Effects of Personal Characteristics on Music Recommender Systems with Different Levels of Controllability

Yucheng Jin
KU Leuven
Leuven, Belgium
yucheng.jin@cs.kuleuven.be

Nava Tintarev
TU Delft
Delft, Netherlands
n.tintarev@tudelft.nl

Katrien Verbert
KU Leuven
Leuven, Belgium
katrien.verbert@cs.kuleuven.be

ABSTRACT

Previous research has found that enabling users to control the recommendation process increases user satisfaction. However, providing additional controls also increases cognitive load, and different users have different needs for control. Therefore, in this study, we investigate the effect of two personal characteristics: *musical sophistication* and *visual memory capacity*. We designed a visual user interface, on top of a commercial music recommender, with different controls: interactions with *recommendations* (i.e., the output of a recommender system), the *user profile* (i.e., the top listened songs), and *algorithm parameters* (i.e., weights in an algorithm). We created eight experimental settings with combinations of these three user controls and conducted a between-subjects study (N=240), to explore the effect on *cognitive load and recommendation acceptance* for different personal characteristics. We found that controlling *recommendations* is the most favorable single control element. In addition, controlling *user profile* and *algorithm parameters* was the most beneficial setting with multiple controls. Moreover, the participants with high *musical sophistication* perceived recommendations to be of higher quality, which in turn lead to higher recommendation acceptance. However, we found no effect of *visual working memory* on either cognitive load or recommendation acceptance. This work contributes an understanding of how to design control that hits the sweet spot between the perceived quality of recommendations and acceptable cognitive load.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**;

KEYWORDS

User control, personal characteristics, cognitive load, recommendation acceptance

1 INTRODUCTION

Traditional user interfaces of recommender systems present the recommendation results with limited feedback possibilities, only allowing the users to indicate how much they like a recommendation. In contrast, Interactive Recommender Systems have been found to improve user satisfaction and perceived effectiveness by providing a visualization where users can inspect the recommender process and control the system to receive better recommendations [11].

Controllability indicates how much the system supports the user to configure the recommender process to improve the recommendations [11]. It has been regarded as an important index to evaluate the overall user experience of recommender systems, as lack of user control can negatively influence the perceived quality of recommendations [9], such as not allowing the user to reject repeating recommendations. In order to address this problem, a variety of recommender systems have components to provide feedback to recommendations [6, 8, 21, 26, 28], modify the user profile [2, 13, 15, 30], and adjust various settings of the recommender engine itself, such as parameter weights [3, 4, 10, 19, 25].

On the one hand, controls empower users to influence their recommendations to a greater extent. On the other hand, additional controls increase cognitive load [1, 12]. The user preference for interaction methods in recommender systems depends on several personal characteristics such as domain knowledge, trust propensity, and choice persistence [17].

However, recommender systems typically offer a "one-size-fits-all" approach with the same user controls for all users. To better understand the interaction between different controls, it is necessary to take into account the influence of user's *personal characteristics* and test combinations of control components.

Our study aims to provide the groundwork for developing recommender systems which **offer rich user control while ensuring acceptable cognitive load**. More specifically, we investigate the following questions:

RQ1: How do different settings of user controls influence cognitive load?

RQ2: How do different settings of user controls influence recommendation acceptance?

RQ3: How do personal characteristics — musical sophistication and visual memory capacity — influence *cognitive load*?

RQ4: How do personal characteristics — musical sophistication and visual memory capacity — influence *recommendation acceptance*?

Our contributions in this paper are the following:

(1) A systematic full factorial user study to investigate the interaction effects of three levels of control on cognitive load and recommendation acceptance. We discuss three main effects with three two-way interactions and one three-way interaction and show which a particular control works well or only when combined with another.

(2) We find that a significant effect of musical sophistication on acceptance, which is mediated by perceived quality. While visual memory capacity has no significant effect on cognitive load and acceptance.

(3) Based on the framework of the user-centric evaluation of recommender of Knijnenburg et al. [18], we construct a user-control-aware model for recommender systems, which provides the theoretical basis for designing personalized user control over recommender systems.

This paper is organized as follows: we first introduce related work. We then describe the implemented user controls in the music recommender system. The next section introduces our research methodology, followed by the design of study and results. Finally, we conclude with a discussion of study findings and limitations.

2 RELATED WORK

This section discusses related work on interactive recommender systems that support user control and research on personal characteristics that influence the user experience of recommender systems.

2.1 User Control in Recommender Systems

Like other branches of human-computer interaction research [23, 31], controllability has been an integral part of research on interactive recommender systems. Previous work shows a positive effect of user control on user satisfaction [9, 25] and perceived quality [27] of recommendations. TasteWeights [3], LinkedVis [4] and SetFusion [25] use sliders to revise user profile data and adjust the weights of the recommender engine components, thereby improving recommendation accuracy and user experience. As a result, users gain insight into how their actions affect the recommendations in real-time.

Some systems use the distance between data nodes and the active user to represent the weight of the selected node, which allows users to modify recommendation preferences by adjusting the distances [2, 15, 24]. PARIS-Ad researches the effects of user control on targeted advertising [13]. The approach allows the user to adjust his/her profile with drop-down lists and check-lists and visualizes the recommendation process in a flowchart. MusiCube refines the recommendations by asking the user to rate as many of the resulting items as possible [29]. All these systems demonstrate that user control has a prominent impact on the accuracy and effectiveness of recommendations.

