How Do Different Levels of User Control Affect Cognitive Load and Acceptance of Recommendations?

Yucheng Jin
Department of Computer Science, KU Leuven
Leuven, Belgium
yucheng.jin@cs.kuleuven.be

Bruno Cardoso
Department of Computer Science, KU Leuven
Leuven, Belgium
bruno.cardoso@cs.kuleuven.be

Katrien Verbert
Department of Computer Science, KU Leuven
Leuven, Belgium
katrien.verbert@cs.kuleuven.be

ABSTRACT
User control has been recognised as an important feature in recommender system, as it allows users to steer the recommendation process. Most typical user controls relate to providing ratings, editing user data, and adjusting weights of the algorithm. The cognitive load of the user may increase when using more advanced user controls. We divided common user controls into three levels (high, middle, and low) and conducted a study (N=90) to investigate how different levels of user control affect cognitive load and quality of recommendations. We designed a visualisation on top of a music recommender system that incorporates three levels of control. The study results show that high level control tends to produce the best recommendations, while requiring the highest cognitive load. However, only participants with rich experience in recommender systems are more likely to tweak such high level control, while the majority of participants still prefers low and middle level control. We validated the robustness of our findings with three different algorithms.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
User control; Cognitive load; Acceptance of recommendations.

INTRODUCTION
Recommender systems are ubiquitous today and we can find them in many application domains. These recommendation algorithms and powerful big data technologies allow applications to provide high quality recommendations to users, increasing their acceptance potential and, in turn, leading to improved user satisfaction and perceived effectiveness. Extensive research has been conducted in the past decades to develop and enhance algorithmic techniques such as content-based filtering, collaborative filtering, knowledge-based filtering and hybridisations. However, many researchers have argued that other factors beyond accuracy may influence the user experience with recommender-based platforms [22, 17].

Recently, user-centred research has gained a lot of attention in the field of recommender systems and various metrics [21, 16] of user experience assessment have been proposed, including diversity, serendipity, trust, transparency, and controllability. Enhancing the user experience from these perspectives requires effective user interaction with the system. Much of the existing literature proposes to address the well-known "black-box" issue by focusing on providing visualisations that expose the recommender algorithm to the user. Such visualisations empower the user to inspect the recommender process and further tune the system to receive better recommendations.

The metric of controllability is of particular relevance to this work and indicates how much the system supports the user to configure the recommender process to improve the recommendations. It has been regarded as an important index to evaluate the overall user experience of recommender systems, as lower levels of user control negatively influence the perceived quality of recommendations [10]. For example, a system that keeps recommending hotels to a user who has booked a hotel recently may annoy the user if the system does not provide a mechanism to reject recommendations or adjust her preferences. In order to address this problem, a variety of recommender systems have components to rate recommendations, modify user data, and adjust various settings of the recommender engine itself, such as parameter weight [8]. However, user interfaces may become difficult to understand when containing many control components [3]. Therefore, we assume that levels of user control may influence the cognitive load of the user when using the system.

To investigate this hypothesis, we used the Spotify API [1] to design a music recommender system and to explore how different levels of user control influence the cognitive load of system use. We visualise recommendations by a column based diagram and use colour to link related items in each column. It is suitable for representing the relationship between user data and recommendations. The recommender system integrates three recommender algorithms. The first one is based on the

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

Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, Como, Italy.

Control level	*Recommender components*	*Explanation*
Low level	Recommendations	Sort and rate the recommendations
Middle Level	User data	Select which user data will be used in the recommender engine and check additional info of user data
High level	Medium data	Modify the weight of the selected or generated data in the recommender engine

Table 1. Three levels of user control are defined in our study.

top seeds (top artists, top tracks and top genres) generated by the user. The second one is an item-item collaborative filtering algorithm that lists the top tracks of artists who are related to followed artists. The third one is a hybrid algorithm that combines these two algorithms.

