoSystems, MEDES 2011. San Francisco, California, USA, 21-24 November 2011. ACM.

[9] SOA WORLD MAGAZINE, *"Enterprise Mashups: The New Face of your SOA"* 2010-03-03.

[10] Richardson, Leonard; Sam Ruby (2007). *"Preface"*. RESTful web service. O'Reilly Media. ISBN 978-0-596-52926-0. Retrieved 18 January 2011.

[11] W3C. *"SOAP Version 1.2 Part 1: Messaging Framework (Second Edition)"*. April 27, 2007. Retrieved 2011-02-01.

[12] RSS 2.0 specification http://www.rssboard.org/rss-specification

[13] RFC 4287 "The ATOM Syndication Format" http://tools.ietf.org/html/rfc4287

[14] Panoramio, http://www.panoramio.com/

[15] Yahoo Pipes, http://pipes.yahoo.com/pipes/

[16] Impure, http://www.impure.com/

[17] ServFace Mashup Builder, http://www.servface.eu/

[18] MyCocktail, http://www.ict-romulus.eu/web/mycocktail

[19] ROMULUS project, http://www.ict-romulus.eu/web/romulus

[20] OMELETTE project, http://www.ict-omelette.eu/home

[21] NewsML standards, http://www.iptc.org/cms/site/single.html?channel=CH0087&document=CMS1206527546450

[22] SportsML standard, http://www.iptc.org/site/News_Exchange_Formats/SportsML-G2/

[23] Alfresco CMS, http://www.alfresco.com/es/

[24] Apache Mahout, http://mahout.apache.org/

[25] Java Genetic Algorithms Package, http://jgap.sourceforge.net/

[26] Google TV, http://www.google.com/tv/

[27] Yahoo Connected TV, http://connectedtv.yahoo.com/

[28] Microsoft Mediaroom, http://www.microsoft.com/mediaroom/

Automation of learning materials: from web based nomadic scenarios to mobile scenarios

Carles Fernàndez Barrera
Avinguda Tibidabo, 47
(+34) 93 3402316

Joseph Rivera López
Avinguda Tibidabo, 47
(+34) 93 3402316

ABSTRACT

This paper presents the experience of the UOC (Universitat Oberta de Catalunya) in its continuous processes of innovation of learning resources. In this case, the University has developed a software that allows the automatic transformation of web-based learning materials into mobile learning materials, more specifically for Android and iOS platforms.

Categories and Subject Descriptors

D.3.3 [**Programming Languages**]: Language Contructs and Features – *abstract data types, polymorphism, control structures.* This is just an example, please use the correct category and subject descriptors for your submission. The ACM Computing Classification Scheme: http://www.acm.org/class/1998/

General Terms

Your general terms must be any of the following 16 designated terms: Algorithms, Management, Measurement, Documentation, Performance, Design, Economics, Reliability, Experimentation, Security, Human Factors, Standardization, Languages, Theory, Legal Aspects, Verification.

Keywords

OCW, Opencourseware, e-learning, Android, Iphone

Introduction. About the UOC

The UOC is what we call a fully online learning university, meaning that the whole learning system and its services allows students to learn beyond the boundaries of time and space. The UOC is a cross-cutting Catalan university, less than fifteen years old, with a worldwide presence, aware of the diversity of its environment and committed to the capacity of education and culture to effect social change. At the moment the University has 56.000 students and offers 1,907 courses in various master's degree, postgraduate and extension programmes.

The next step: towards mobile scenarios

The UOC has a specific team working on innovation, consisting on developers, pedagogists, psychologists, graphic designers, etc., and this structure has facilitated the evolution of such resources, from the initial paper based scenarios to the most innovative mixed scenarios. Our students use the learning materials we provide them as the main tool for learning. The University, from time to time, develops several studies in order to understand what kind of tools and what kind of use our learners do of what we provide them, and we have observed that the last trend is that they use mobile devices (specially netbooks, Ipads and mobile phones) to access the wide range of tools we offer them. As an example, we have confirmed that a considerable number of students are using their mobile phones to access services like email, teacher spaces, subject forums, marks, etc. One of the demands of some of these students was to also have a proper mobile version of their learning materials, and is because of that that we decided to develop a way to provide them.

In our studies, we studied what mobile devices our students have, being Android devices and Iphones the most common ones. As most of these devices have screens smaller that 4 or 5", we planned to create a version of learning materials where learners not only have text but other multimedia resources like images, audio and video.