Jin et al. have investigated the effects of three levels of user control on cognitive load and recommendation acceptance in an additive way [12]. However, to the best of our knowledge, no work has yet investigated how interaction among different control components influences the cognitive load and recommendation acceptance, for users with different personal characteristics. We elaborate on these characteristics in the next section.

2.2 Personal Characteristics

Improving the user experience of interactive recommender systems requires understanding the controllability needs of individual users. Knijnenburg et al. [17] investigated how three personal characteristics, domain knowledge, trusting propensity and choice persistence, influence the user preference of interaction methods in recommender systems. We propose a control-aware recommender model that aims to find the proper settings of control that increases recommendation acceptance and requires the moderate cognitive load. Several works found that *visual working memory* is a factor that affects cognitive load in adaptation of interactive systems [7, 20, 32]. Moreover, the *musical sophistication index* (MSI) is regarded as an effective way to measure domain knowledge for music recommenders, and it shows a strong correlation with music preferences [22]. Both characteristics have been found to influence perceptions of diversity in music recommendation [14].

These previous studies suggest that *visual working memory and musical sophistication* may be individual characteristics that influence the effectiveness of adaptation. In our study, we investigate the effects of these two characteristics on cognitive load and recommendation acceptance.

3 SYSTEM DESIGN AND USER CONTROL

We used the Spotify API [1] to design a music recommender system and to present the user controls for three distinct recommender components. In this section, we describe how we use this API for controlling the recommendation process as well as our user interface and interaction controls.

3.1 Recommendation Algorithms

Our system generates a play-list style listening experience based on three types of *seeds*: artists, tracks, and genres. We use the active user's top artists, tracks, and genres as input seeds. It is worth noting that the top artists and tracks are calculated by affinity, which is a measure of expected user preference for a particular track or artist based on her/his listening history. The number of songs recommended through the use of a particular seed depends on the weight of the seed's type, and the priority of the used seed among the seeds of the same type.

Moreover, it is possible in the Spotify API to specify the track attributes which affect recommendations such as *loudness, danceability, and valence*. Tracks with the attribute values nearest to the target values will be preferred and all target values will be weighted equally in ranking results.

We created four scenarios for the user task of selecting music, with each scenario represented by setting a pair of audio feature values between 0.0 and 1.0. We set a value for each scenario based on the explanation of audio feature value in Spotify API. The used scenarios include: *"Rock night - my life needs passion"* assigning attribute "energy" between 0.6 and 1.0; *"Dance party - dance till the world ends"* setting "danceability" between 0.6 and 1.0; *"A joyful after all exams"* with "danceability" between 0.6 and 1.0; *"Cannot live without hip-hop"* with "speechiness" from 0.33 to 0.66.

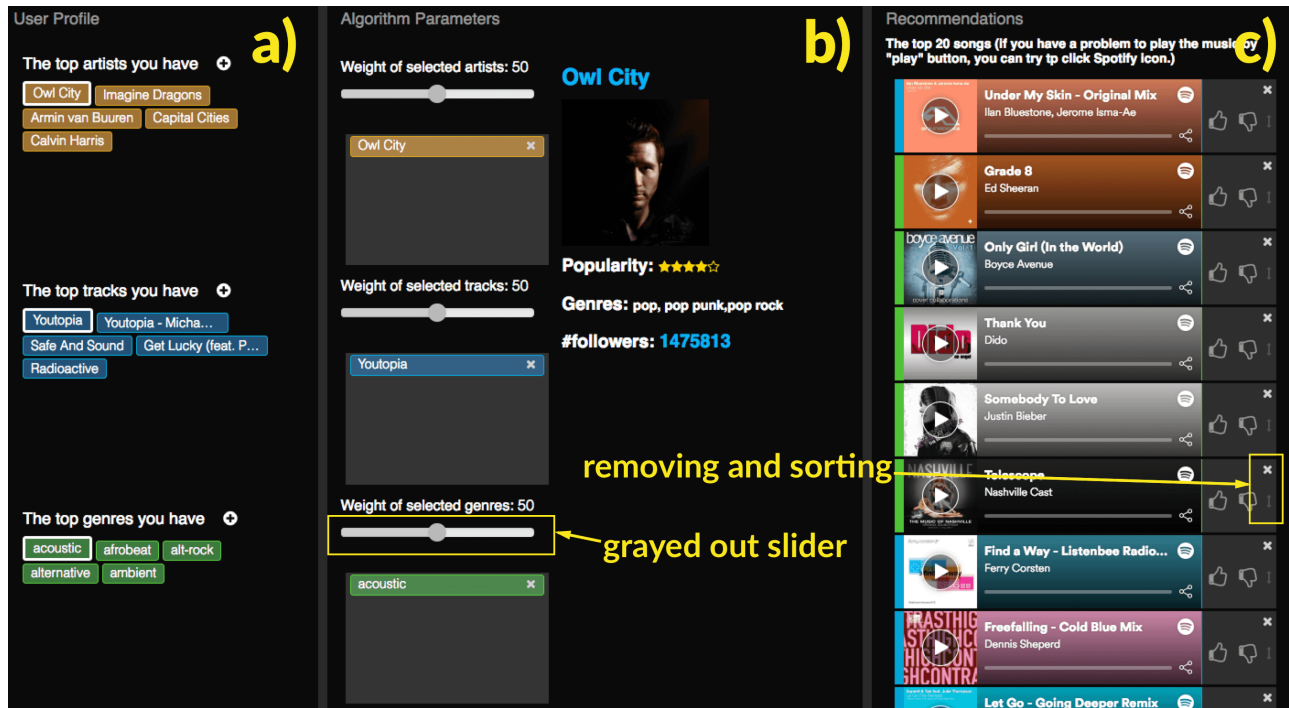[1]https://developer.spotify.com/web-api

Figure 1: a): the recommendation source shows available top artists, tracks and genre tags. b): the recommendation processor enables users to adjust the weight of the input data type and individual data items. c): play-list style recommendations. Some UI controls are disabled in specific settings of user control, e.g., the sliders in b) are grayed out in the setting 5: REC*PRO.

