User Modeling and User-Adapted Interaction (2020) 30:199–249
https://doi.org/10.1007/s11257-019-09247-2

Effects of personal characteristics in control-oriented user interfaces for music recommender systems

Yucheng Jin[1] · Nava Tintarev[2] · Nyi Nyi Htun[1] · Katrien Verbert[1]

Received: 15 August 2018 / Accepted in revised form: 18 September 2019 / Published online: 25 October 2019

Abstract

Music recommender systems typically offer a "one-size-fits-all" approach with the same user controls and visualizations for all users. However, the effectiveness of interactive interfaces for music recommender systems is likely to be affected by individual differences. In this paper, we first conduct a comprehensive literature review of interactive interfaces in recommender systems to motivate the need for personalized interaction with music recommender systems, and two personal characteristics, *visual memory* and *musical sophistication*. More specifically, we studied the influence of these characteristics on the design of (a) *visualizations* for enhancing recommendation diversity and (b) the optimal level of *user controls* while minimizing cognitive load. The results of three experiments show a benefit for personalizing both visualization and control elements to musical sophistication. We found that (1) *musical sophistication* influenced the acceptance of recommendations for *user controls*. (2) *musical sophistication* also influenced recommendation acceptance, and perceived diversity for *visualizations* and the UI combining user controls and visualizations. However, musical sophistication only strengthens the impact of UI on perceived diversity (moderation effect) when studying the combined effect of controls and visualizations. These results allow us to extend the model for personalization in music recommender systems by providing guidelines for interactive visualization design for music recommender systems, with regard to both visualizations and user control.

Keywords User control · Personal characteristics · Recommender systems · Perceived diversity · Acceptance · Cognitive load · User experience

✉ Yucheng Jin
yucheng.jin@cs.kuleuven.be

1 KU Leuven, Leuven, Belgium

2 TU Delft, Delft, The Netherlands

1 Introduction

Music recommender systems suggest items that might be suitable for individual users using a range of different recommendation techniques. With the proliferation of music streaming platforms such as Spotify and Apple Music, users can easily access a large number of songs. More importantly, these platforms can provide users with personalized music recommendations based on their listening history, taste profile, etc. These platforms have influenced the way users search and explore music. For instance, the streaming platform Spotify currently has 180 million active users and provides a collection of more than 35 million songs (July 2018). As a result, it may be more meaningful to recommend songs that fit user's temporal preferences and context rather than showing songs based on user search requests (Lee et al. 2016).

To date, the user interface (UI) elements of most commercialized music recommender systems provide limited ability to control recommendation results, only allowing users to indicate whether they like or dislike a song. This limited ability to control additionally may lead to users perceiving the recommender system as a "black box" and lead to trust issues when recommendations fail (Herlocker et al. 2000). Previous research has shown many benefits for supporting controllability and transparency in several application domains such as music recommendations (Bostandjiev et al. 2012), career recommendations (Bostandjiev et al. 2013), and academic talk recommendations (Verbert et al. 2013). Having more control can increase users' perceived quality of recommendations (O'Donovan et al. 2008). In addition, users tend to be more satisfied when they have control over how recommender systems make suggestions (Konstan and Riedl 2012).

On the other hand, although controls empower users to influence the recommendation process to a greater extent, a high level of control may increase their cognitive load (Jin et al. 2017; Andjelkovic et al. 2016). The preference for interaction methods in recommender systems also depends on several personal characteristics such as domain knowledge, trust propensity, and persistence (Knijnenburg et al. 2011). In the music recommender domain, personal characteristics such as familiarity and visual memory have been shown to influence users' music choice and interaction with visual elements (Kamehkhosh and Jannach 2017; Millecamp et al. 2018, 2019). Besides, previous studies have shown positive effects of visualization on perceived diversity (Tsai and Brusilovsky 2017) and controllability on users' trust and recommendation acceptance (de Vries 2004; Fitzsimons and Lehmann 2004). However, the effects of different personal characteristics on perceived diversity, recommendation acceptance and cognitive load have not been investigated in the music recommender domain.

Based on the interactive recommendation framework proposed by Chen et al. (2016), we previously devised different levels of user control (low, middle, and high) associated with various components of a recommender system (Jin et al. 2017). In this paper, we propose a comprehensive version of the framework which exhibits the most common interaction and visualization elements, and their association with the three levels of control. This framework is presented in Fig. 1, which shows three main components of an interactive recommender system (Chen et al. 2016): *recommendation data*, *user profiles*, and *algorithm parameters*. A detailed explanation of the

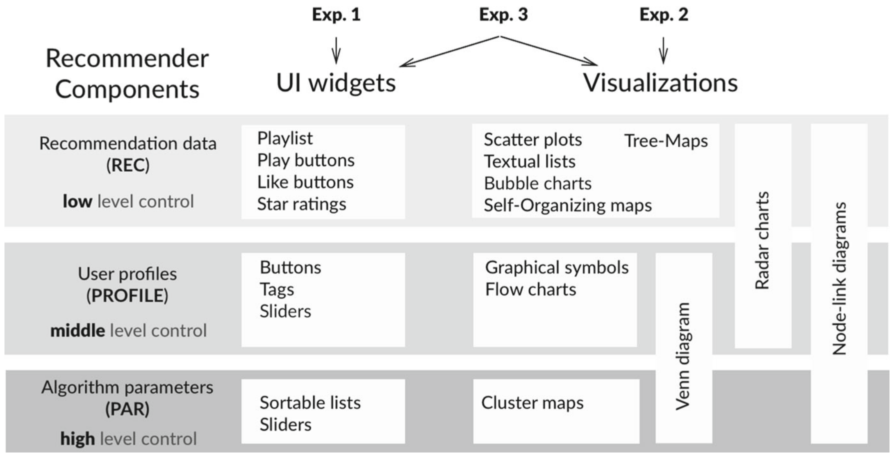

Fig. 1 User control-oriented UI framework for music recommender systems

framework is provided in Sect. 3.1. Depending on the level of control, a number of UI widgets are available that can be used to mediate between users and the recommender components. These UI widgets can be seen in Fig. 1 and are associated with three different recommender components and levels of control: control over recommendations (low level of control), control over the user profile (middle level of control), and control over the recommendation algorithm (high level of control). In addition to the levels of control, Fig. 1 shows the most commonly used visualizations in recommender systems.