The strategy: offering Open Source Materials to the world

In line with the current trends in Education and more specifically Online Education, the UOC is offering these materials in the form of Creative Common licenses, allowing other users to work more freely with our materials. The learning materials will be uploaded to the Apple Store and the Android Market in order to make them available. The UOC is a member of the Open Content movement and own a OpenCourseWare portal (www.ucw.uoc.edu) , where many of our authors and teachers may upload their creations. The materials have a Creative Commons Attributive License 2.5, and so, can be offered without authoring conflicts. In this OpenCourseWare, students and other users can find resources that we currently use in our graduate and undergraduate courses, are free of charge and available exactly as our students receive them. In total, there are 240 materials in several languages: Catalan, Spanish and English, meaning all of them are potentially available

The software

In short, due to practical, pedagogical and economic reasons, we are interested to facilitate the creation of hundreds or thousands of learning materials for mobile platforms, avoiding the creation from scratch of these versions, which of course would be

inefficient and non possible because of the high cost. On the one hand, the Office of Learning Technologies of the UOC has used one of its previous developments, called *MyWay*, consisting on a ser of applications that allowed us to transform docbook documents (used by our materials) into different outputs like *epub, mobipocket or pdf*, and automatically producing several other type of outputs (for example, videomaterials, materials for ebooks, audiomaterials, etc.) from our web-based materials. Myway is distributed under Open License (GLP) and can be downloaded in the Google code website.

In conclusion, our web based materials (consisting on XML documents) can be automatically adapted to Android and Iphone devices (and their style guides), through a specific application consisting on a compiler that generates a set of files that we can upload to the mobile markets (iOS and Android), where they will be downloadable as applications.

From the point of view of our online students, who are very prone to move and study far from other locations that home, the functionality of having the learning materials "as they move" or "where they move" and available from several devices is a new world of flexibility and follows the requirements of a society where lifelong learning increases with new spaces for learning.

As for the concrete application, the compiler, it will be the piece that enhances the process of creation for new learning materials for new mobile platforms, avoiding the complex edition processes behind and saving all the money and time that our developers would need to create something from scratch.

Also, regarding guidestyles, the application we created follows the behaviors of Iphone and Android applications, facilitating the quick use of our materials by users who are used to work with these two common mobile phones. Following, we show a few images of one of our materials as a product of the use of the compiler.



Figure 1. Example of materials in Android version



Figure 2. Example of materials in Iphone version

The process of evaluation

The application is just being upload to the Android and Iphone markets, meaning at the moment we do not have results yet. As method for evaluation, we will analize the statistics of use, as well as the number of downloads and the number of visits that the UOC website receives from the two markets. By doing that we expect to be able to calculate how many users are interested to join UOC due to having used our materials. From the qualitative point of view, we will send questionnaires to users and will also interview material authors about the type of uses and quality of the learning materials.

Conclusions

Mobile phones and many other types of mobile devices are going to be (if they are not now) the dominants in the the short and mid term of the e-learning field. As an example, the last Horizon Report in 2011 stated that in both developed and developing countries, mobile devices are and will be the main trendy technology in the next years. Many Universities are currently conscious of the demands and requests of their students to have decent mobile versions of learning resources, meaning that there is necessary a reflection about the process of automation, and also other issues like authoring rights and the use of existing open licenses. For most of our students, always living between the transition between home or nomadic situations to mobile situations, *MyWay* has been the solution.

Quality Control, Caching and DNS - Industry Challenges for Global CDNs

Stef van der Ziel
Jet Stream B.V.
Helperpark 290, 9723 ZA Groningen,
The Netherlands
+31 50 5261820
stef@jet-stream.nl

ABSTRACT

When CDNs emerged in the late nineties[1], their purpose was to deliver content as best as possible on a best effort network: the Internet. Even though global CDNs offer SLA's, these SLA's only cover their own pipes, servers and support. Global CDNs dump their traffic on the Internet, via carriers or peering links, on internet exchanges, into ISP networks. Their SLA's don't cover any capacity guarantee, delivery guarantee or quality guarantee. So it made a lot of sense for these global CDNs to use matching technologies: caching and DNS. Both technologies are best effort technologies because they do not offer guarantee or control. Which wasn't a requirement on the Internet.