3.2 User Interface and Visualizations

The user interface of the recommender is featured with 'drag and drop' interactions. The interface consists of three parts, as presented in Figure 1.

(a) The *user profile* works as a warehouse of source data, such as top artists, top tracks, and top genres, generated from past listening history.

(b) The *algorithm parameters* shows areas in which source items can be dropped from part (a). The dropped data are bound to UI controls such as sliders or sortable lists for weight adjustment. It also contains an additional info view to inspect details of selected data items.

(c) The *recommendations*: the recommended results are shown in a play-list style.

As presented in Figure 1, we use three distinct colors to represent the recommendation source data as visual cues: brown for artists, green for tracks, and blue for genres. Additional source data for a particular type is loaded by clicking the "+" icon next to the title of source data type. Likewise, we use the same color schema to code the seeds (a), selected source data and data type slider (b), and recommendations (c). As a result, the visual cues show the relation among the data in three steps of the recommendation process. When users click on a particular data item in the recommendation processor, the corresponding recommended items are highlighted, and an additional info view displays its details.

Components	User controls
User profile (**PRO**)	Select which user profile will be used in the recommender engine and check additional info of the user profile
Algorithm parameters (**PAR**)	Modify the weight of the selected or generated data in the recommender engine
Recommendations (**REC**)	Remove and sort recommendations

Table 1: The three types of user control employed in our study.

3.3 Interactions and User Controls

The interactive recommendation framework proposed by He et al. [11] defines three main components of interactive recommenders. Based on this framework, we defined three user control components in our study: (1) user profile (PRO), (2) algorithm parameters (PAR), (3) recommendations (REC) (see Table 1).

*3.3.1 Control for user profile (**PRO**).* This type of control influences the seed items used for recommendation. A drag and drop interface allows users to intuitively add a new source data item to update the recommendations (Figure 1(a)). When a preferred source item is dropped to the recommendation processor, a progress

animation will play until the end of the processing. Users are also able to simply remove a dropped data item from the processor by clicking the corresponding **"x"** icon. Moreover, by selecting an individual item, users can inspect its detail: artists are accompanied by their name, an image, popularity, genres, and number of followers, tracks are shown with their name, album cover, and audio clip, and genres are accompanied by a play-list whose name contains the selected genre tag.

*3.3.2 Control for algorithm parameters (**PAR**).* This type of control allows users to tweak the influence of different underlying algorithms. To support this level of control, multiple UI components are developed to adjust the weight associated with the type of data items, or the weight associated with an individual data item. Users are able to specify their preferences for each data type by manipulating a slider for each data type. By sorting the list of dropped data items, users can set the weight of each item in this list (Figure 1(b)).

*3.3.3 Control for recommendations (**REC**).* This type of control influences the recommended songs directly. Since the order of items in a list will affect the experience of recommendations [33], manipulations on recommendations include reordering tracks in a play-list. It also allows users to remove an unwanted track from a play-list. When doing so, a new recommendation candidate replaces the removed item. The action of removing can be regarded as a kind of implicit feedback to recommendations. Although a rating function has been implemented for each item in a play-list, the rating data is not used to update the user's preference for music recommendations. Therefore, user rating is not considered as a user control for the purposes of this study.

4 RESEARCH METHODOLOGY

Our experiment is designed based on a user-centric evaluation framework for recommender systems of Knijnenburg et al. [18]. The framework provides a way to measure the influences of objective and subjective factors on the user experience and interaction. By employing this framework, we can explore the *relationships* between Objective System Aspects (OSA), Subjective System Aspects (SSA), User Experience (EXP), Interaction (INT), Situational Characteristics (SC), and Personal Characteristics (PC). Figure 2 summarizes the interaction between components in this framework. OSA relates to the underlying recommendation algorithms and user interfaces, which influences the users' perception (e.g., perceived quality of the recommendations) of the recommender. In this framework, the users' subjective factors are defined through SSA, which has a direct influence on EXP (e.g., the satisfaction of recommendation) and Interaction (e.g., rating the recommended songs). SSA is used as the moderators for the OSA's influence on EXP and INT, which means that the influence from OSA to EXP and INT is through SSA. Moreover, this framework also considers the effects of situational characteristics, such as trust, and personal characteristics, such as demographics, on EXP and INT.

More specifically, we consider *musical sophistication* and *visual memory capacity* as PC, *perceived quality, accuracy, and diversity* as SSA, and *satisfaction, effectiveness, choice difficulty,cognitive load* as EXP. We use the different settings of user controls as OSA in the
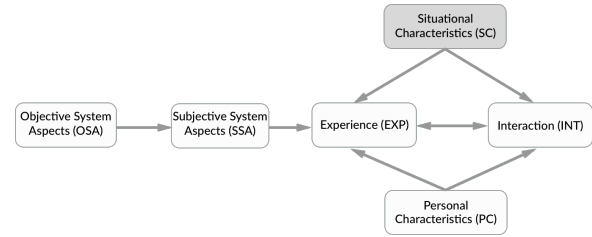
Figure 2: The user-centric evaluation framework of recommenders as used in our experiment, while our model does not contain situational characteristics (SC).

framework. We also count the participants' likes of the recommendations and the number of interaction (i.e., the number of clicks) into our model as INT.