A number of our previous works have focused on a variety of gaps in the music recommender domain (Jin et al. 2017; Millecamp et al. 2018). However, a better understanding of the effects of personal characteristics in association with the three levels of user control on music recommender systems has yet to be realized. Another strand of research has also focused on the effects of personal characteristics on the perception of visualizations (Conati et al. 2014, 2015; Tintarev and Masthoff 2016; Tintarev 2017), but to the best of our knowledge this has yet to be investigated in the music recommender domain. To address these gaps, we conducted three different experiments (*Experiment 1*, *Experiment 2*, and *Experiment 3*), investigating various aspects of the proposed framework.

While versions of Experiment 1 (Jin et al. 2018a) and Experiment 2 (Jin et al. 2018b) have both been previously published, this paper replicates Experiment 2 with a larger sample and also introduces the results of Experiment 3. Furthermore, this paper presents a deeper analysis of the experiments. These additional contributions help us to gain a comprehensive understanding of the effects of personal characteristic on music recommender systems under the proposed UI framework.

Consequently, the contributions of this paper are threefold:

– We review the effect of personal characteristics on the effectiveness of both visualizations and user control in music recommender systems (Sect. 2).

- Building on previous empirical work (Chen et al. 2016; Jin et al. 2017), we introduce our UI framework for personalized interface design in music recommender systems, which considers both control and visualizations (Fig. 1).
- We describe findings of three rigorous experiments conducted using the framework to evaluate the effectiveness of various user controls and visualizations in music recommender systems (Sects. 4, 5, and 6).

These novel contributions allow us to address the following research questions.

RQ1: How do *personal characteristics* influence user perception of recommendations (diversity, acceptance, and cognitive load)?

RQ2: How do personal characteristics *moderate* the effect of the user interface (user controls / visualizations) on user perception of recommendations (diversity, acceptance, and cognitive load)?

RQ3: How does the *complexity* of the user interface (user controls / visualizations) influence user perception of recommendations (diversity, acceptance, and cognitive load)?

Overall, we saw that combining multiple levels of user control tends to increase recommendation acceptance but does not lead to higher cognitive load (Experiment 1). While a more sophisticated visualization seems to have little impact on user perception. Moreover, combining full user control and visualization tends to increase the perceived diversity (Experiment 3).

When studying the main effects of personal characteristics on user perception, we found that musical sophistication positively affects recommendation acceptance through several mediators such as perceived quality (Experiments 1–3) and positively influences perceived diversity (Experiment 2 and 3).

For the moderating effect of personal characteristics, we only found that music sophistication positively moderates the impact of the user interface on diversity (Experiment 3)

The rest of the paper is organized as follows: In Sect. 2, we discuss previous work related to personal characteristics, user control, and visualization techniques in recommender systems. In Sect. 3, an overview and methodology of the three experiments, including experimental procedure, materials and evaluation metrics, are described. Sections 4, 5, and 6 present details and results of the three experiments (*Experiment 1*, *Experiment 2*, and *Experiment 3*, respectively).

Finally, we conclude the paper in Sect. 7 with a discussion of results and the limitations of our approach. We also highlight the implications of our findings for personalized music recommender systems.

2 Background

In this section, we review related work on personal characteristics, user control, and visualizations in recommender systems. In Sect. 2.1, we give an overview of previous literature on how *personal characteristics* interact with the performance of users, which include level of experience, trust, demographics, personality traits, and cognitive skills. In Sect. 2.2, we discuss previous work that looked at *user control* of recom-

Table 1 Overview of the personal characteristics discussed and their related example measures

Personal characteristics	Example measures
Level of experience	Musical sophistication (Müllensiefen et al. 2014; Ferwerda and Graus 2018; Millecamp et al. 2018, 2019)
Personality traits	The Big-Five (Chen et al. 2015; Ferwerda et al. 2017b), locus of control (Millecamp et al. 2019)
Demographic characteristics	Age, gender (Ferwerda et al. 2017a; Millecamp et al. 2018)
Cognitive skill	Visual working memory (Lallé et al. 2017; Tintarev and Masthoff 2016; Millecamp et al. 2018, 2019)

mender systems at three different levels: controlling recommendation results, user profiles, and algorithm parameters. In Sect. 2.3, we present *visualization techniques* that have been used in the recommender systems domain to support transparency and user interaction with such systems.

2.1 Personal characteristics

The influence of personal characteristics on the performance of users in interactive systems has been researched in depth. These works have investigated a variety of personal characteristics, which we describe below using the classification of Aykin and Aykin (1991): level of experience, personality traits, demographic characteristics and cognitive ability. An overview of the personal characteristics discussed in this section and their example measures are highlighted in Table 1.

2.1.1 Level of experience

Level of experience is one of the most commonly studied characteristics in the literature (Toker et al. 2012; Carenini et al. 2014; Conati et al. 2014; Domik and Gutkauf 1994; Inoue et al. 2011; Al-Maskari and Sanderson 2011; Zhang and Chignell 2001; Aykin and Aykin 1991). It is represented on a continuous scale with novice at one end, and expert at the other end of the scale. The level of experience may be expressed in different ways based on the area of research. For example, when investigating interactive user interfaces, users' experience may be seen as their level of familiarity with computers (Zhang and Chignell 2001) or with visualizations (Carenini et al. 2014; Conati et al. 2014).

Many studies have shown significant effects of the level of experience when interacting with recommender systems. For example, level of experience influences the users' choices of interaction methods (Knijnenburg et al. 2011). Novice users prefer simple and transparent interaction methods (Kramer 2007). These users typically lack the attribute knowledge (Alba and Hutchinson 1987; Chernev 2003), which may prohibit them to effectively use a personalized attribute-based recommender system that leverages such knowledge (Komiak and Benbasat 2006; Perera 2000; Randall et al. 2007). In the music recommender domain, Kamehkhosh and Jannach (2017) discovered that users' familiarity with a recommended song influences their choice, i.e., users tend to like a recommendation when they already know the song. The *Goldsmiths Musical Sophistication Index* (Gold-MSI)[1] is regarded as an effective way to measure domain expertise of users and has shown a strong correlation with individuals' music preference (Müllensiefen et al. 2014), listening behavior (Ferwerda and Graus 2018) and interaction with visual elements in recommender interfaces (Millecamp et al. 2018, 2019). Hence, in our studies, we used the Gold-MSI to measure musical sophistication (MS) of the participants.