Usually, measuring the cognitive load relies on self-reported data or analysis of physiological data. The approach of self-reporting uses questionnaires such as NASA-TLX [2] to ask users about their experience after performing tasks. In turn, the physiological data approach usually analyses EEG and eye-tracking data to predict cognitive load during the tasks. Both approaches have their strengths and weaknesses. Although using physiological data can provide real-time information, it is difficult to set up for online studies. Therefore, we use a classic cognitive load testing questionnaire, the NASA-TLX, to assess cognitive load on six aspects: mental demand, physical demand, temporal demand, performance, effort, and frustration. In addition, we also investigate the effects of different levels of user control on acceptance of recommendations by asking users to rate recommended songs.

The interactive recommendation framework proposed by He et al. [11] defines three main components in interactive recommenders: user data and context, medium, and recommendations. We therefore define different levels of user control for each component in Table 1.

Our study aims to provide the groundwork for developing high-quality recommender systems offering sufficient user control, while demanding acceptable cognitive load. Specifically, we investigate the following questions:

RQ1: Do different levels of user control have an impact on the cognitive load of using recommender systems and, if so, what is the impact?

RQ2: Do different levels of user control have an effect on acceptance of recommendations?

RQ3: Will different recommender algorithms influence the answers to RQ1 and RQ2?

Andjelkovic et al. [3] already show that users spend more effort with systems offering higher levels of user control than with systems with lower levels of user control. However, to the best of our knowledge, no comprehensive work has yet investigated to what extent varying levels of user control influence the cognitive load of using recommender systems and the perceived quality of their recommendations. With regards to related work, our contributions are the following:

1. We define three levels of user control (low, middle, high) based on estimated work load of tweaking each level of control.
2. By leveraging the metaphors of "processing" and "production", we design and develop an interactive music recommender with a drag and drop user interface to help the user understand the recommendation process.
3. We conduct a user study to investigate the user cognitive load and the perceived quality of recommendation under the three defined levels of user control. We also validate our findings with three recommender algorithms.
4. Based on our findings, we discuss the possible ways to balance levels of user control and required cognitive load in the recommendation process. In addition, we also demonstrate what kind of users are more likely to benefit from each level of user control.

This paper is organised as follows: we first introduce related work covering interactive recommenders that support user control, and research on cognitive load of recommender visualisations. We then describe the system design of our recommender system. The next section introduces the design of study, followed by results of the user study. Finally, we conclude with a discussion of study findings and limitations.

RELATED WORK

User Control in Recommender Systems

Many HCI researchers [25, 18] count controllability as one of most prominent factors that influence overall user experience with recommender systems. Current user control research focuses on rating recommendations, revising the user profile, and adjusting recommendation parameters such as weight [11]. User control has been an integral part of research on interactive recommender systems. Previous work shows a positive effect of user control on user satisfaction [20, 10] and perceived quality [22] of recommendations. We review several typical systems that increase user involvement in various stages of the recommendation process, through different levels of user control.

TasteWeights [5], LinkedVis [6] and SetFusion [20] 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 [19, 4, 14] 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. PARIS-Ad [12] researches the effects of user control on targeted advertising. It allows the user to adjust her profile with drop-down lists and check-lists, and visualises the recommendation process in a flowchart. MusiCube [24] refines the recommendations by asking the user to rate as many of the resulting items as possible. All these systems demonstrate that user control has