We measure musical sophistication using the Goldsmiths Musical Sophistication Index (Gold-MSI) [2], consisting of ten selected questions with seven-point Likert scales. The visual memory capacity is measured by a visual memory capacity challenge [3] represented by a level ranging from 1 to 48. In the test, a number of tiles are highlighted at a time, and participants need to select the tiles in the correct order afterward. The number of tiles increases until the participant makes three mistakes. The memory test was also used as a quality control: both in terms of participant motivation, and the effect of poor memory on the main task. The rest of the above listed factors, *trust, perceived quality, perceived accuracy, perceived diversity, satisfaction, effectiveness and choice difficulty*, are measured by the questionnaires provided by the framework of Knijnenburg et al. [18]. Each of these factors contains four question items. We use a classic cognitive load testing questionnaire, the NASA-TLX [4], to assess the cognitive load on six aspects: mental demand, physical demand, temporal demand, performance, effort, and frustration. All question items are measured on a seven-point Likert scale from Completely disagree to Completely agree.

5 EXPERIMENTAL DESIGN

We employed a between-subjects study to investigate the effects of interactions among different user control on cognitive load and recommendation acceptance. We consider each of three user control components as a variable. By following the 2x2x2 factorial design we created eight experimental settings (Table 2), which allows us to analyze three main effects, three two-way interactions, and one three-way interaction. Each experimental setting is evaluated by a group of participants (N=30). Of note, to minimize the effects of UI layout, all settings have the same UI and disable the unsupported UI controls, e.g., graying out sliders.

5.1 Procedure

The **experimental task** is to compose a play-list for the chosen scenario by interacting with the recommender system. The play-list is generated automatically by Spotify API with specifying the value

---

	REC	PRO	PAR
Setting 1			
Setting 2	*		
Setting 3		*	
Setting 4			*
Setting 5	*	*	
Setting 6	*		*
Setting 7		*	*
Setting 8	*	*	*

Table 2: Experimental settings: a cell filled by "*" indicates this control feature is available in the corresponding setting. Setting 1 is a baseline.

range of an audio feature for the scenario. The procedure contains the following steps:

(1) *Tutorial of study* - Participants were invited to read the description of the user study and to choose a scenario for generating a play-list. Then, they were asked to watch a task tutorial. Only the features of the particular setting were shown in this video. The "Start" button of the study was only activated after finishing the tutorial. Users logged in with their Spotify accounts to our experimental system, so that our recommenders could leverage the Spotify API and user listening history to generate "real" recommendations.

(2) *Pre-study questionnaire* - This questionnaire collects user demographics and measures the user's personal characteristics such as musical sophistication, visual memory capacity, and their trust in recommender systems.

(3) *Manipulating Recommender and rating songs* - Participants were presented with play-list style recommendations (Figure 1 c). The play buttons allow users to listen to 30-second excerpts for the selected song. While users can access to a full track by clicking the Spotify icon beside the song title. Conditions were altered on a between-subjects basis. Each participant was presented with only one setting of user control. For each setting, initial recommendations are generated based on the selected top three artists, top two tracks, and top one genre. According to the controls provided in a particular setting, participants were able to manipulate the recommendation process. To ensure that participants spent enough time to explore recommendations, the questionnaire link was only activated after 10 minutes. After tweaking the visualization, participants were asked to rate the top-20 recommended songs that resulted from their interactions.

(4) *Post-study questionnaire* - Participants were asked to evaluate the perceived quality, perceived accuracy, perceived diversity, satisfaction, effectiveness, and choice difficulty of the recommender system. After answering all the questions, participants were given opportunities to provide free-text comments on their opinions and suggestions about our recommender.

5.2 Hypotheses

In this study, we evaluated a music recommender system with eight settings of user control to address the questions of how different settings of user control affect cognitive load (RQ1) and acceptance (RQ2). We also answer the questions of which specific *personal characteristics* (musical sophistication, visual memory capacity) influence cognitive load (RQ3) and recommendation acceptance (RQ4). Therefore, we have six hypotheses:

H1: The settings of user control significantly influence cognitive load.

H2: The settings of user control significantly influence recommendation acceptance.

H3: The *visual memory capacity* will *negatively* correlate to cognitive load.

H4: The *visual memory capacity* will *positively* correlate to recommendation acceptance.

H5: The *musical sophistication* will *negatively* correlate to cognitive load.

H6: The *musical sophistication* will *positively* correlate to recommendation acceptance.

6 RESULTS

6.1 Analytical approaches

We employ several validated questions [18] to measure each subjective factor in a questionnaire such as perceived quality, perceived diversity, perceived accuracy, effectiveness, satisfaction, and choice difficulty. To establish the validity of these question items, we perform a Confirmatory Factor Analysis (CFA) before evaluation. We eliminate the factors *perceived diversity* and *trust* from the model because of low AVE [5] (0.41 and 0.48) which are lower than the recommended value 0.5. Besides, we also remove the factor *satisfaction* based on the modification indices, because all the items of satisfaction load on perceived quality are large. As a result, we refine the answers to our questions and establish the validity of the factors in our study.