2.1.2 Personality traits

In the Chambers concise dictionary,[2] personality is defined as "a person's nature or disposition; the qualities that give one's character individuality" and it is a key area of research in user modeling and user adaptive systems. Personality traits can affect the performance and preference of a user (Aykin and Aykin 1991). In the recommender systems domain, a number of previous works have studied the correlation between personality traits and user preference (Perik et al. 2004; Hu and Pu 2011; Tkalcic et al. 2009, 2011). Similarly, in more related studies in the domain of music recommendation, the Big-Five personality traits have been shown as an influencing factor for users' recommendation preference (Chen et al. 2015; Ferwerda et al. 2015, 2017b). On the other hand, it has been found that other personality traits such as locus of control and need for cognition do not affect users' interaction with the system, trust or perceived recommendation quality (Millecamp et al. 2019). Nevertheless, to clearly differentiate from the extensive research on personality-based recommendations, personality traits were deemed out of the scope of this paper and thus were not measured in the studies presented herein.

2.1.3 Demographic characteristics

Demographic characteristics have also been researched extensively in the literature on adaptive interactive systems (Brusilovsky and Millán 2007; Gauch et al. 2007; Champiri et al. 2015; Lekkas et al. 2011; Domik and Gutkauf 1994; Aykin and Aykin 1991). While some research has focused only on basic demographics such as age, sex and gender (Lekkas et al. 2011; Domik and Gutkauf 1994; Aykin and Aykin 1991), others went deeper and investigated characteristics such as personal interests, goals, background, country, education level, marriage status, job sector, income and first language (Brusilovsky and Millán 2007; Gauch et al. 2007; Champiri et al. 2015; Zhang and Chignell 2001). In a similar study in the domain of music recommendation, Ferwerda et al. (2017a) found varying musical preference of different age-groups. However, the effect of age and gender on users' interaction with recommendations, acceptance and diversity has not been found in previous research (Millecamp et al. 2018). For

[1] https://www.gold.ac.uk/music-mind-brain/gold-msi/, accessed June 2018.

[2] http://www.chambers.co.uk.

this reason, we did not further study the effect of demographic characteristics in this paper.

2.1.4 Cognitive skills

Cognitive skills have been investigated by numerous previous works (Brusilovsky and Millán 2007; Toker et al. 2012; Carenini et al. 2014; Conati et al. 2014; Domik and Gutkauf 1994; Al-Maskari and Sanderson 2011). In particular, *working memory* is a commonly measured cognitive skill. It can be categorized into visual and verbal working memory. Previous work has *repeatedly* found *visual working memory (VM)* to be a factor that affects cognitive load when interacting with adaptive interactive systems (Lallé et al. 2017; Conati et al. 2014; Tintarev and Masthoff 2016). Besides, visual working memory and visual literacy have been found to affect users' interaction with visual elements in recommender interfaces (Millecamp et al. 2018, 2019). As the main focus of our studies involved comparing interactive UI elements and visualization techniques, we also measured *visual working memory* of participants.

2.2 User control in recommender systems

In the previous section, we reviewed personal characteristics that may influence interactions with music recommendations. In this section, we categorize and define the type of interactions we are considering. First, we review why these interaction types are considered as beneficial for recommender systems.

Many recommender systems are inscrutable to users, and users often need more control in order to increase the perceived quality of recommendations (O'Donovan et al. 2008). In addition, users tend to be more satisfied when they have control over how recommender systems produce suggestions for them (Konstan and Riedl 2012). Controllability often allows users to steer the recommendation process to obtain suggestions that are better suited to them (He et al. 2016). This, in turn, promotes trust in the system (de Vries 2004; Fitzsimons and Lehmann 2004), hence leading to increased acceptance of recommendations.

Controlling the recommendation process may range from providing ratings for an item to adjusting algorithm parameters and may take place at any stage of the system's life cycle (Chen et al. 2016). The UI elements of existing recommender systems allow users to interact with three distinct components of the systems (see Fig. 1); these are: *recommendation results* (i.e., the output of a recommender system), *user profile* (e.g., the user's most listened songs), and *algorithm parameters* (i.e., the weights in an algorithm). As highlighted in Sect. 1, each component has also been associated with either low, middle, or high control levels that are provided to users (Jin et al. 2017). These levels are further explained in Sect. 3.1. In the following subsections, a detailed description for each of the components is provided.

2.2.1 Controlling recommendation results

This particular type of interaction involves users evaluating the output of a recommender system and steering the recommendations toward their desired outcomes similar to critiquing (Chen and Pu 2012; McCarthy et al. 2010). The systems that support control of recommendations initially produce one or more recommendations based on user preferences. From this initial list of recommendations, users then select an item that represent their desired outcomes. The system then updates the settings and provides users with another set of recommendations. This cycle may be iterated a number of times until users find their desired outcomes. An advantage of controlling recommendations directly is that the user does not need to specify the exact features of the desired outcomes; thus, the demand for product knowledge and domain expertise is lowered (McCarthy et al. 2010). According to an experiment conducted by Pommeranz et al. (2012), familiarity with an item is an important factor mediating the trade-off between providing detailed preference feedback and required effort. Finally, Gena et al. (2011) also found that controlling recommendations can reduce the task time and error rate of users while increasing decision accuracy.

In the domain of music recommendation, Saito and Itoh (2011) implemented MusiCube in which users first evaluate an initial set of music with binary feedback. The system then produces a new set of recommendations and categorizes them into 11 musical features (i.e., root-mean-square energy, low energy, tempo, zero crossing, roll off, brightness, roughness, spectral irregularity, inharmonicity, and mode). Among these features, the user can select any two and the system visualizes the respective songs in a 2D graph. Although MusiCube has an advantage of control over recommendations, we believe that the majority of musical features provided may be unfamiliar for most users.