[2] https://humansystems.arc.nasa.gov/groups/tlx

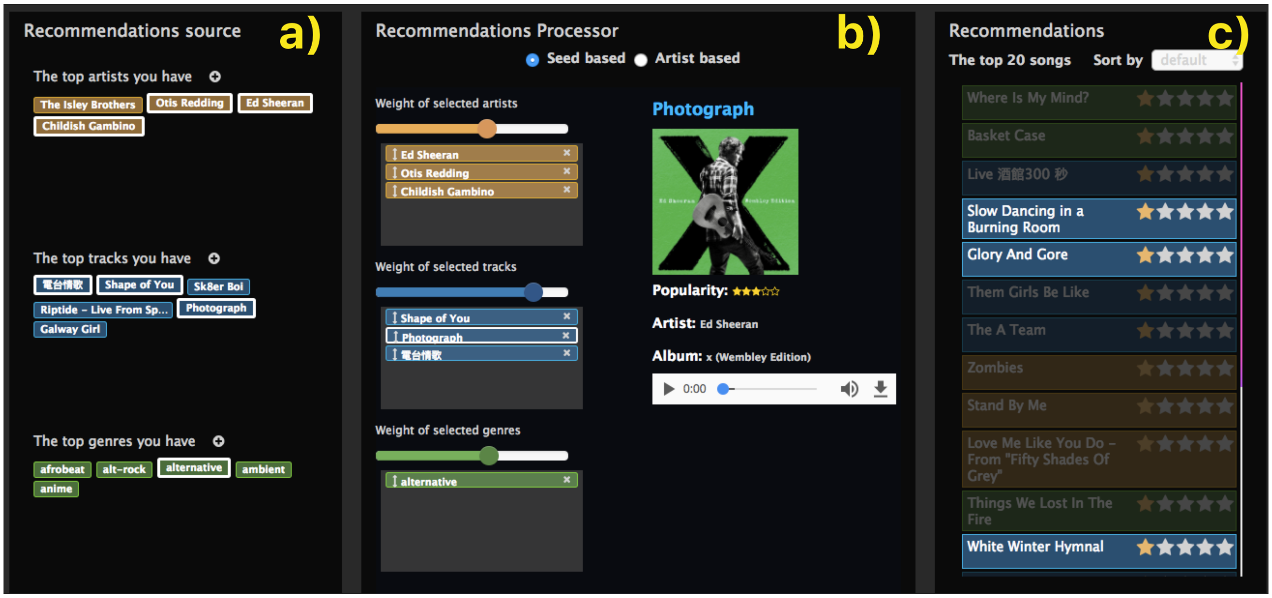

Figure 1. Visualisation of the seed based algorithm. a): the recommendation source shows available top artist, tracks and genre tags. b): the recommendation processor enables users to adjust the weight of the input data type and individual data items. c): play-list style recommendations.

a prominent impact on the accuracy and effectiveness of recommendations. However, it is not clear if varying levels of user control affect the robustness of the findings. We therefore intend to compare recommendation ratings in different experimental tasks entailing different levels of user control.

Cognitive Load

The construct of "cognitive load" is usually used to measure how many cognitive resources are taken up by activities that facilitate learning [9]. In general, cognitive load measurement is performed through the application of a post-study in the form of a self-assessment questionnaire, or the analysis of physiological data collected during task execution. The NASA task-load index (NASA-TLX) is one of the most widely used questionnaires to measure cognitive load, along six dimensions: mental demand, physical demand, temporal demand, own perception of performance, effort and frustration. Although it is not designed to measure cognitive load in real-time, it is easy to apply and reliable in many conditions.

The information visualisation community has adopted various physiological data to measure the cognitive load of using different visualisation techniques [27]. Typically, researchers analyse eye tracking [1] and brain activity [2] data to estimate the cognitive load while performing tasks. However, even though physiological methods provide the means to estimate cognitive load in real-time, the cost of hardware such as eye trackers and electroencephalography (EEG) systems and professional training for analysing produced data are substantial barriers to the widespread adoption of this approach.

Previous work has demonstrated various ways of decreasing cognitive load while improving the performance of interactive recommender systems. Schnabel et al. [9] use shortlists as digital short-term memory. Since users do not need to keep the considered items in their minds, the cognitive load is reduced. Quiroga et al. [23] pointed out that information filtering and building profiles on users' organisational behaviour is essential to reduce cognitive load.

Although we do not find that related work reveals the relation between levels of user control and cognitive load, Andjelkovic et al. [3] observed in their music recommender that additional control to new aspects such as avatars might increase cognitive load. In addition, Adil Yalçinn et al. [26] presented the Cognitive Exploration Framework, providing guidelines to reduce the cognitive load in their defined six stages of cognitive activities in visual data exploration.

In our study, we not only aim to provide effective user control to enhance the user experience with recommender systems, but also to investigate how different levels of user control affect the cognitive load and recommendation quality. Moreover, we provide groundwork for designing user-centred recommender systems that also adapt to different levels of user cognitive load.