Categories and Subject Descriptors

C.2.4 [**Computer Communication Networks**]: Distributed Systems Network - operating systems

General Terms

Measurement, Design, Economics, Standardization

Keywords

CDN, Content Distribution Network, Caching, DNS

1. INTRODUCTION

Imagine this scenario: you are paying some of your hard earned money to watch a football match, or a blockbuster via an OTT (over the top) content provider. But the stream underperforms. It was advertised as being high quality, but the quality constantly changes from HD to SD quality. That's not what you paid for: you paid for a cinematic experience. So you want a refund. Who are you going to call? The OTT provider? If they have a customer service desk at all, their response will be: our CDN says everything is working fine, it must be the Internet or your local network, so don't call us. Your broadband access provider? Their response will be: hey the broadband service works, and you didn't buy the video rental service from us anyway, so don't call us. The subscriber will get frustrated. After another attempt which goes wrong, the subscriber walks away from the service, for good. Consumers will not accept a best effort, QoE service, especially not when they pay for a broadcast grade service, so they expect, demand and deserve broadcast grade quality. Everyone of them. Every view. Broadcast grade means that even the smallest outage is unacceptable. That one minute downtime may not occur during an advertisement window. That one minute downtime may not occur during that major sports event. Broadcast grade means that the quality of the image and the audio is constant and is guaranteed. It is QoS, not QoE. QoE cannot replace QoS. Quality of Experience is a euphemism because there is no constant quality of experience. QoE is a workaround for the real problem. That problem is that the Internet is simply not capable of delivering a broadcast grade service. And neither are Internet CDNs. The Internet is a collection of patched networks, without any capacity, performance or availability guarantee. There are no SLAs between CDNs, carriers and access providers. That by itself is a blocking issue for anyone trying to offer a premium service via Internet based CDNs. So is the answer in putting CDN capacity within a telco network? Yes and no. Opposed to the Internet, a telco network is a managed network. So a telco can actually offer end to end QoS and SLAs to OTT providers and subscribers. But putting an Internet CDNs technology or a vendor CDN technology that is heavily based upon Internet CDNs on top of a telco network does not solve the problem. Since Internet CDNs are built upon best effort technologies such as caching and DNS. It is like putting a horse carriage on the German Autobahn.

2. CACHING

Caching does not guarantee that every user always gets the requested content in time. It is a passive and unmanaged technology, assuming that caches can pull in content from origins and pass it through in realtime to end users. It is an assumption but there's no guarantee. No management, no 100% control. It usually works great. Best effort. Internet grade. Not broadcast grade. Caching assumes that an edge cache can pull in an object from an origin. But what if the origin or the link to the origin is unavailable? Caching assumes that an edge cache can pull in an object in real time. But what if the origin or the link to the origin is slow? There are too many assumptions but there is no 100% guarantee that every individual viewer can get access to the requested object, there is no 100% guarantee that the file wasn't corrupted in the caching process, there is no 100% guarantee that the end user gets the object in realtime. Caching is a passive distribution technology without any guarantees. Best effort.

3. DNS

DNS does not guarantee that every user request is always redirected to the right delivery node. DNS is a passive, unmanaged technology, assuming that third party DNS servers work along and assuming that end users in bulk are nearby a specific DNS server thus nearby a specific delivery node. These are assumptions but there is no guarantee. No management, no 100% control. It usually works. Best effort. Internet grade. Not broadcast grade. DNS assumes that an end user is in the same region as their DNS server. But what if the user is using another DNS server? The end user will connect to a remote server, with

dramatic performance reduction and higher costs for the CDN. DNS assumes that other DNS servers respect their TTLs. But what if they don't? End users will connect to a dead or overloaded cache, dramatically degrading the uptime of the CDN. There are many more downsides to DNS. DNS is a passive request routing technology without any guarantees. Best effort.

4. MARKET WEAKNESS

Akamai was the first mover for CDNs on the web. They pioneered. And most other Global CDNs have basically copied the same fundamental Akamai approach (with mild variations). Best effort stuff, which is great for the web, but not for premium content. We would have expected a smarter approach from telco technology vendors. However, they too simply copied the best effort approach of global CDNs: their solutions still heavily rely on caching and DNS. They lock telcos into their proprietary caches, storage vaults, origins, control planes and pop heads. This isn't innovative technology. It is stuff from the nineties. That is a prehistoric era in Internet years. This is a fundamental problem: they just don't understand what premium delivery really means.

5. MARKET DEMAND

Subscribers and content publishers demand more. They want their all-screens retail content and also the OTT content to be delivered with the same quality as they get from digital cable. Content publishers demand SLA's with delivery guarantee, delivery capacity guarantee, delivery performance guarantee.

1. Premium content requires premium delivery.
2. Premium delivery requires a premium network.
3. Not the internet, but on-net delivery by operators who own and control the last mile.
4. Operator CDNs require a premium CDN.
5. Not a CDN that is built upon best effort stone age technologies such as caching and DNS.
6. Global CDNs technologies make no sense in a telco network.
7. Vendors who offer CDNs based upon caching and DNS are totally missing the point.

The key is to, of course, support caching and DNS, but with smarter caching features such as thundering herd protection, logical assets for HTTP adaptive streaming, session logging for HTTP adaptive streaming, and multi-vendor cache support where you can mix caches from a variety of vendors in a single CDN. However, the future of CDNs is in premium delivery on premium networks with premium, active and managed technologies for request routing and content distribution. The future is in QoS, the future is in SLAs. Basically you can't monetize a CDN if you can't offer Qos and a proper SLA.