Figure 3 shows our fitted SEM model which consists of eight experimental conditions and five subjective factors; perceived accuracy, perceived quality, effectiveness, choice difficulty, and satisfaction. Objective system aspects (OSA) are represented by experimental conditions. Based on previous studies [18], we choose two factors for subjective system aspects (SSA): perceived accuracy and perceived quality. In addition, we define three factors: effectiveness, choice difficulty, and satisfaction for user experience (EXP). In interaction (INT), we count the number of likes and the interaction times. Moreover, this model also takes the cognitive load as a component.

The fit of our SEM model is adequate: χ^2_{98} = 257.410, p <.001; root mean squared error of approximation (RMSEA)= 0.083; Comparative Fit Index (CFI) = 0.980; Turker-Lewis Index (TLI) = 0.968.

To investigate the effects between different factors, we conducted a structural equation model (SEM) analysis for the logged data and questionnaire results by using the R toolkit Lavaan [6]. All answers to the questions are modeled as ordinal variables. We introduce three

---

[5] AVE is short for average variance extracted. For a given factor, it is the average of the R^2 values of the factor's question items.
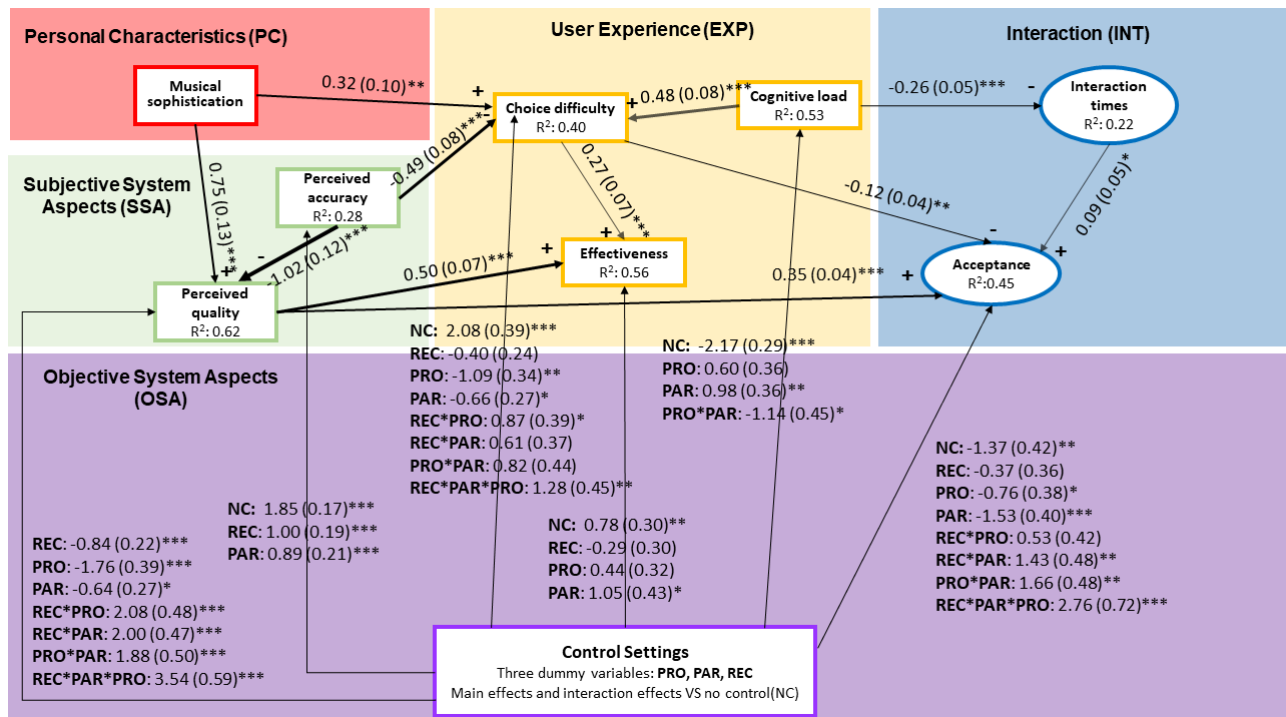[6] http://lavaan.ugent.be/, retrieved February 2018

Figure 3: The structured equation modeling (SEM) results. The number (thickness) on the arrows represents the β coefficients and standard error of the effect. Significance: *** p < .001, ** p < .01, * p < .05. R^2 is the proportion of variance explained by the model. Factors are scaled to have an SD of 1.

dummy variables *REC* (control for recommendations), *PRO* (control for user profile), and *PAR* (control for algorithm parameters) to represent the settings of user control for our music recommender. SEM is able to analyze the effects in an integrative structure where we can associate all the detected effects.

6.2 Participants

240 valid participants (Age: Mean = 27.98, SD = 7.07; Gender: Female = 55.42%, Male = 44.58%) were recruited with Amazon Mechanical Turk (mTurk), and were paid $2.00 USD for an estimated completion time of 30 minutes. Three participants results were rejected due to outliers data. The participants were required to have a minimum approval rating of 90%. We recorded the unique worker IDs of participants who completed the experiment to avoid repeated participation. In addition, to ensure the quality of responses, we rejected the responses which contained contradicting answers to the questions measuring the same factor.

6.3 General results

In this section, we present the results of acceptance and cognitive load for each setting across all users.