SYSTEM DESIGN AND INTERACTIONS

Recommendation Algorithms

In order to validate our research findings with different recommender approaches, we implemented three different algorithms to generate music recommendations by using the Spotify API.

Seed based algorithm

The Spotify API provides a recommender service that 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 listening history. The number of songs

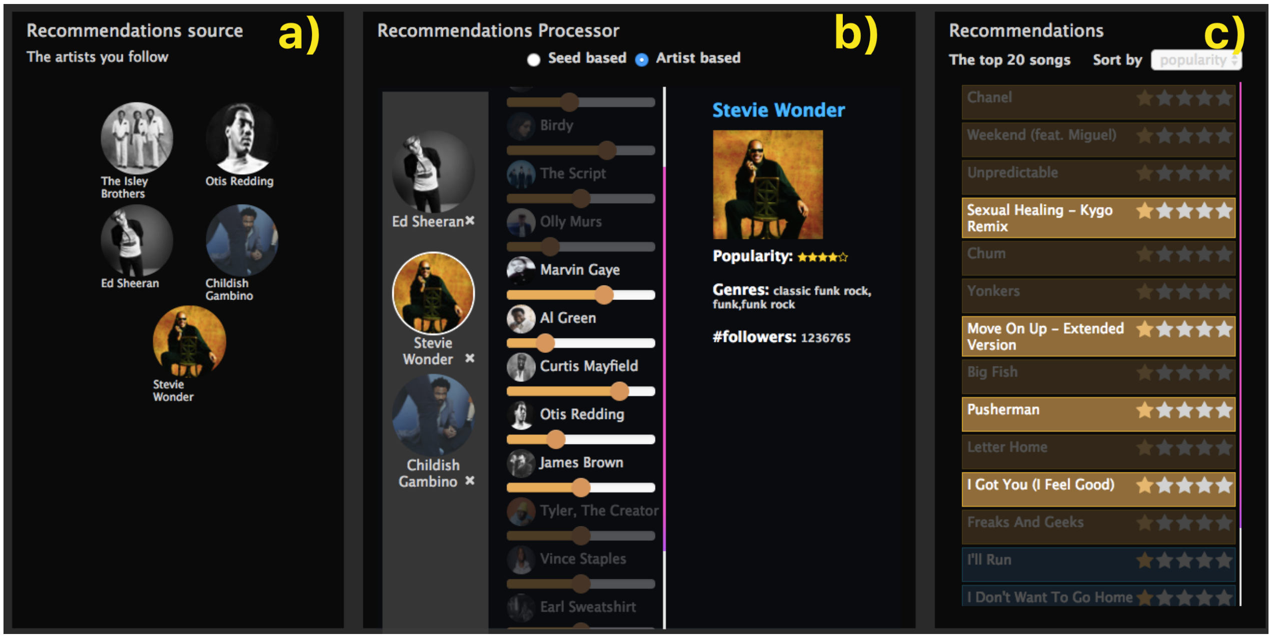

Figure 2. Visualisation of the artist based algorithm. a): the recommendation source panel shows available followed artists, b): the recommendation processor enables users to adjust the weight of related artists of selected followed artists, and c): play-list style recommendations.

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 same type.

Artist based algorithm

The artist-based algorithm uses the item-item collaborative filtering approach. First, the algorithm reads the list of user-followed artists. Then, the Spotify API allows us to find artists related to a followed artist by calculating the similarity between them, which is based on analyses of the Spotify community listening history. The top 20 tracks of these related artists are returned. The number of recommendations by an artist is proportional to the weight of the artist.

Hybrid based algorithm

The hybrid based algorithm combines the seed based algorithm and artist based algorithm. The same weight is assigned to both algorithms.

User Interface and Visualisations

The user interface of the recommender was designed using the metaphor of "processing" and "production". It consists of three parts:

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

(b) The *recommendations processor* 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. In addition, a pair of radio buttons allows the user to switch between different algorithms.

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

Visualisation of the seed based algorithm

As presented in Figure 1(a), we use three distinct colors to represent types of recommendation source data as visual cues (yellow 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 encode the data type slider and selected source data (Figure 1 (b)), and recommendations (Figure 1 (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 will be highlighted, and an additional info view displays its details.