6. INDUSTRY CHALLENGE

The real challenge for this industry is to make sure that we get end to end SLAs, from CDN down to the subscriber. It is not a matter of getting subscribers to pay more for a premium service, it is a matter of getting them to pay at all for such a service. That is impossible when OTT providers use Internet CDNs. That is impossible when access providers use transparent internet caching. That is impossible when access providers use DNS and caching based CDNs. Premium content deserves premium delivery.

7. CDN COMPLIANCY

This section will help developers of Smart TVs, Smart Phones, Tablets, SetTopBoxes and Software video clients to better understand the dynamics of a Content Delivery Network so they can improve the way their client integrates with Content Delivery Networks.

7.1 Implement the full HTTP 1.1 specification: HTTP redirect

HTTP adaptive bit rate streaming is great technology. But for many vendors, it is quite new technology. If you implement HTTP adaptive bit rate streaming, whether it is Apple HLS, Adobe HDS, Microsoft Smooth Streaming or MPEG DASH, make sure you fully support HTTP redirect. Come on, it's not new, it is from 1999! Why do clients need to implement HTTP redirect? Many portals and CDNs use HTTP redirect. For instance for active geo load balancing. Or in CDN federation environments. f you claim to support HTTP streaming, you claim to support the formal HTTP specs. Which includes these redirect functions. Users, CDNs and portals always must be able to assume that your client complies to the full spec. If you don't your users will end up with broken video streaming after paying for that premium video. We were quite surprised when we learned that Adobe implemented their own HTTP Dynamic Streaming technology in Adobe Flash but forgot to implement HTTP redirect in the very same Adobe Flash player. Whoops.

7.2 Respect HTTP redirect status

To be more specific, make sure you respect the HTTP redirect status as being provided by portals and CDNs. We have seen clients assume that any 30x response always is a permanent move. However for instance, 307 means temporarily redirect: so go back to the original URI. Stick to the rules! Never assume that any 30x response is a specific redirect instruction. If you wrongly implement redirect, end users will be affected.

7.3 Retry after HTTP 503 status response

In rare occasions, a portal, a CDN request router, a Federated CDN or a CDN delivery node may be unable to process a clients request and responds with 503 Service Unavailable. These CDNs can add a 'retry-after' statement, allowing the client to wait and retry the request, allowing the client to retry. However we have seen some clients fully drop the request after a 503 response, ignoring the retry-after statement. Never assume that a 50x response immediately means complete unavailability, please check the header for retry information and respect the retry-after time.

7.4 Respect DNS TTL

Many generic CDNs use DNS tricks. Their master DNS tricks third party DNS servers into believing that for a certain domain, their can be specific IP addresses. That is how they let users connect to local edge servers. If for any reason the CDN needs to send users to an alternative server, they update their records so your DNS provider gets the latest update for the right IP address. That is why CDNs typically use a low TTL. Although DNS is a poor man's solution that never really can guarantee any controlled

request routing, it is still fairly common with website accelerating Internet CDNs. So respect the DNS TTL statements: don't override this TTL. If you cache DNS records for a longer period than stated in the TTL, your viewers risk being sent to an outdated, out of order or overloaded delivery node. Meaning no video, or a crappy stream. CDNs already have the risk that DNS servers override their TTLs. Don't let this happen for media clients (or their OSes). We have seen quite some SetTopBoxes that override DNS TTL or have poor DNS client implementations. Caching DNS records sounds like a great idea, but there is a reason why CDNs use a low TTL! To add to the problem: many applications (not resolver libraries but video playback software) request an IP address once for a hostname. And as long as this application is running, the app will use the same IP address. We see this behavior quite a lot. So even when DNS servers and the local resolver libary on the client all respect the DNS TTL as they should, the app will still use an outdated IP address. Resulting in broken or underperforming streams. So what the application should do is re-request the right IP address for every new session.

7.5 Random use of multiple IP addresses in DNS record

CDNs can return a list of IP addresses for a single DNS record. DNS servers typically put these IP addresses in a random order, so end users are load balanced quite nicely over the delivery nodes. However you cannot assume this. And you can't guarantee or control or measure this from a CDNs perspective: third party DNS servers can have their issues. We have seen quite some examples where the ISPs DNS servers always respond with the same order of IP addresses. If media clients don't randomly pick an IP address, they will all connect to the first request router or to the first delivery node in the list, resulting in some servers being overloaded, while others are idle. This happened in a large OTT CDN where the SetTopBox from a respected vendor had a poor DNS implementation. It resulted in sub optimal performance for all clients, even ones that were not affected. Even though this is a limitation of how DNS works (DNS was never intended for CDN usage, so there are no clear instructions to clients about how to behave in these conditions), it is advised to vendors to make sure they do not assume that the list of IP addresses is random, but select IP addresses randomly themselves.

