# ALADIN: Adaptive Voice Interface for People with Disabilities

**Jonathan Huyghe, Jan Derboven, Dirk De Grooff**
CUO|Social Spaces, iMinds-KU Leuven
Parkstraat 45, bus 3605, 3000 Leuven (B)
firstname.lastname@soc.kuleuven.be

**ABSTRACT**
This position paper gives an overview of our ongoing work within the ALADIN project, which aims to develop an assistive vocal interface for people with physical impairments. Unlike most current Automatic Speech Recognition solutions, the system is trained by the user, which provides extra challenges to the design of the interface. We describe three iterations of our user tests, showing how constraints and multimodal design influence the user expectations and interactions.

**Author Keywords**
Assistive technology; Speech interaction; Learning systems

**ACM Classification Keywords**
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

## INTRODUCTION

While a lot of systems in the home can now be remotely controlled by people with motor impairments, the means of controlling them are not always ideally suited to their abilities. Current solutions rely mostly on button-based remote controls or a graphical user interface operated using switches or other means such as a sip-and-puff device (see figure 1), which are controlled with varying ease-of-use.

A voice-operated interface could help regain people with motor impairments the ability to control their home, domestic appliances, or entertainment devices, voice control, contributing significantly to their independence of living and quality of life [7].

In this paper, we describe the speech recognition system for people with disabilities developed in the ALADIN project. We describe the overall project goals and three iterations of user tests, focusing on how our methodology of testing influenced the way in which users interact with the vocal interface.



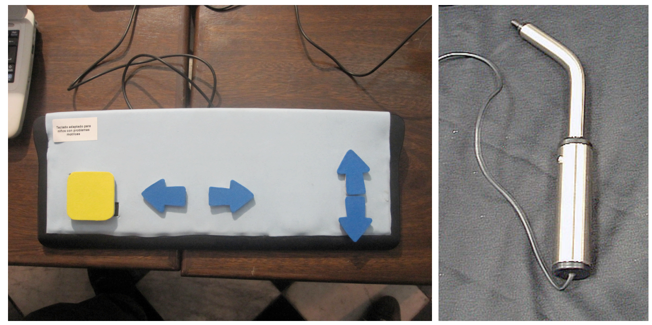

**Figure 1. Adapted keyboard for children [3] (left) and a sip and puff device [4] (right), two examples of existing interfaces for people with disabilities**

## AIM OF THE PROJECT

The ALADIN project was set up to create an adaptive, learning speech recognition system for people with disabilities, offering control over a wide range of applications. So far, vocal interfaces have not yet seen a wide adoption in assistive technologies, despite the obvious advantages as an interface for people whose impairment restricts (upper) limb use and thus their ability to use more traditional remote controls. There are several reasons why speech recognition is difficult to implement for this target group:

- A lot of users who could benefit from voice control due to motor impairments also suffer from a speech pathology, making state-of-the-art speech recognisers unusable for them.
- Current vocal remotes require the user to use pre-defined commands, forcing them to adapt to the system and learn the proper commands.
- Progressive diseases often lead to changing speech patterns, which requires a constant adaption of the system.

There are already a number of solutions that address some of these problems, but are lacking in other aspects: the Pilot Pro [5], for example, offers a fixed number of pre-programmed functions and a very hard to use training method. Castle OS [1], a more recent solution that is not aimed specifically at people with disabilities, features a more intelligent and expandable set of controls, but uses natural language recognition, unsuitable for people with speech impairments.

The aim of the ALADIN project then, is to provide users with a system that can be adapted to their specific living situation and can learn their commands instead of the other way around, deducing grammar and vocabulary from the user's speech. This deduction happens based on examples of commands, which are given using traditional remote controls and via speech at the same time. The system then links sounds from the user's speech with concepts in their commands (such as open/kitchen/door).

## DESIGN PROCESS

As this project is a collaboration between HCI researchers and speech recognition researchers, the work on both sides necessarily runs parallel to each other. During the course of the project, we started off with user research (both contextual inquiries in the homes of users, and context mapping sessions [8]), moving to several user tests informing the system's design.

### User Research

We visited the residences of 10 people (aged 9-48, average age 30), all of whom had physical limitations: half of the participants lived independent or assisted, and half of them were physically completely dependent, and lived at their parents' homes or in a residential care center. Points of interest in this contextual inquiry were physical and cognitive possibilities and limitations, the living environment (adjustments made, tips and tricks of the user to adapt to his environment, most important/preferred places in the house), problems encountered, organization of daily tasks, and assistance of devices or caregivers.