Cognitive load. The results of SEM (see Figure 3) show that *the control settings directly affect cognitive load.* More specifically, controls having PRO and PAR positively influence cognitive load, while

the interaction effect of PRO*PAR has a significantly negative effect on cognitive load (also see Figure 4c). In turn, the increased cognitive load also increases the choice difficulty and decreases the interaction times.

Thus, we can accept the hypothesis **H1**: *The settings of user control significantly influence cognitive load.*

Acceptance. In addition, this model shows that the *settings of control significantly affect acceptance* directly or through the mediator 'perceived quality'. For the direct influence, the main effect of two control components PRO and PAR show significantly negative effects on acceptance. In contrast, the two-way interaction effects and three-way interaction effects show significantly positive effects on acceptance (Figure 4a shows the marginal effects of control settings and their interaction on acceptance). For the indirect influence mediated by perceived quality, three main effects show significantly negative effects of control conditions on quality, while all interaction effects show significantly positive effects on quality (Figure 4g shows the marginal effects of control settings and their interaction on perceived quality). Moreover, quality positively influences acceptance.

Thus, we can accept the hypothesis **H2**: *The settings of user control significantly influence recommendation acceptance.*

Other interactions. Additionally, the results of SEM (see Figure 3) show that the settings of control (OSA) significantly correlate with
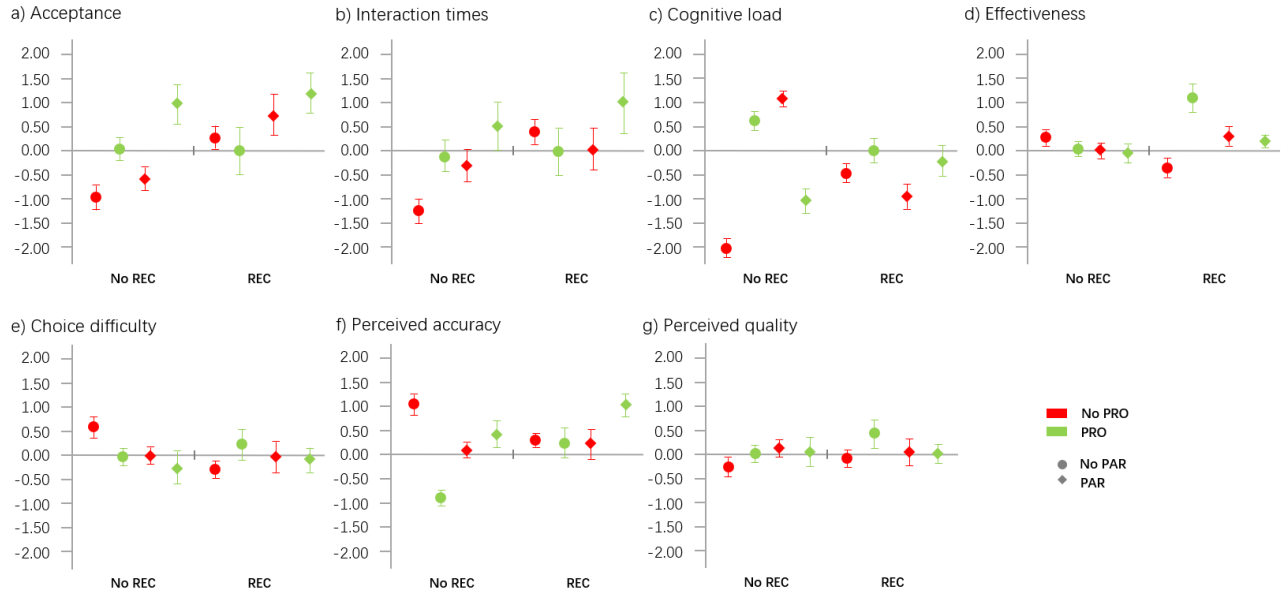
Figure 4: Marginal effects for three control components (REC, PRO, and PAR) on user interactions: a) acceptance and b) interaction time; user experience: c) cognitive load, d) effectiveness, and e) choice difficulty; and subjective aspects: f) accuracy and g) quality. Legend given for PRO and PAR.

all the measured factors of Subjective System Aspects (SSA) and User Experience (EXP) directly.

6.4 Personal characteristics

In this section, we summarize the effects of two personal characteristics, visual memory, and musical sophistication, on cognitive load and recommendation acceptance.

6.4.1 Visual memory. The SEM did not show a significant effect of visual memory on cognitive load or acceptance (INT), and is not depicted in Figure 3 (PC). This suggests that users' visual memory does not correlate with cognitive load or acceptance. Therefore, we remove the visual memory in our model. As a result, we reject two hypotheses **H3:** the *visual memory capacity* will *negatively* influence cognitive load; and **H4:** the *visual memory capacity* will *positively* influence recommendation acceptance.

6.4.2 Musical sophistication. Musical sophistication (PC) has a positive effect on perceived quality, which in turn leads to a higher recommendation acceptance (PC→SSA→INT). Meanwhile, the high perceived quality resulting from high musical sophistication may also increase effectiveness and acceptance. In contrast, increasing choice difficulty leads to high effectiveness, and lower acceptance (PC→SSA→EXP→INT). Thus, choice difficulty acts as a mediator.

We reject the hypothesis **H5:** the *musical sophistication* will *negatively* influence cognitive load; and accept the hypothesis **H6:** the *musical sophistication* will *positively* influence recommendation acceptance.