2.2.2 Controlling user profiles

Certain recommender systems allow users to view and modify their profile and personalization assets (i.e., data matrix used to provide recommendations) to their requirements and preferences. Bakalov et al. (2013) implemented a recommender system for biochemical literature which allows users (mainly biologist, biochemists and genomicists) to access and modify their user models. Their evaluation results showed that in addition to improved quality of recommendations, this approach also helps to solve the typical black box issue of recommender systems. Schaffer et al. (2015) implemented a movie recommender system using the MovieLens dataset in order to investigate the impact of profile manipulation (i.e., adding, deleting, re-rating). Their results showed that users were able to identify sources of bad recommendations and remove them. Overall, allowing users to access and modify their profile has been found to improve transparency and acceptance of recommendations (Jin et al. 2016).

2.2.3 Controlling algorithm parameters

Recommender systems may also allow tweaking the underlying algorithm such as adjusting the weight of an item, which is usually invisible to users. SetFusion (Parra and Brusilovsky 2015), for example, allows users to search academic papers by adjusting a number of attributes such as: "most bookmarked papers," "similar to favorite articles," and "frequently cited authors in ACM DL." Evaluation results showed that this approach, similar to profile manipulation, has a positive outcome in terms of trust and acceptance toward the recommendations (Parra and Brusilovsky 2015). LinkedVis (Bostandjiev et al. 2013) is a social recommender system that used the same approach. Based on LinkedIn, LinkedVis allows users to adjust the weights of their profile items (i.e., schools, degrees, skills, etc.), as well as the weights with their connection in order to get a suggestion for new connections. Like SetFusion, evaluation results of LinkedVis also showed higher acceptance and satisfaction (Bostandjiev et al. 2013).

In the context of music recommendation, TasteWeights (Bostandjiev et al. 2012) has a similar design as LinkedVis. Users are able to change the weights of their favorite artists, together with trending items from Wikipedia, Facebook, and Twitter. This approach allows users to find music not just from their favorite artists, but also from trending topics. This idea can be traced to the work of Schaffer et al. (2015) on meta-recommendation systems, where users are provided with personalized control over the generation of recommendations by altering the importance of specific factors on a scale from 1 to 5. Although TasteWeights provides an innovative approach in the music recommendation domain, further studies are required to explore other interaction techniques and their effects on users' cognitive load, acceptance, satisfaction, and trust (He et al. 2016).

2.3 Visualizations for recommender systems

The UI element of most recommender systems contains a type of visualization to enable users to view and interact with system components. While some systems use visualizations just to present the results to users, other systems also allow users to directly interact with the recommender components described in Sect. 2.2.

We highlighted various kinds of visualization that are commonly used in recommender systems in Fig. 1. In the following subsections, we explain these visualizations and how they are related to each of the three recommender components.

2.3.1 Visualizations of recommendation results

Recommendation results are typically presented using scatter plots (Saito and Itoh 2011), textual lists (Torrens et al. 2004), bubble charts (Jin et al. 2018a), self-organizing maps (Knees et al. 2007; Pampalk et al. 2002), TreeMaps (Torrens et al. 2004), and radar charts (Hilliges et al. 2006). Figure 2 shows an example for each of these visualizations.

The self-organizing map (Fig. 2a) shows an arrangement of a music collection on a 2D map. Music from the collection is represented in the "islands of music," which fluctuates according to the music's rhythmic model. The radar chart (Fig. 2b) places similar songs close together. The radar is subdivided into a number of sections based on musical attributes such as: melodic, slow, rough, turbulent, rhythmic, fast, clean, and calm. The scatter plot (Fig. 2c) displays a set of colored icons corresponding to different music. The color defines four states of a song: “positively listened,” “negatively listened,” “being suggested,” and “not suggested yet.” As with all scatter plots, on the X- and Y-axes, users can choose to view any two of the four states at a time. The textual list (Fig. 2d) is the most basic way of visualizing music libraries. In essentially a table-like grid, each row shows a song with its associated attributes such as genre, artist, composer, album, and year. The tree-map (Fig. 2e) shows music genres in various sizes of rectangle that are proportional to the number of songs in the given genre. The bubble chart (Fig. 2f) is one of the simplest visualizations that can show clusters of music by their genre and reflects popularity using the size of the bubbles. Thus, it is able to visualize multiple dimensions, such as popularity, similarity and genre, at a time. Due to these advantages and simplicity, we used various versions of bubble charts in our studies.

2.3.2 Visualizations of user profiles

User profiles are presented using Venn diagrams (Andjelkovic et al. 2016), graphical symbols (Bogdanov et al. 2013), and radar charts (Millecamp et al. 2018). The overall objective of these visualizations is to give the user insight into the user model that is used by the recommender system. Figure 3 shows an example for each of these visualizations.

The radar chart (Fig. 3a) allows users to constantly modify their musical taste using attributes such as acousticness, energy, valence, danceability, and instrumentalness. These modifications are immediately reflected in the recommendations. The Venn diagram (Fig. 3b) shows a user’s songs clustered into three sets based on their mood tags: vital, uneasy, and sublime. The clusters are differentiated using separate colors. The graphical symbols (Fig. 3c) are used to build an avatar of a user which represents the user’s musical taste preference. Each graphic element box represents an attribute of the user (e.g., body, head, eye, etc.), and the values inside the boxes represent all possible descriptor values associated with the presented element.

2.3.3 Visualizations of algorithm parameters

Algorithm parameters are represented using Venn diagrams (Parra and Brusilovsky 2015) and cluster maps (Verbert et al. 2013). Figure 4 shows an example for each of these visualizations. The Venn diagram (Fig. 4a) shows color-coded ellipses representing recommendation methods, and small circles within the ellipses represent recommended items. In this example, a comparison of three different recommendation methods can be seen, represented by the three ellipses. Items located on the intersections are recommended by more than one method. The cluster map (Fig. 4b) shows different clusters of recommendation linked by connected components. The circles at the connection points represent input entities such as other users, recommender agents or tags. Yellow circles within each bubble represent the recommended items. This visualization allows users to explore relationships between items that are associated with different entities (i.e., recommended by an agent, bookmarked by a user, tagged with a tag).