Apart from the contextual inquiry, two context mapping sessions were organized. These sessions focussed on the current quality of life of people with physical limitations, and how speech technology could help them achieve a better quality of life. The first session was held with two caregivers and two occupational therapists, while the second session was held with six people with physical limitations (aged 23-53, average age 43), and three caregivers that accompanied them. The context mapping sessions further enriched the results of the contextual inquiry by offering a more detailed insight in what is important in the lives of people with physical limitations, and how speech technology can be useful to them (see figure 2).

Both the contextual inquiry and the context mapping sessions were used as input to determine the application that was to be developed in the ALADIN project. The conclusion was that the main areas in which voice control would be useful were home automation (opening doors, blinds, switching lights, etc.), entertainment and communication. For the young participants in our research, entertainment was more important than home automation, as they liked receiving a lot of attention from their parents and their caregivers. However, for the adult participants, living as independently as possible clearly was an important factor determining their quality of life. For this reason, home automation was selected as the focal use case in the ALADIN design process. This process, starting from scenario sketches and Wizard-of-Oz prototyping of home automation, will lead to a field test involving the entertainment component, using voice-controlled televisions installed in the homes of several users.

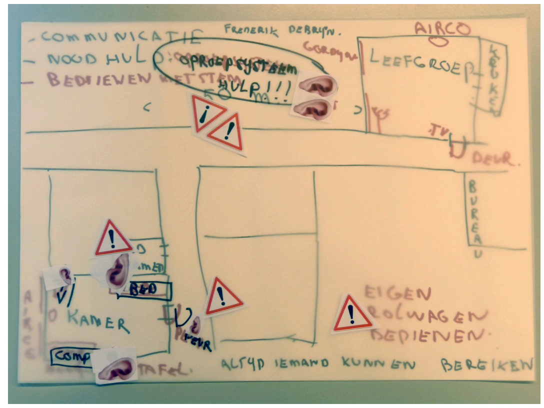


**Figure 2. Context mapping outcome: a participant's living environment. The exclamation marks indicate dangerous situations, the ears indicate potentially interesting uses for speech technology.**

### Sketched Scenarios

To gauge how people would want to use a voice-controlled home automation system, we used the 'sketched scenario' method, in which we presented a group of users, some of whom were part of the first stage of user research, with visualisations of interactions, and asked them to utter the voice commands they would use to control this interaction (see figure 3).

The focus of this study was not to simulate system interaction in a very realistic way, but rather to explore the variation in how the targeted user group addresses a voice interaction system. Significant diversity was found in interaction styles: voice commands ranged from a purely 'technical', command-style interaction to a more anthropomorphized, natural communication with the system. In addition, some respondents addressed individual devices, without addressing the voice-control system separately, while other respondents addressed the voice-controlled system as a whole, telling it to act on the environment and control other devices. For this last group, addressing separate objects such as doors felt very unnatural. On the other hand, the device-oriented way of thinking implies a different technology approach, in which the system can identify the users' intentions based on their location and context (for instance, 'light on' turns on the lights in the room where the user is, without specifying which particular lamp).

Figure 3. Sketched scenario used during our first tests.

### Wizard of Oz

The second medium-fidelity approach to user tests came in the form of Wizard of Oz testing, which has its roots in the testing of speech recognition applications [6]. Because we mainly focused on home automation at this point, an efficient way of simulating typical home automation tasks was needed that could also be used on location with test users, whose mobility was often limited. This simulation was made in the form of a virtual 3D environment, modeled after an adapted home for people with disabilities (figure 4). Using the Unity3D application, we could open doors, turn on lights, adjust the bed, etc. from a separate interface, allowing a researcher behind to scenes to manipulate the 3D home based on voice commands from the user, who was taken through a scenario with a moderator. This proved to be a much more immersive experience for users, and created a much more realistic representation of the envisioned interaction.

Figure 4. Screenshot from the 3D virtual home.

In the Wizard-of-Oz tests, participants were asked to address the system using a system name before uttering their commands. While this was necessary primarily for technical reasons, this change resulted in a smaller diversity of command styles: as participants had to name the speech system, they no longer addressed individual devices in the environment, but addressed the system as a whole. In other words, this primarily technical constraint limited the users' interaction styles, making the participants' commands more coherent. This meant it was easier to have a uniform starting word/phrase, which taken together with the smaller variation in commands, aligns better with the capabilities of the speech recognition system.

However, some problems remained, as users did not always use consistent words to denote the different devices that they could control. For instance, a few people used "Aladin, turn on the light" as a command turning on a specific light, which would lead to an ambiguous input for the system. Other types of commands proved confusing as well, such as doors which were often addressed using the name of the room on the other side of the door (e.g. "Aladin, open the bedroom door") but, when addressed from the other side, the same door would be called by a different name ("Aladin, open the living room door"). For this reason, and because of the need to provide an easy way to teach the system new commands and correct wrongly interpreted commands, we developed a tablet interface that could be used alongside ALADIN.