7.6 Implement the full protocol spec

It is a problem of embedded technology vendors in general: they do a poor implementation of a protocol. We see it all the time in delivery appliances, but we also see it in media clients such as Smart TVs. One example is where Samsung released a new Smart TV series not so long ago, claiming to fully support Windows Media Streaming. However they only tested with a Windows 2003 server in their lab. The TV could only connect via the MMS protocol on port 1755. Which was declared by Microsoft as legacy in 2003 and isn't even supported in Windows Media Services 2008 anymore. The result was that quite some premium OTT providers had invested in Samsung TV apps but could not launch their service. And because it was all hardcoded there was no way to force the client to connect to the right ports with the right protocols. Another example is where Nokia phones with Real clients falsely interpret SDP session bit rate information. This leads to constant buffer underruns and buffer overruns. CDNs and streaming server vendors had to jump through hoops and adapt their SDP sessions constantly for these specific clients because Nokia and Real ignored the problem. They should have used a proper modern server environment in their lab. Or even better, they should have worked with a real life CDN to test the product before putting it on the market. They should have checked about legacy and unsupported technologies. They should have implemented the Windows Media clients ability to process MMS requests but connect to RTSP on port 554. Don't assume compliancy on paper or from a lab environment.

7.7 Implement protocol rollover

Even though many protocols don't formally spec protocol rollover, it is a must have for media clients. For instance, the 3GPP RTSP streaming spec formally prescribes RTSP via UDP on port 554. However, in many environments, UDP is blocked due to NAT routing. That is why most CDNs and streaming services also support rollover to HTTP via TCP on port 80. One specific example is where a mobile phone vendor forgot to implement such a rollover function. It isn't formally part of the standard, but the entire industry does it anyway to make the standard work in real life. End users of these smart phones constantly complain that streams don't work on their devices. Vendors must do more real life field tests. In lab enviroments, cell phones always work great. If they had worked with a CDN, their product would have been more rugged and tested. And used. Now they lose to competition. Jet-Stream also supports various interprotocol redirect functions, which are great to redirect clients, for instance HTTP->RTSP redirect. The more protocols your client supports, the better the chance that your end users will get a decent stream.

7.8 Do not cache playlists

A very common mistake is that media players tend to cache playlists. We have seen it in virtually all WinAmp-like clients, including iTunes. These clients assume that URLs in the playlist are static. Wrong! A CDN is not a static environment. You cannot assume that a URL has an unlimited validity. Most CDNs use anti deep linking technologies, involving tokens that expire after a few minutes. If these clients cache a playlist, then the user will be able to stream or download the first object, but in the meantime the token expires for all the other objects. The result: the end user can't download or stream what they paid for. Whoops.

7.9 Stick to your own specs

We have seen quite some vendors who designed their own media player technology. And then they suddenly changed the client or the server without informing the industry. Which can be quite frustrating to CDN operators and content publishers: end users install a new client or upgrade their smart phone and suddenly the stream breaks. We have seen all major vendors pulling these tricks unfortunately. Whether they changed something in how HTTP adaptive streaming behaves (Apple) or changed their protocol spec over night (Adobe). Backward compliancy is extremely important!

7.10 Support referrer files

Referrer files (RAM, QTL, ASX, XSPF, SMIL for instance) are a great tool for CDNs to redirect clients to specific delivery nodes. Almost every media streaming technology has it's own implementation and Jet-Stream supports them all. The CDN doesn't passively use DNS or responds with a simple HTTP redirect: the CDN generates an instruction file, which can be

parsed by the client so it will always connect to the right delivery node. Referrer files are the most underestimated technology in CDN-client integration! Referrer files are the most ideal way for CDNs to instruct clients how to respond. Instead of having to rely on DNS (best effort technology) or on HTTP redirect, CDNs can put quite some intelligence into referrer files, such as protocol, server and port rollover, bringing much higher uptime (broadcast grade uptime) to clients than possible with DNS or redirect based CDNs. Unfortunately many embedded clients claim to fully support a specific streaming technology. Even when they implemented all kinds of protocols and ports, many of them simply forgot to implement the referrer file. Which really limits the CDN and client interaction possibilities. The best referrer file spec is Microsoft Windows Media ASX (WMX). Unfortunately Microsoft decided not to support their own technology with Smooth Streaming. Doh! It is on-the-shelves technology that Microsoft could have used to smooth (pun intended) the migration from all those Windows Media based services to Smooth Streaming. It could have brought them back World Domination in streaming, after losing their market share to Flash. Also if you decide to implement support for referrer files, make sure you fully implement according to the specs. For instance, if the referrer file contains a list of servers, make sure you rollover to alternative servers, or protocols as described in the referrer file.