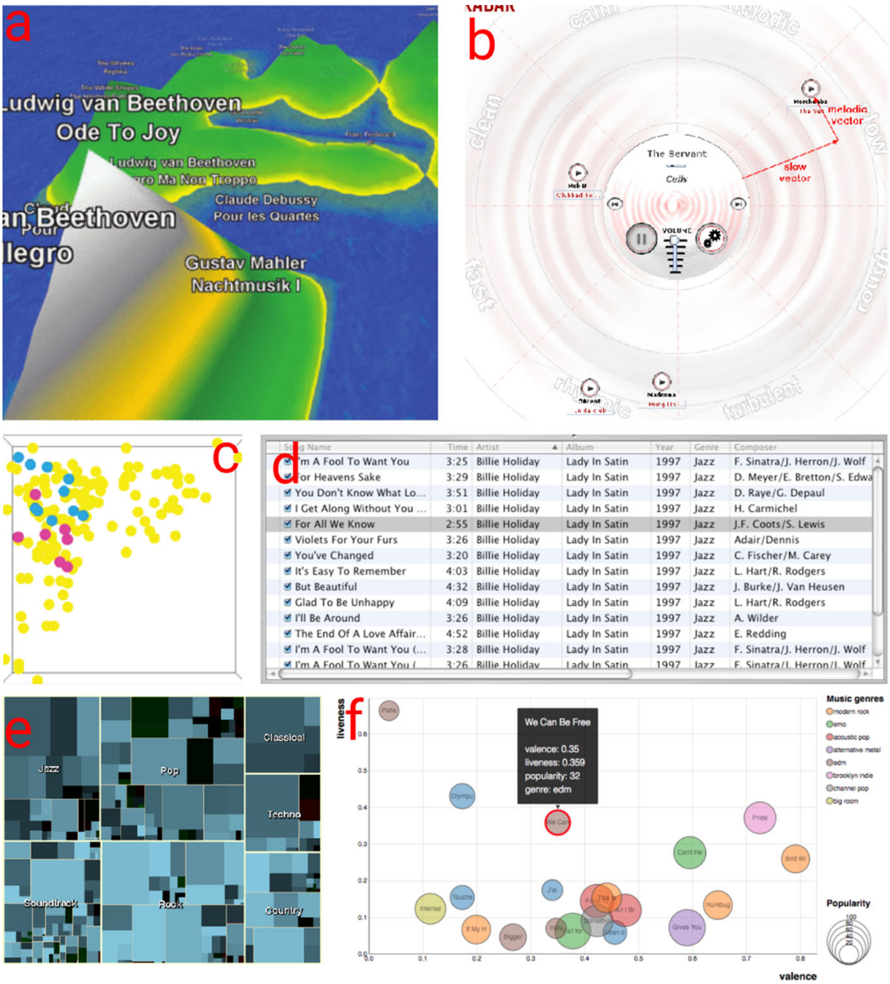

Fig. 2 Visualizations used to present music recommendation results. **a** Self-organizing maps (Knees et al. 2007), **b** radar charts (Hilliges et al. 2006), **c** scatter plots (Saito and Itoh 2011), **d** textual lists (Torrens et al. 2004), **e** TreeMaps (Torrens et al. 2004), **f** bubble charts (Jin et al. 2018a). (Color figure online)

Unlike these visualizations, node-link diagrams (Bostandjiev et al. 2012, 2013; O'Donovan et al. 2008) have been used to represent all three recommender components. Figure 5 shows an example for this type of visualization. As seen in the figure, the user is presented with (1) recommendation results, (2) their profile (i.e., favorite artists), and (3) algorithm parameters (i.e., weighting of top trending items from Wikipedia, Facebook, and Twitter).

Usually, a visualization allows users to not just inspect but also manipulate a particular recommender component, which may influence different aspects of recommendations. For instance, by visualizing the user profiles and recommendation process, *transparency and user control* of the system can be improved significantly.

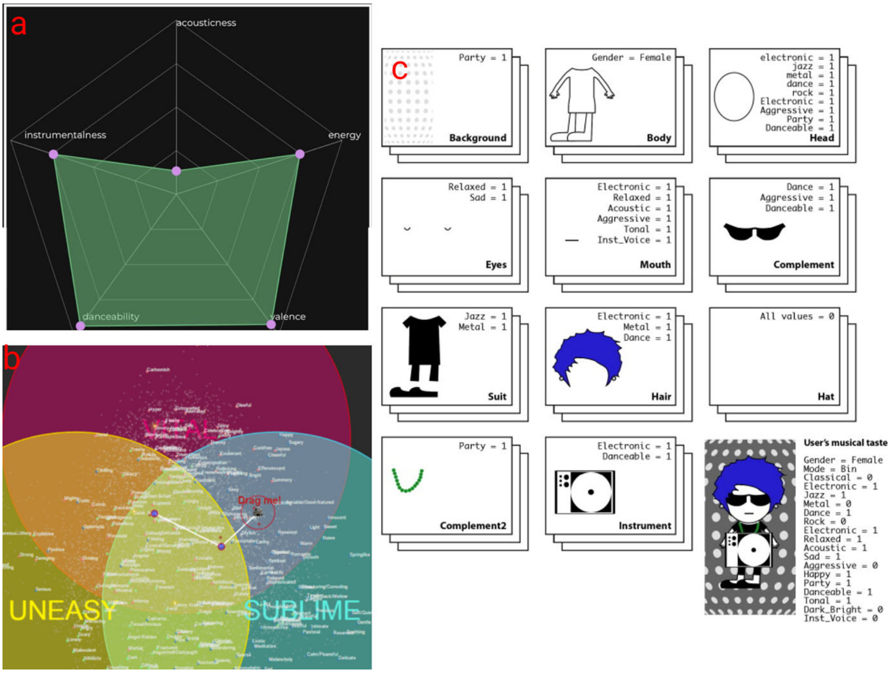

Fig. 3 Visualizations used to present user profiles. **a** Radar charts (Millecamp et al. 2018), **b** Venn diagrams (Andjelkovic et al. 2016), **c** graphical symbols (Bogdanov et al. 2013)