Figure 5. Mockup of the tablet application

### Multimodal interface

The usefulness of a multimodal interface has previously been shown in the area of assistive technology, and especially when used in combination with speech recognition [2]. The main functions of the tablet interface (figure 5) are to (1) provide richer feedback from both the system (showing when the system is listening, or reporting possible problems) and the devices (showing which lights in the house are still on, the temperature of the thermostat, etc.), (2) function as a back-up interface for correcting misunderstood commands or as a fallback, and (3) provide an easier system for training the system.

While using this input method seemingly defeats the purpose of having voice control, we have adapted it to our user group by using large vertical buttons which can be activated using swabbing, which means a button is selected upon release of a finger input, rather than on the first contact, a method originally developed and successfully tested for older people with tremors [9]. Furthermore, the interface can be used by caregivers during the heaviest training period, or by users themselves using their existing scanning/switch inputs.

During the development of this tablet interface, we used an interactive mock-up of the application which could send and receive information from the 3D home used earlier. Because we did not yet have a functional speech recognition system, a second researcher controlled the application and navigated the 3D home from a separate interface in a Wizard of Oz setup.

The extra information offered on the tablet interface limited the variation in commands even further. By seeing feedback about the system state, users also get information about devices that can be controlled, and which states are available. For instance, in a home automation environment, users get feedback on which lights they can control, how they can address them, and which states are available (e.g. different brightness levels for dimmable lamps vs. binary on/off for non-dimmable lamps). During the tests with the tablet interface, the swabbing proved to be effective in helping users suffering from tremors, while the interface was adjusted to provide full-screen instructions during use of the voice interface, to support the training and usage process of the system.

## FUTURE WORK

Our future work will involve a series of 2-week long field trials in the homes of users, offering them a way to control their television sets using a working prototype of the ALADIN system working via infrared commands. While this provides a 'plug and play' solution to integrating with existing systems, the system will not know what the infrared commands mean, so no correction is possible, and the system will not be able to share common concepts in different but similar commands (such as "*volume* up" and "*volume* down"). More work will be needed to integrate with existing systems for home automation, entertainment and ICT, such as mapping the available commands using meaningful data structures.

## CONCLUSION

The results from the ALADIN project have shown that there is a lot of potential for speech interaction to improve the quality of life of people with physical impairments. Speech interaction could provide significant added value in domains such as home automation, communication and entertainment. However, the ALADIN design process has also shown that adaptation to the specific target group is necessary. To be optimally accessible, speech systems benefit from adaptation to the users' specific voice characteristics, as well as from a multimodal setup providing rich feedback and an alternative input modality.

When speech systems are adapted to the specific needs of people with physical impairments, they can become important instruments for self-care. As such, the technology can allow caretakers to focus more on the people with impairments themselves instead of operating devices. For people with physical impairments themselves, the technology can be instrumental in becoming more independent, and in achieving a better quality of life.

## REFERENCES


1. CastleOS Software, LLC. CastleOS – Home automation made simple, http://www.castleos.com
2. Christensen, H., Casanuevo, I., Cunningham, S., Green, P., Hain, T. homeService: Voice-enabled assistive technology in the home using cloud-based automatic speech recognition. SLPAT 2013, Grenoble, France.
3. Derndorfer, C. 2011. "Adapted keyboard for children with motor disabilities" https://www.flickr.com/photos/christophd/5691855731/in/photostream is licensed under CC BY SA 2.0, retrieved October 6, 2014.
4. Hall, J.L. (joebeone) 2006. "sip-and-puff device" http://en.wikipedia.org/wiki/File:Sip-and-puff_device.jpg is licensed under CC BY 2.5, retrieved October 6, 2014.
5. NanoPac, Inc. Sicare Pilot Pro – Control Your Environment, http://www.nanopac.com/
6. John D. Gould, John Conti, and Todd Hovanyecz. 1983. Composing letters with a simulated listening typewriter. *Commun. ACM* 26, 4 (April 1983), 295-308.
7. Noyes, J. and Frankish, C. Speech recognition technology for individuals with disabilities. *Augmentative and Alternative Communication 8*, 4 (1992), 297–303.
8. Sleeswijk Visser, F., Stappers, P.J., van der Lugt, R., Sanders, E.B.N. (2005). Contextmapping: Experiences from practice. CoDesign: International Journal of CoCreation in Design and Arts, 1(2), 119-149.
9. Wacharamanotham, C. et al. 2011. Evaluating swabbing. *Proc. CHI '11,* (2011), 623.