Visualisation of the artist based algorithm

To emphasise the concept of artist relations, this algorithm only contains artist data items represented by the corresponding artists' portraits in addition to their names (Figure 2 (a)). When users drag an artist and drop it in the selected artists block, the top five related artists of the dropped artist are shown, each with a slider to adjust its weight (Figure 2 (b)). Similar to the first visualization, recommendations are highlighted when users click on a particular artist in the recommendation processor (Figure 2 (c)) to depict their relation.

Interactions and User Controls

Our system offers several interactions to support our three levels of user control.

Low level of user control

In this level, users can sort the recommendation results by preference through a drop-down menu. Although ratings normally have no immediate effects on recommendations, we still regard recommendations feedback as a kind of low level user control. The star rating widget beside song title allows users to rate the songs in the recommendation list (Figure 1(c), Figure 2(c)).

Middle level of user control
In general, manipulating source data and checking details compose the middle level of user control. A drag and drop interface allows users to intuitively add a new source data item to update recommendations (Figure 1(a), Figure 2(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, popularity, and audio clip; and genres are accompanied by their top related artists and tracks.

High level of user control
The high level of user control allows users to tweak the underlying algorithm as a basis to further manipulate the recommendation process. 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 to an individual data item. In the seed based algorithm, users are able to specify their preferences for each data type by manipulating a slider for each data type. By sorting a list of dropped data items, users can set the weight of each item in this list (Figure 1(b)). Similarly, the weight of related artists can be manipulated by moving its associated slider in the artist based algorithm visualisation (Figure 2(b)).

EVALUATION
We evaluated our system by conducting a study on Amazon Mechanical Turk (MTurk) with 107 participants who are all active users of Spotify. 17 of our participants were rejected because of their repetitive and invalid answers. In the end, we had 90 valid participants (48 female, 42 male), their ages ranged from 20 to 48 years (mean age = 29.8 years, SD = 7.51, Median = 28). 86.67% of participants are familiar with recommender system. We paid $ 1 for each study. The average study completion time was around 33 minutes (SD = 7.23, Median = 33).

Evaluation Design
We designed a within-subjects study to investigate the effects of different levels of user control on cognitive load and acceptance of recommendations. Therefore, we created three experimental tasks T1, T2, and T3 corresponding to the different levels of user control.

T1 Users were only allowed to interact with recommendations by *sorting recommendations* (low level control) in a list. In the end, they rated each song in the list of recommended items.

T2: Users were asked to interact with recommendations by *sorting* (low level control) recommendations and *modifying recommendation source* (moderate level control). Finally, they rated each song.

T3: Users were asked to interact with recommendations by *sorting* (low level control) recommendations, *modifying recommendation source* (moderate level control), *tweaking the parameters of algorithms* (high level control). Once again, participants were asked to rate each song.

We split the 90 participants equally into three groups to validate the results with three different settings of recommender algorithms: the seed-based algorithm (*Setting 1*), the artist-based algorithm (*Setting 2*), and a hybrid of the two algorithms with equal weight (*Setting 3*).

Participants of each group tested one algorithm setting with three experimental tasks. The order of the three tasks has been mixed to avoid learning effects.

Evaluation Procedure
The participants were asked to watch a task tutorial. Only the features of the particular setting were shown in this video. After interacting with the visualization, participants were asked to rate the top-20 recommended songs that resulted from their interaction, and to fill out the NASA-TLX questionnaire to measure their cognitive load. Users had to complete this questionnaire in the three experimental tasks. At the end of the task, they were asked to fill out a questionnaire that was based on a part of the *ResQue* to evaluate the perceived quality of the recommender with all levels of user control. To assess the validity of the responses, we set contradictory questions in this questionnaire. In addition, user interactions with the different components of the visualization were recorded in a log.

RESULTS
To analyze the cognitive load, we calculated the score from participant responses to the NASA-TLX questionnaire, which ranges from 0 to 100. The higher the score is, the more cognitive load is required. Since we intend to measure the overall accuracy of a recommendation list, we apply the Breese's R-Score "utility" metric [7] to calculate a utility score. The rating score for a song ranges from 1 to 5, and the default score is 1. We also analyze responses to the *ResQue*-based questionnaire, and report the results separately for each recommender algorithm.