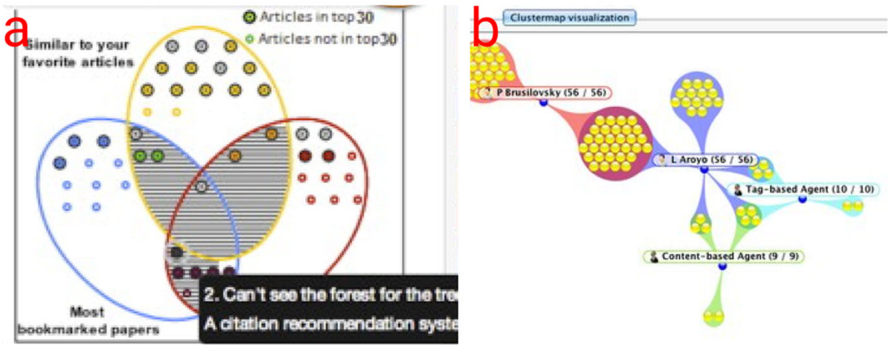

Fig. 4 Visualizations used to present algorithm parameters: **a** Venn diagrams (Parra and Brusilovsky 2015) and **b** cluster maps (Verbert et al. 2013). (Color figure online)

Jin et al. (2016), for instance, demonstrate an interactive flowchart-based visualization that explains how a selected ad is filtered for the targeted user profile. MoodPlay (Andjelkovic et al. 2016) is an emotion-based music recommender system. It allows users to explore music by modifying affective data and inspect the explanation of recommendations, which increases *acceptance and understanding* of recommendations. Verbert et al. (2013) present a system that increases the *effectiveness of making a choice* by explaining the provenance of recommendations and offering control to users. Some

Fig. 5 Visualization used to present data from all three recommender components: node-link diagram (Bostandjiev et al. 2012)

systems show increased *accuracy* by enabling users to inspect the recommendation process (Bostandjiev et al. 2012; O'Donovan et al. 2008).

In addition, several studies have shown the positive effects of visualization on *perceived diversity*. Hu and Pu (2011) proposed an organization-based interface to increase users' perceived diversity of recommendations. Wong et al. (2011) presented a system named Diversity Donut that allows users to indicate the level of diversity for the recommended items. Tsai and Brusilovsky (2017) also presented a diversity-enhanced interface that presents recommendations with multiple attributes in a two-dimensional scatter plot inspiring our approach.

2.4 Evaluation metrics

Performance of a recommender system can be measured in a number of ways. Some of the most frequently used metrics have been highlighted by Schedl et al. (2018) and categorized into accuracy-related metrics (e.g., mean absolute error, precision, recall, etc.) and beyond-accuracy metrics (e.g., perceived diversity, acceptance, serendipity, etc.). In this section, we focus on the beyond-accuracy metrics that we believe will influence primarily acceptance of recommendations and perceived recommendation diversity. We also discuss the importance of measuring cognitive load.

Diversity As mentioned above, several studies have shown the positive effects of visualization on *perceived diversity*. In addition, one paper considered individual user tendencies, with regard to diversity of content consumed, when applying algorithmic re-ranking (Jugovac et al. 2017). Previous work has also found a relationship between personality and individuals' attitudes toward new or diverse recommendations (Tintarev et al. 2013). Jointly these results suggest a variance in users' needs for diversity, and a gap in understanding what this means in terms of the requirements on visualizations and control in music. Therefore, in order to study the effect of individual differences for music, we used diversity as one of the evaluation metrics in our studies. There are two commonly used user-centric evaluation frameworks for recommender systems, by Knijnenburg et al. (2012) and Pu et al. (2011)—this paper applies the framework of Knijnenburg et al. (2012) which differentiates between different user-centered aspects of the recommender systems such as objective system aspects, user experience, subjective system aspects, interaction, and personal characteristics.

Acceptance Allowing control over the system and users' acceptance of the system or advice are correlated (de Vries 2004). Distrusting the system and being unable to control it typically lead to lower satisfaction or even cause reactions where users actively counter the system's advice (Fitzsimons and Lehmann 2004). Users may therefore prefer systems that are easy to understand and ones they can control. The various user control elements investigated in our studies are discussed in Sect. 2.2. The user-centric frameworks proposed by Knijnenburg et al. (2012) and Pu et al. (2011) contain measurements for users' trust, but do not contain specific measurements for acceptance. We are primarily interested in the situations where trust leads to recommendation acceptance. Thus, we measured percentage of acceptance for recommendations and investigated factors of the Knijnenburg et al.'s framework that interacts with acceptance.

Cognitive load Research has shown correlations between cognitive load and user satisfaction (Bradford 2011), as well as choice accuracy of users (Aljukhadar et al. 2012). Cognitive load of a user is usually determined by how many cognitive resources are taken up by activities that facilitate problem-solving (Chandler and Sweller 1991; Sweller 1988). Cognitive load can be measured by means of self-assessment questionnaires or by analyzing physiological data collected during a task. The NASA task-load index (NASA-TLX)[3] is one of the most widely used questionnaires to measure cognitive load. It comprises six dimensions: mental demand, physical demand, temporal demand, perception of self-performance, effort and frustration. We used the NASA-TLX due to its simplicity and versatility for on-the-web experiments.

In our studies, following a well-validated conceptual model (Fig. 6) introduced in a user-centric evaluation framework (Knijnenburg et al. 2012) we analyze how the above three factors interact with other user experience factors. In Fig. 6, the directional path from objective system aspects (OSA) to subjective system aspects (SSA) and to experience (EXP) implies the potential effect of user control and visualization on perceived diversity and cognitive load. The relation between personal characteristics (PC) and interaction (INT) suggests the potential effect of musical sophistication and visual memory on acceptance and cognitive load.