So make sure your client supports referrer files to make sure your viewers get the best streaming experience out there.

8. ACKNOWLEDGMENTS

Our thanks to Richard Kastelein for helping prepare this paper for submission.

9. REFERENCES

[1] Tim Siglin, What is a Content Delivery Network (CDN)? - A definition and history of the content delivery network, as well as a look at the current CDN market landscape - http://www.streamingmedia.com/Articles/Editorial/What-Is-.../What-is-a-Content-Delivery-Network-%28CDN%29-74458.aspx.

Grand Challenge

EuroITV Competition Grand Challenge 2010-2012

Artur Lugmayr
EMMi Lab.
Tampere Univ. of Technology
POB. 553, Tampere, Finland
+358 40 821 0558
lartur@acm.org

MilenaSzafir
Manifesto 21.TV, Brazil

milena@manifesto21.tv

Robert Strzebkowski
Dept. of Comp. Science and Media
Beuth Hochschule für Technik Berlin
Luxemburger Str.10, 13353 Berlin, Germany
+49 (0) 30-4504-5212
robertst@beuth-hochschule.de

ABSTRACT

Within the scope of this paper, we discuss the contribution of the EuroITV Competition Grand Challenge that took place within the EuroITV conference series. The aim of the competition is the promotion of novel service around the wider field of digital interactive television & video. Within the scope of this work, the EuroITV Competition Grand Challenge is introduced and the entries from the year 2010-2012 are presented.

Categories and Subject Descriptors

D.2.4 **[Computer-Communication Networks]**: Distributed systems – *client/server*, *distributed applications*. H.3.5 **[On-line Information Services]**: *commercial services*, *web-based services*. J.4 **[Social and Behavioral Sciences]**: *economics*, *sociology*.

General Terms

Management, Performance, Economics, Experimentation, Security, Human Factors.

Keywords

Broadcasting Multimedia, EuroITV Grand Challenge, Interactive Television, Multimedia Applications, Broadcasting Content

1. INTRODUCTION

The idea of the competition was the promotion of practical applications in the field of audio-visual services and applications. Thus pressing play and watching as moving images unfold on a screen is ultimately an experience for the audience. We enjoy watching a movie to be entertained – or simply enjoy staying informed on current events by watching news broadcasts. Interacting with the content also adds up to the total experience, as for example playing online games parallel to watching a television documentary.

Thus not solely the entertainment part is in the foreground, also the learning effect via e.g. serious games. Shared experiences are another form of entertainment, which are enabled by adding social media type of services to the television platform. One prominent example is the exchange of video clips based on a friend's recommendation via mobile phone that let a shared emerge and allows to adding the feeling of relatedness between people. Another example for potential service scenarios viable for submitting to the competition are mashups of social media, which present ones' competence and creativity.

The competition therefore aims to premier creators, developers, and designers of interactive video content, applications, and services that enhance the television & video viewing experience for the audience worldwide. The entries shall help to answer the following questions:

- How does the entry add to the positive viewing experience?

- How does the entry improve digital storytelling?

- How do new technologies improve the enjoyment?

- What are the opportunities for innovative, creative storytelling?

- How can content be offered by a multitude of platforms?

An international jury composed of leading interactive media experts honor and award the best entries with a total price sum of 3.000 Euro, as well as additional prices for excellence in enhancing the viewing experience. The price sum is divided between several entries. Entries that are accepted had to be piloted within the two previous years.

To allow the jury to evaluate the various entries, each competition entry had to include:

- Demonstration, production, pilot, or demo as 3-5 minute video submission;
- Description of the entry in form of sketches, storyboards, images, power points, or word documents;
- Link to project websites, and production teams;
- Filling out of a brief questioner about the entry.

The international jury has to evaluate the entries according the following criteria:

- Innovativeness
- Commercial impact
- Entertainment value/usefulness/artistic merit/creativity
- Usability and user experience

2. THE COMPETITION 2010-2012

In total the competition attracted 35 entries during the years 2010-2012 (14 in 2010; 6 in 2011; and 15 in 2012). Table 1 presents the organizing team of the competition during these years. The winners of the competition are shown in Table 2. The organizers of the competition compromised an international team from Finland, Germany, Brazil, USA, and UK. The total entries of the competition are presented in the appendix of this publication.