Cognitive load
Setting 1: seed based
Descriptive statistics show that participants have the highest cognitive load in T3 (M=57.14), followed by T2 (M=46.11) and T1 (M=31.43). We performed a one-way repeated ANOVA to test for significance. There was a significant effect for cognitive load, $F(2, 58) = 44.47$, $p<.001$. Bonferroni-corrected pairwise comparisons (sig. level = .016) revealed that T3 requires significantly higher cognitive load than T2 ($p<.001$) and T1 ($p<.001$). T2 required a significantly higher cognitive load than T1 ($p<.001$).

Setting 2: artist based
Descriptive statistics show that participants in T3 (M=50.32) have the highest cognitive load, followed by T2 (M=38.57) and

T1 (M=30.24). To test for significance, we performed a one-way repeated ANOVA test. To compensate for violations of the sphericity assumption (Mauchly's (W(df=2) = .721, (p=.010), the significance levels were corrected by Greenhouse-Geisser. The corrected score shows a significant effect for cognitive load, $F(1.56, 45.36) = 15.42$, $p<.001$. Bonferroni-corrected pairwise comparisons (sig. level = .016) revealed that T3 required significantly higher cognitive load than T2 (p=.001) and T1 (p<.001), and T2 required significantly higher load than T1 (p=.009).

Setting 3: hybrid
Descriptive statistics show that T3 (M=52.14) requires the highest cognitive load, followed by T2 (M=45.87) and T1 (M=34.44). To test for significance, we performed a one-way repeated ANOVA test. The corrected score shows a significant effect for cognitive load, $F(2, 58) = 8.54$, p=.001. Bonferroni-corrected pairwise comparisons (sig. level = .016) revealed that both T3 (p<.001) and T2 (p=.001) require significantly higher cognitive load than T1.

In general, T3 requires a significantly higher cognitive load in all three settings. But the differences between T2 and T1 and between T3 and T2 are not always significant.

Acceptance of recommendations

Setting 1: seed based
Descriptive statistics show that the list of recommendations in T3 (M=3.49) was rated higher than T2 (M=2.95) and T1 (M=2.08). A one-way repeated ANOVA test was conducted for examining significance. A significant effect is found for user rating, $F(2, 58) = 25.04$, p<.001. Bonferroni-corrected pairwise comparisons (sig. level = .016) revealed that ratings of recommendations in T3 was rated significantly higher than those in T2 (p=.003) and T1 (p=.001), and recommendations in T2 were rated significantly higher than those in T1 (p=.001).

Setting 2: artist based
Descriptive statistics show that the list of recommendations in T3 (M=3.54) was rated higher than in T2 (M=2.92) and T1 (M=2.41). The result of a one-way repeated ANOVA test shows a significant effect for user rating, $F(2, 58) = 14.68$, p<.001. Bonferroni-corrected pairwise comparisons (sig. level = .016) revealed that recommendations in T3 were rated significantly higher than in T2 (p<.001) and T1 (p=.001).

Setting 3: hybrid
Descriptive statistics show that the list of recommendations in T3 (M=3.27) was rated higher than in T2 (M=3.21) and T1 (M=2.59). The result of a one-way repeated ANOVA test shows a significant effect for user rating, $F(2, 58) = 7.80$, p<.001. Bonferroni-corrected pairwise comparisons (sig. level = .016) revealed that the lists of recommendations in T3 (p=.002) and T2 (p=.004) were rated significantly higher than in T1.

By comparing the findings of different settings, we find that the list of recommendations in T3 was always rated significantly higher than in other settings.

Settings	*Low level*	*Middle level*	*High level*
Setting 1	60.5%	27.9%	11.6%
Setting 2	54.4%,	28.3%	17.3%
Setting 3	51.9%,	26.9%	21.2%

Table 2. Percentage of interactions with each level of control in task 3.