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

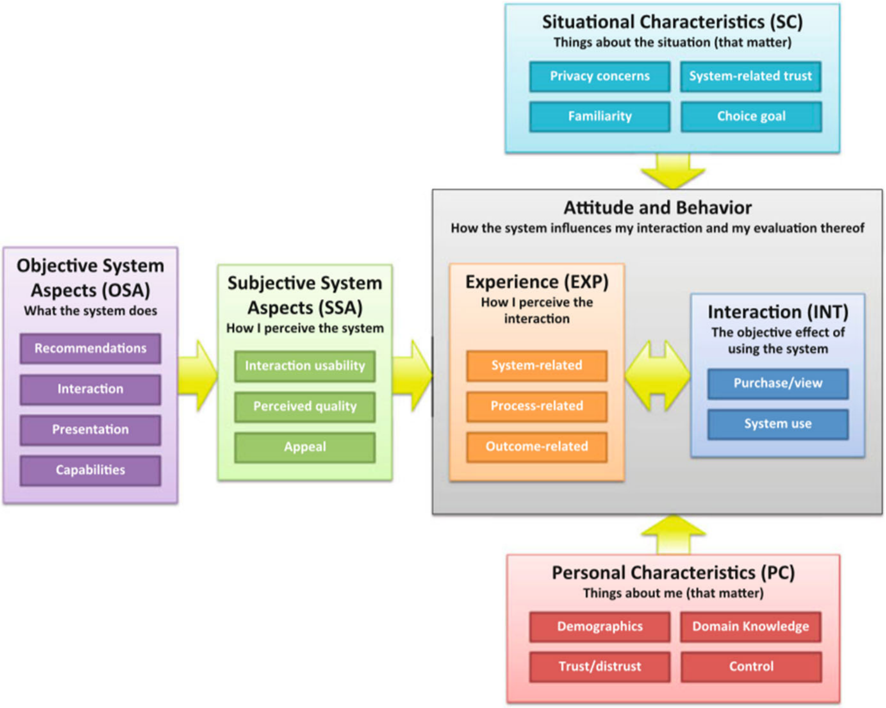

Fig. 6 The conceptual model in a framework for the user-centric evaluation of recommender systems (Knijnenburg et al. 2012)

2.5 Research gap

Previous studies have shown the potential of using visualization and interaction techniques to improve acceptance and perceived diversity of recommendations (Hu and Pu 2011; Wong et al. 2011; Tsai and Brusilovsky 2017; de Vries 2004; Fitzsimons and Lehmann 2004). In addition, a number of studies in the music recommender domain have discovered that certain personal characteristics, familiarity and visual memory, can influence users' music choice and interaction with visual elements (Kamehkhosh and Jannach 2017; Millecamp et al. 2018, 2019). To the best of our knowledge, little is known about the influence of musical sophistication (i.e., experience/familiarity) and visual memory, on the effect of visualizations and control elements on perceived diversity, acceptance and cognitive load.

Reviewing previous work, we found that users have different preferences and methods of interaction with a recommender system depending on their differing characteristics, but that this work does not systematically compare the effectiveness of visualization and control elements for different types of users (Knijnenburg et al. 2011). Similarly, while some of the previous work e.g., (Conati et al. 2014, 2015; Tintarev and Masthoff 2016; Tintarev 2017) has found an effect of personal characteristics on the perception of visualizations, their findings may not be directly applicable to the music recommender domain.

Several of the reviewed studies have shown positive effects of visualization on perceived diversity. Therefore, we measured diversity as one of the outcomes in our studies, to assess the extent to which individuals' differences in personal characteristics can impact their perceptions of diversity in recommendation results. Additionally, since controllability is often associated with users' trust in the system and recommendation acceptance (de Vries 2004; Fitzsimons and Lehmann 2004), we chose to study how UI elements impacted the final interactions of users with systems. Consequently, we also measured how individuals' differences in personal characteristics can impact their acceptance of recommendations.

We also identified *musical sophistication* and *visual working memory* as personal characteristics that may influence the effectiveness of visualizations and control. The musical sophistication index was chosen since it has shown a strong correlation with individuals' music preferences, e.g., Müllensiefen et al. (2014) and interaction with visual elements in recommender interface (Millecamp et al. 2018, 2019). *Visual working memory* has been previously found to be a factor that affects cognitive load in adaptation of interactive systems (Conati et al. 2014; Lallé et al. 2017; Mayer and Moreno 2003; Tintarev and Masthoff 2016), and our studies involved comparing interactive UI elements and visualization techniques.

To summarize, in this paper we aim to address a number of research gaps in the music recommender domain by assessing the influence of personal characteristics on user perception of recommendations. Furthermore, we measure how do personal characteristics moderate the effect of user controls, visualizations, and their combination on acceptance, perceived diversity and cognitive load. More specifically, we research the influence of user controls, visualizations, as well as their combination.

3 Overview and methodology of experiments

In this section, we first explain in detail the framework (Fig. 1) that is introduced in Sect. 1 and how each of the three experiments is related to this framework. We then describe the methodology that is shared between the three experiments.

3.1 User control-oriented UI framework for music recommender systems

The first component of the framework (see Fig. 1), *recommendation data*, involves interactions with recommendation results, such as sorting and rating. This type of interaction typically grants users with a low level of control over the recommendation results as they can only indicate whether they like or dislike a particular result.

The second component involves interactions with *user profiles*. These middle-level interactions allow users to view and adjust their profile and personalization matrix that the system uses to calculate recommendations, in order to closely match their constantly developing preferences.

Interactions with the third component, *algorithm parameters*, grant users with control over the algorithm. This is a high level of control which allows users to manipulate parameters such as item weight and item attributes that are usually invisible to them.

Table 2 Independent variables, dependent variables, experiment design in three experiments (Experiments 1–3), and the factors that influence UI complexity

Experiment	Independent Var.	Dependent Var.	Factors of UI complexity
Experiment 1: User controls	Settings of user control realized by UI widgets	Diversity, acceptance, cognitive load	User control levels
Experiment 2: Visualizations	Types of bubble charts	Diversity, acceptance, cognitive load	Dimensions of visualized data
Experiment 3: Controls + Vis.	Types of user interfaces	Diversity, acceptance, cognitive load	UI components

In addition to the level of control, the framework considers what kind of visualizations accompany the recommendations. The majority of visualizations used to present *recommendation data* include scatter plots, textual lists, bubble charts, self-organizing maps, tree-maps, and radar charts. Some visualizations used to present *user profile* data include Venn diagrams, graphical symbols, and radar charts. Unlike these visualizations, node-link diagrams have been used to present data from all three recommender components: *recommendation data*, *user profiles*, and *algorithm parameters* (Bostandjiev et al. 2012, 2013).

As mentioned earlier, three experiments were conducted using this framework in order to address a number of gaps in the music recommender domain. Table 2 summarizes the experiments and highlights the UI components and control levels implemented for each experiment.

In *Experiment 1* (see Sect. 4 for details), we investigated how personal characteristics (i.e., visual memory and musical sophistication) influence user perception and moderate the effect of the three levels of control on diversity, acceptance, and cognitive load.