EuroITV Competition Jury Members 2010-2012

Award Chair:

- Artur Lugmayr, Tampere Univ. of Technology, Finland

Competition Chairs:

- Susanne Sperring, Åbo Akademi University, Finland (2010)
- Milena Szafir, Manifesto21.TV, Brazil (2010-2012)
- Robert Strzebkowski, Beuth Hochschule für Technik, Germany (2010-2012)

Previous and Current Jury Members:

- Simon Staffans, Format developer at MediaCity, Åbo Akademi University, Finland
- William Cooper, Chief Executive of the newsletter informitv.com, UK
- Nicky Smyth, Senior Research Manager of BBC Future Media & Technology, UK
- Sebastian Moeritz, President of the MPEG Industry Forum, US/UK (2010, 2011)
- Carlos Zibel Costa, Professor of Inter Unities Post Graduation Aesthetics and Art History Program of University of São Paulo, BRA
- Esther Império Hamburger, television and cinema course's coordinator of University of São Paulo, BRA
- Jürgen Sewczyk, SmartTV Working Group at German TV-Platform, Germany
- Almir Antonio Rosa, CTR-USP, Brazil
- Rainer Kirchknopf, ZDF, Germany
- Alexander Schulz-Heyn, German IPTV Association

Table 1. EuroITV Competition Jury Members 2010-2012

EuroITV Competition Grand Challenge Winners

2010:

Waisda?

- Netherlands Institute for Sound and Vision, KRO Broadcasting - *for Excellence Achieved in Innovative Solutions of Interactive Television Media*

2011:

1st Place: Leanback TV Navigation Using Hand Gestures

- Benny Bing, Georgia Institute of Technology, USA

2nd Place: Video-Based Recombination For New Media Stories

- Cavazza & Lenoardi, Teeside Univesity, UK and University of Brescia, Italy

3rd Place: Astro First

- ASTRO Holdings Snd Bhd, Malaysia

Table 2. EuroITV Competition Grand Challenge Winners

3. EuroITV COMPETITION GRAND CHALLENGE 2012

The competition in 2012 attracted 15 challenging projects, which are listed in Table 3 in the appendix of this publication. At the stage of writing this publication, the winners have not be evaluated yet, but can be found on the EuroITV 2012 [3] website after the conference date.

4. CONCLUSIONS

For more information about the competition, please visit the EuroITV websites of 2010 [1], 2011 [2], and 2012 [3]. We are also aiming at creating a website that contains several competition entries of the previous years.

REFERENCES

[1] EuroITV 2010. www.euroitv2010.org

[2] EuroITV 2011. www.euroitv2011.org

[3] EuroITV 2012. www.euroitv2012.org

APPENDIX: COMPETITION ENTRIES

Competition Entries 2010

Crossed Lines	Sarah Atkinson
clipflakes.tv	clipflakes.tv GmbH, Germany
movingImage24	movingIMAGE24, Germany
Smeet	Communication GmbH, Germany
5 Interactive Channels in HD for “2010 FIFA World Cup”	Astro, Malaysia
Zap Club	Accedo Broadband, International
Active English	Tata Sky Ltd., Bangalore, India
Simple HbbTV Editor	Projekt SHE
LAN (Living Avatars Network) TV	M Farkhan Salleh
Waisda?	Netherlands Inst. for Sound and Vision / KRO Broadcasting, Netherlands
Cross Media Social Platform - Twinners Format	Sparkling Media-Interactive Chat Systems / Richard Kastelein
Smart Video Buddy	German Research Center for Artificial Intelligence (DFKI), Germany
KIYA TV	KIYA TV Production und Web Services
Remote SetTopBox	Univ. Ghent University - IBBT, Belgium

Competition Entries 2011

Hogun Park	KIST, Seoul, Korea
Leanback TV Navigation using Hand Guestures	Georgia Institute of Technology, USA
Astro First	Astro, Malaysia
meoKids Interactive Application	Meo, Portugal
Peso Pesado	Meo, Portugal
Video Based Recombination for News Media Stories	Univ. of Bresia, Italy and Teesside University, UK

Competition Entries 2012

The Berliner Philharmoniker's Digital Concert Hall	Berlin Phil Media / Meta Morph, Germany
illico TV New Generation	Videotron & Yu Centrik, Canada
The Interactive TV Wall	University Stefan cel Mare of Suceava, Slovenia
BMW TV Germany	SAINT ELMO'S Entertainment, Germany
mpx	thePlatform, USA
Social TV für Kinowelt TV	nacamar GmbH, Germany
Phune Tweets	Present Technologies, Portugal
QTom TV	Qtom GmbH, Germany
Linked TV	Huawei Software Company, International
mixd.tv	MWG Media Wedding GmbH, Germany
non-linear Video	BitTubes GmbH, Germany
lenses + landscapes	Power of Motion Pictures, Japan
Pikku Kakkonen	YLE, Finland
3 Regeln	Die Hobrechts & Christoph Drobig, Germany
CatchPo -Catch real-time TV and share with your friends	Industrial Technology Research Institute (ITRI), Taiwan