Overall user experience

The left bar chart (Figure 3) plots users' attitudes towards the various controls of the recommender systems. Participants seem to enjoy using a drag-and-drop interface to manipulate the recommendation process. The system also allows users to express their preferences easily. In general, users like to give feedback and modify their data. However, it seems that only a part of participants would like to control more components of the system. It is worth noting that 91.1% of the participants who would like to tweak the high level control have experience with recommender systems and 95.6% of them enjoy listening to music online.

The chart on the right side illustrates the users' positive responses to our system in terms of other user experience aspects such as novelty, diversity and confidence. Users indicated that using our system was fun and that they easily became familiar with the system. Despite these merits, some users are not sure they would use this system frequently to listen to music.

Log file data

Since we intend to know how often users will interact with each user control, we also analyzed interaction data.

We report the percentage of interactions for each level of control in T3, where all levels of control are presented (Table 2). More than half of the interactions are related to low level controls, and around a quarter of clicks are related to middle level. Only a small part of clicks were done with high level controls.

DISCUSSION

In this section, we discuss the results presented in the previous section, thereby answering the research questions and evaluating the proposed hypothesis.

Overall, the results of NASA-TLX show that the higher level of user control tends to increase cognitive load (RQ1). Specially, we see that the high level user control has significantly higher cognitive load than the low level, in all of the three settings. Previous work [12, 10, 20] has reported that user control improves the accuracy of recommendations. Furthermore, the results of user ratings indicate that the level of control has a significant influence on the acceptance of recommendations (RQ2). The poor result in T1 may suffer from the unmodifiable tags for bootstrapping the system. Also in our data, we can observe that the high level user control increases the quality of recommendations in all three settings. The effects of levels of control on cognitive load and acceptance of recommendations are not statistically significant when we compare the high level control to the middle level, and the middle level control to the low level in some settings. By comparing the mean value of each result, our findings can be validated with different algorithms (RQ3). Besides, it seems that participants have

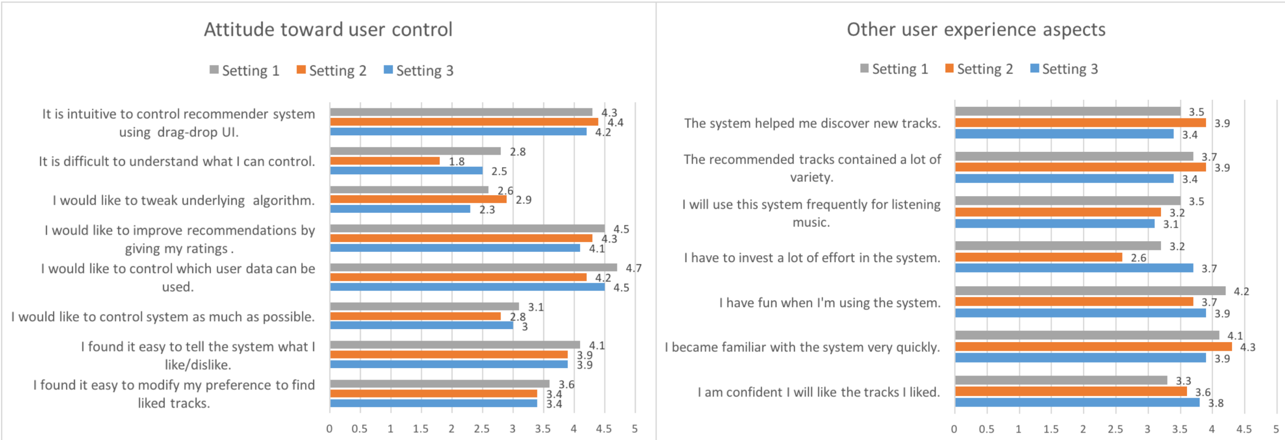

Figure 3. User responses to the *ResQue* based questionnaire in the three settings.

difficult in understanding what they can control and have less interest while performing T3 in Setting 2. A possible explanation is that the current visualization does not clearly plot the relations among artists, suggesting that a network graph could be a better option.