7 DISCUSSION

Our results show that the settings of user control significantly influence cognitive load and recommendation acceptance. We discuss the results by the main effects and interaction effects in a 2x2x2 factorial design.

Moreover, we discuss how visual memory and musical sophistication affect cognitive load and recommendation acceptance.

7.1 Main effects

We discuss the main effects of three control components. Increased control level; from control of recommendations (REC) to user profile (PRO), to algorithm parameters (PAR); leads to higher cognitive load (see Figure 4c). The increased cognitive load, in turn, leads to lower interaction times. Comparing to the control of algorithm parameters (PAR) or user profile (PRO), the control of recommendations (REC) introduces the least cognitive load and supports users in finding songs they like.

We observe that most existing music recommender systems only allow users to manipulate the recommendation results, e.g., users provide feedback to a recommender through acceptance. However, the control of recommendations is a limited operation that does not allow users to understand or control the deep mechanism of recommendations.

7.2 Two-way interaction effects

Adding multiple controls allows us to improve on existing systems w.r.t. control, and do not necessarily result in higher cognitive load. Adding an additional control component to algorithm parameters increases the acceptance of recommended songs significantly.

Interestingly, all the settings that combine two control components do *not* lead to significantly higher cognitive load than using only one control component. We even find that users' cognitive load is significantly *lower* for (PRO*PAR) than (PRO, PAR), which shows a benefit of combining user profile and algorithm parameters in user control. Moreover, combing multiple control components potentially increases acceptance without increasing cognitive load significantly. Arguably, it is beneficial to combine multiple control components in terms of acceptance and cognitive load.

7.3 Three-way interaction effects

The interaction of PRO*PAR*REC tends to increase acceptance (see Figure 4a), and it does not lead to higher cognitive load (see Figure 4c). Moreover, it also tends to increase interaction times and accuracy. Therefore, we may consider having three control components in a system.

Consequently, we answer two research questions. **RQ1**: *How do interactions with different control components influence cognitive load?* Combing either two or three control components does not significantly increase cognitive load. **RQ2**: *How will interactions with different control components influence recommendation acceptance?* It seems that combining PAR with a second control component or combing three control components increases acceptance significantly.

7.4 Effects of personal characteristics

Having observed the trends across all users, we survey the difference in cognitive load and item acceptance due to personal characteristics. We study two kinds of characteristics: visual working memory and musical sophistication.

7.4.1 Visual working memory. The SEM model suggests that visual memory is not a significant factor that affects the cognitive load of controlling recommender systems. The cognitive load for the type of controls used may not be strongly affected by individual differences in visual working memory. In other words, controlling the more advanced recommendation components in this study does not seem to demand a high visual memory. In addition, we did not find an effect of visual memory on acceptance (or perceived accuracy and quality).

7.4.2 Musical sophistication. Our results imply that high musical sophistication allows users to perceive higher recommendation quality, and may thereby be more likely to accept recommended items. However, higher musical sophistication also increases choice difficulty, which thereby decreases acceptance.

One possible explanation is that users with higher musical sophistication are able to leverage different control components to explore songs, and this influences their perception of recommendation quality, thereby accepting more songs.

Thus, we answer two other research questions. **RQ3**: *How do personal characteristics influence cognitive load?* Musical sophistication does not significantly influence cognitive load. **RQ4**: *How do personal characteristics influence acceptance?* Musical sophistication positively influences recommendation acceptance indirectly through perceived quality. While musical sophistication also negatively influence acceptance via choice difficulty.

8 LIMITATIONS

First, to control the duration of the user study, by default, participants were provided with only 30-second excerpts provided by the Spotify service. Although we think the excerpts are able to represent the tracks, they may present incomplete audio features such as tempo.

Second, to ensure sufficient user engagement in testing the system, we required users to spend at least ten minutes to control the system and listen to recommended songs. Thus, the recorded user actions may not reflect the real users' behavior in natural music listening environments.

In addition, we present the same user interface in all experimental settings to avoid the potential effects of variation of the layout. For some unsupported features, we simply grayed out the interface components which might be inconsistent with user expectations while seeing these components.

Lastly, some researchers [5, 16] have argued the potential issues of using the crowd-sourcing platform like Amazon Mechanical Turk to evaluate a system with complex tasks. Therefore, we cannot ignore these issues while interpreting some users' responses. However, best practices to ensure high-quality data in crowd-sourcing tasks have been used in this experiment.

9 CONCLUSIONS

We have presented an in-depth study to investigate the interaction effects of different user controls on cognitive load and acceptance. We categorize the settings of control by the number of control components implemented over the systems:

- The setting of no user control (baseline) has the lowest cognitive load, but it also receives the lowest acceptance. The control of *REC* is the winner among the settings having a single control component.
- For the settings of having two control components, the control of *PRO*PAR* results in higher acceptance and lower cognitive load – arguably the sweet spot of the two variables.
- The setting with REC*PRO*PAR leads to the highest acceptance but does not significantly increase cognitive load.

Moreover, we also find that *musical sophistication* positively (indirectly) influences recommendation acceptance. The settings of user control had direct effects on all investigated subjective factors and acceptance.

The findings shed light on designing personalized user controls that allow users to perceive high-quality recommendations with the moderate cognitive load.