EuroITV 2012 – Organizing Committee

General Chairs: Stefan Arbanowski, Fraunhofer *(Fraunhofer FOKUS, Germany)*
Stephan Steglich, Fraunhofer *(Fraunhofer FOKUS, Germany)*

Program Chairs: Hendrik Knoche *(EPFL, Switzerland)*
Jan Hess *(University of Siegen, Germany)*

Tutorial & Workshop Chairs: Maria da Graça Pimentel *(Universidade de São Paulo, Brazil)*
Regina Bernhaupt *(Ruwido, Austria)*

Doctoral Consortium Chairs: Marianna Obrist *(University of Newcastle, UK)*
George Lekakos *(Athens University of Economics and Business, Greece)*

Demonstration Chairs: Jean-Claude Dufourd *(France Telecom, France)*
Lyndon Nixon *(STI, Austria)*

iTV in Industry: Patrick Huber *(Sky Germany)*

Grand Challenge Chair: Artur Lugmayr *(Tampere University of Technology, Finland)*
Milena Szafir *(Manifesto 21.TV, Brazil)*
Robert Strzebkowski *(Beuth Hochschule für Technik Berlin, Germany)*

Track Chairs: David A. Shamma *(Yahoo! Research, USA)*
Teresa Chambel *(Universidade de Lisboa, Portugal)*
Florian 'Floyd' Mueller *(RMIT University, Australia)*
Gunnar Stevens *(University of Siegen, Germany)*
Célia Quico *(Universidade Lusófona de Humanidades e Tecnologias, Portugal)*
Filomena Papa *(Fondazione Ugo Bordoni, Roma, Italy)*

Sponsor Chairs: Oliver Friedrich *(Deutsche Telekom/T-Systems, Germany)*
Shelley Buchinger *(University of Vienna, Austria)*
Ajit Jaokar *(Futuretext, UK)*

Publicity Chairs: David Geerts *(CUO, IBBT / K.U. Leuven, Belgium)*
Robert Seeliger *(Fraunhofer FOKUS, Berlin, Germany)*

Local Chair: Robert Kleinfeld *(Fraunhofer FOKUS, Berlin, Germany)*

Reviewers:

Jorge Abreu
Pedro Almeida
Anne Aula
Frank Bentley
Petter Bae Brandtzæg
Regina Bernhaupt
Michael Bove
Dan Brickley
Shelley Buchinger
Steffen Budweg
Pablo Cesar
Teresa Chambel
Konstantinos Chorianopoulos
Matt Cooper
Cédric Courtois
Maria Da Graça Pimentel
Jorge de Abreu
Lieven De Marez
Sergio Denicoli
Carlos Duarte
Luiz Fernando Gomes Soares
Oliver Friedrich
David Geerts
Alberto Gil-Solla
Richard Griffiths
Gustavo Gonzalez-Sanchez
Rudinei Goularte
Mark Guelbahar
Nuno Guimarães
Gunnar Harboe
Henry Holtzman
Alejandro Jaimes
Geert-Jan Houben
Iris Jennes
Jens Jensen
Hendrik Knoche
Marie José Monpetit
Omar Niamut
Amela Karahasanovic
Ian Kegel
Rodrigo Laiola Guimarães
George Lekakos
Bram Lievens
Stefano Livi
Martin Lopez-Nores
João Magalhães
Reed Martin
Oscar Martinez-Bonastre
Judith Masthoff
Maja Matijasevic
Thomas Mirlacher
Lyndon Nixon
Marianna Obrist
Corinna Ogonowski
Margherita Pagani
Isabella Palombini
Jose Pazos-Arias
Chengyuan Peng
Jo Pierson
Fernando Ramos
Mark Rice
Bartolomeo Sapio
David A. Shamma
Mijke Slot
Mark Springett
Roberto Suarez
Manfred Tscheligi
Alexandros Tsiaousis
Pauliina Tuomi
Tomaz Turk
Chris Van Aart
Wendy Van den Broeck
Paula Viana
Nairon S. Viana
Petri Vuorimaa
Zhiwen Yu
Marco de Sa
Francisco Javier Burón Fernández
Rong Hu
Johanna Meurer

Supporters and Partners:























EuroITV 2012

10th European Interactive TV Conference

» Bridging People, Places & Platforms «

Adjunct Proceedings, ISBN: 978-3-00-038715-9