Log file data also suggests that users are more likely to tweak the low level and the middle level control. By looking at the user profile, we find that participants with rich experience with recommender systems and online music tend to tweak the high level user control more frequently. The majority of participants prefers to have only low and middle level user control. This may depend on user personal characteristics and domain knowledge [15]. In addition, a drag-and-drop UI seems to allow users to interact with the system intuitively. In spite of the merits in our system, users hesitate to use it for listening to music. A potential reason is that many users prefer to listen and discover music on mobile devices with simple interactions rather than on large screens with complex interaction [13].

CONCLUSION

We define three levels of user control to investigate the effects of levels of control on cognitive load and acceptance of recommendations. We designed and implemented a music recommender with three distinct settings of recommender algorithms. An online study was performed to answer research questions. We conclude with the following findings:

- By incorporating higher level of user control, cognitive load tends to increase.
- By incorporating higher level of user control, the recommendations are more likely to be accepted.
- Our research findings are generalizable to different recommender algorithms.

Our study has three main limitations: first, although we have excluded unqualified users by setting contradictory questions in questionnaires, the validity of study results may still suffer from inattentive or "spamming" users. Second, the research finding should be validated in other application domains. Third, the research findings were found based on specific user control mechanisms implemented in the study system. Our future work will focus on adapting the user interface of recommender systems to address the individual needs and preferences of users.

Acknowledgements

The research has been partially financed by the KU Leuven Research Council (grant agreement no. C24/16/017).

REFERENCES

1. Erik W Anderson. 2012. Evaluating visualisation using cognitive measures. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization*. ACM, 5.
2. Erik W Anderson, Kristin C Potter, Laura E Matzen, Jason F Shepherd, Gilbert A Preston, and Cláudio T Silva. 2011. A user study of visualisation effectiveness using EEG and cognitive load. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 791–800.
3. Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: interactive mood-based music discovery and recommendation. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. ACM, 275–279.
4. 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 *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 49–56.
5. Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: a visual interactive hybrid recommender system. In *Proceedings of the 6th ACM conference on Recommender systems*. ACM, 35–42.

6. Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2013. LinkedVis: exploring social and semantic career recommendations. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 107–116.
7. John S Breese, David Heckerman, and Carl Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 43–52.
8. Robin Burke. 2002. Hybrid recommender systems: survey and experiments. *User Modeling and User-Adapted Interaction* 12, 4 (2002), 331–370.
9. Paul Chandler and John Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and instruction* 8, 4 (1991), 293–332.
10. F Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting users in control of their recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 3–10.
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, 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 *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 68–75.
13. Mohsen Kamalzadeh, Christoph Kralj, Torsten Möller, and Michael Sedlmair. 2016. TagFlip: active mobile music discovery with social tags. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 19–30.
14. Antti Kangasrääsiö, Dorota Glowacka, and Samuel Kaski. 2015. Improving controllability and predictability of interactive recommendation interfaces for exploratory search. In *Proceedings of the 20th international conference on intelligent user interfaces*. ACM, 247–251.
15. 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 *Proceedings of the 5th ACM conference on Recommender systems*. ACM, 141–148.
16. Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.
17. Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction* 22, 1 (2012), 101–123.
18. Jakob Nielsen. 1999. *Designing web usability: The practice of simplicity*. New Riders Publishing.
19. John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1085–1088.
20. Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: a case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.
21. Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.
22. Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction* 22, 4 (2012), 317–355.
23. Luz M Quiroga, Martha E Crosby, and Marie K Iding. 2004. Reducing cognitive load. In *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)-Track 5-Volume 5*. IEEE Computer Society, 50131–1.
24. Yuri Saito and Takayuki Itoh. 2011. MusiCube: a visual music recommendation system featuring interactive evolutionary computing. In *Proceedings of the 2011 Visual Information Communication-International Symposium*. ACM, 5.
25. Ben Shneiderman. *Designing the User Interface*. Pearson Education India.
26. M Adil Yalçin, Niklas Elmqvist, and Benjamin B Bederson. 2016. Cognitive Stages in Visual Data Exploration. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization*. ACM, 86–95.
27. Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring cognitive load using eye tracking technology in visual computing. In *BELIV'16: Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. 78–85.