Our future work will focus on three directions. First, it is important to extend this model by investigating other potential personal characteristics that influence cognitive load beyond musical sophistication, such as *choice persistence* [17]. Second, based on this extended model, we intend to investigate the adaptive strategies that are suitable to the personal characteristics of users. Third, we plan to validate our research finding in other application domains such as online learning and exploring articles on debated topics.

10 ACKNOWLEDGEMENTS

REFERENCES

[1] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: interactive mood-based music discovery and recommendation. In *Proc. of UMAP'16*. ACM, 275–279.

[2] Fedor Bakalov, Marie-Jean Meurs, Birgitta König-Ries, Bahar Sateli, René Witte, Greg Butler, and Adrian Tsang. 2013. An approach to controlling user models and personalization effects in recommender systems. In *Proc. of IUI'13*. ACM, 49–56.

[3] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: a visual interactive hybrid recommender system. In *Proc. of RecSys'12*. ACM, 35–42.

[4] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2013. LinkedVis: exploring social and semantic career recommendations. In *Proc. of IUI'13*. ACM, 107–116.

[5] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science* 6, 1 (2011), 3–5.

[6] Li Chen and Pearl Pu. 2012. Critiquing-based recommenders: survey and emerging trends. *UMUAI'12* 22, 1-2 (2012), 125–150.

[7] Cristina Conati, Giuseppe Carenini, Enamul Hoque, Ben Steichen, and Dereck Toker. 2014. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 371–380.

[8] Cristina Gena, Roberto Brogi, Federica Cena, and Fabiana Vernero. 2011. The impact of rating scales on user's rating behavior. In *Proc. of UMAP'11*. Springer, 123–134.

[9] F Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting users in control of their recommendations. In *Proc. of RecSys'15*. ACM, 3–10.

[10] Gerald Häubl and Valerie Trifts. 2000. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science* 19, 1 (2000), 4–21.

[11] Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications* 56 (2016), 9–27.

[12] Yucheng Jin, Bruno Cardoso, and Katrien Verbert. 2017. How do different levels of user control affect cognitive load and acceptance of recommendations?. In *Proc. of IntRS co-located with RecSys'17*. CEUR-WS, 35–42.

[13] Yucheng Jin, Karsten Seipp, Erik Duval, and Katrien Verbert. 2016. Go with the flow: effects of transparency and user control on targeted advertising using flow charts. In *Proc. of AVI'16*. ACM, 68–75.

[14] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of individual traits on diversity-aware music recommender user interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, 1–9.

[15] Antti Kangasrääsiö, Dorota Glowacka, and Samuel Kaski. 2015. Improving controllability and predictability of interactive recommendation interfaces for exploratory search. In *Proc. of IUI'15*. ACM, 247–251.

[16] Aniket Kittur, Ed H Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proc. of CHI'08*. ACM, 453–456.

[17] Bart P Knijnenburg, Niels JM Reijmer, and Martijn C Willemsen. 2011. Each to his own: how different users call for different interaction methods in recommender systems. In *Proc. of RecSys'11*. ACM, 141–148.

[18] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *UMUAI* 22, 4-5 (2012), 441–504.

[19] Thomas Kramer. 2007. The effect of measurement task transparency on preference construction and evaluations of personalized recommendations. *Journal of Marketing Research* 44, 2 (2007), 224–233.

[20] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2017. Impact of Individual Differences on User Experience with a Visualization Interface for Public Engagement. In *Proc. of UMAP'17*. ACM, 247–252.

[21] Kevin McCarthy, Yasser Salem, and Barry Smyth. 2010. Experience-based critiquing: Reusing critiquing experiences to improve conversational recommendation. In *Proc. of ICCBR'10*. Springer, 480–494.

[22] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one* 9, 2 (2014), e89642.

[23] Jakob Nielsen. 1999. *Designing web usability: The practice of simplicity*. New Riders Publishing.

[24] John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proc. of CHI'08*. ACM, 1085–1088.

[25] Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: a case study with SetFusion. *IJHCS'15* 78 (2015), 43–67.

[26] Alina Pommeranz, Joost Broekens, Pascal Wiggers, Willem-Paul Brinkman, and Catholijn M Jonker. 2012. Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process.

UMUAI'12 22, 4-5 (2012), 357–397.

[27] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the user's perspective: survey of the state of the art. *UMUAI'12* 22, 4 (2012), 317–355.

[28] Pearl Pu, Li Chen, and Pratyush Kumar. 2008. Evaluating product search and recommender systems for E-commerce environments. *Electronic Commerce Research* 8, 1-2 (2008), 1–27.

[29] Yuri Saito and Takayuki Itoh. 2011. MusiCube: a visual music recommendation system featuring interactive evolutionary computing. In *Proc. of VINCI'11*. ACM, 5.

[30] James Schaffer, Tobias Höllerer, and John O'Donovan. 2015. Hypothetical Recommendation: A Study of Interactive Profile Manipulation Behavior for Recommender Systems.. In *FLAIRS Conference*. 507–512.

[31] Ben Shneiderman. [n. d.]. *Designing the User Interface*. Pearson Education India.

[32] Nava Tintarev and Judith Masthoff. 2016. Effects of Individual Differences in Working Memory on Plan Presentational Choices. *Frontiers in psychology* 7 (2016).

[33] Qian Zhao, Gediminas Adomavicius, F Maxwell Harper, Martijn Willemsen, and Joseph A Konstan. 2017. Toward Better Interactions in Recommender Systems: Cycling and Serpentining Approaches for Top-N Item Lists. In *Proc. of CSCW'17*. 1444–1453.