In *Experiment 2* (see Sect. 5 for details), we investigated how personal characteristics influence user perception and moderate the effect of two types of visualizations on diversity, acceptance, and cognitive load.

In *Experiment 3* (see Sect. 6 for details), we investigate how personal characteristics influence user perception and moderate the effect of two combinations of control and visualizations on diversity, acceptance, and cognitive load.

Together, these experiments allowed us to better understand how personal characteristics influence the interaction with music recommender systems under the proposed framework. In the next section, we provide a detailed description of the methodology used for the experiments.

3.2 Methodology

This section describes the methodological aspects that are shared across the three user-centered experiments.

3.2.1 Study procedure

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 playlist. 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 recommender could leverage the Spotify API and user listening history to generate "real" recommendations.
2. *Pre-study questionnaire*—This questionnaire collects user demographics and measures user's personal characteristics such as musical sophistication and visual memory capacity.
3. *Manipulating the recommender and rating songs*— Participants were asked to interact with the recommender and to rate the generated songs. To ensure that participants spent enough time to explore recommendations, the questionnaire link was only activated after 10 minutes. After tweaking the recommender, 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 of their opinions and suggestions about our recommender.

3.2.2 Experimental platform

As an experimental platform, we chose Spotify because it is one of the largest online music providers and offers a free API.[4] The Spotify API allows to generate recommendations based on up to five favorite artists. In addition, the API also allows modification of 14 musical attributes[5] in order to describe musical preference.

As in the Spotify application, we presented each recommended song by its title and artist. Album art and album name were not displayed in order to have a clean and manageable layout. The Spotify API provides a way to play a preview of up to 30 seconds for each recommended song (complete songs are inaccessible). We attached this feature with a play button in our interfaces which allowed users to listen to a preview of the recommended songs. Similar to the Spotify radio feature, we used "Thumb up" and "Thumb down" buttons to allow users to like or dislike the recommended songs.

4 https://developer.spotify.com/web-api/get-recommendations, accessed June 2018.

5 https://developer.spotify.com/web-api/get-recommendations/#tablepress-220., accessed June 2018.

3.2.3 Recommender algorithm

The recommendation algorithm was implemented by leveraging the Spotify Web API. First, we get the **seeds**, e.g., the top artists of a user, by calculating the user's expected preference to a particular artist according to his/her listening history.[6] Then, we take **seeds** as an input to call a recommendation service[7] (**RS**) that generates a playlist containing 20 songs matching similar artists and tracks. Each recommended song has a *popularity score*, *genres*, and *audio features*.

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.

3.2.4 Independent variables

In each experiment, we varied the interface where users interact with recommenders. While the materials between experiments varied, the same material was used within each experimental setting.

3.2.5 Dependent variables

Each experiment measured three key dependent variables:

- *Perceived diversity* This was a self-reported measure based on diversity-related questionnaire items in a user-centric evaluation framework for recommender systems (Knijnenburg et al. 2012) (see Table 3).
- *Recommendation acceptance* as described in the study procedure, all participants need to rate 20 recommended songs in the three experiments. Thus, the recommendation acceptance was measured by the percentage of liked songs in the playlist.
- *Cognitive load* we employed The NASA task-load index (NASA-TLX) to measure cognitive load from six dimensions: mental demand, physical demand, temporal demand, perception of self-performance, effort and frustration.

3.2.6 Covariates

In all three experiments, we evaluated user characteristics and perceived factors of the recommender system. In all three experiments, we measured two personal characteristics:

- **Musical sophistication (MS)** *measurement of the ability to engage with music in a flexible, effective and nuanced way* (Müllensiefen et al. 2014). We measure

[6] https://api.spotify.com/v1/me/top, retrieved July 2018.

[7] https://api.spotify.com/v1/recommendations, retrieved July 2018.

musical sophistication using the Goldsmiths Musical Sophistication Index (Gold-MSI).[8] We selected the ten most relevant questions items of the subscale *general factor* on a 7-point Likert scale in all three experiments.
- **Visual memory (VM)** *the ability to recall visual patterns* (Tintarev and Masthoff 2016).
  The visual memory capacity is measured by "Corsi block-tapping test."[9] In the test, a number of tiles are highlighted one at a time, and participants are asked to select the tiles in the correct order afterward. The number of highlighted tiles increases until the user makes too many errors. This test allows us to better distinguish participants by the level of visual memory capacity.

Additionally, we measured user-centered factors that might have an effect on recommendation acceptance and perceived diversity. In all three experiments, we use Knijnenburg et al.'s framework (Knijnenburg et al. 2012) to measure the factors presented in Table 3.

The questions were in the form of 7-point Likert scales, and the answers ranged from 1 (strongly disagree) to 7 (strongly agree). Following the above questions, a number of open-ended questions were also administered to capture feedback from the participants about the most and the least useful parts of each interface.

3.2.7 Interaction log

Since user perceptions may differ from behavior, we also recorded a log of the participants' interactions with different UI components. Specifically, the log captured:

- The number of times the algorithm parameters (weights) were modified (*parChange*).
- The number of times the seeds were added or removed (*proChange*).
- The number of times the recommendation items were removed and sorted (*recChange*).
- The number of times the dislike button was clicked (*disliked*).
- The number of times the like button was clicked (*liked*).
- The total number of times the items on the visualizations were clicked (*visClick*).
- The total number of times the items on the visualizations were hovered (*visHover*).

This log was then used to understand the impact of participants' personal characteristics on their interactions with the interfaces.

3.2.8 Hypotheses

We discuss the existing effects of user interface and personal characteristics on user perception from previous research (Sect. 2.5), which allows us to propose nine research hypotheses as below:

- **H1:** The more sophisticated UI (user control/visualizations) will increase recommendation acceptance.

8 http://www.gold.ac.uk/music-mind-brain/gold-msi/, accessed June 2018.

9 https://www.humanbenchmark.com/tests/memory, accessed June 2018.

Table 3 The questionnaire constructed based on the user-centric evaluation framework for recommender systems (Knijnenburg et al. 2012)

Perceived quality: participants' perceived quality of the recommended songs	
	Q1: The recommended songs fitted my preference
	Q2: Each of the recommended songs was well-chosen
	Q3: The provided recommended songs were interesting
Perceived accuracy: participants' perceived accuracy of the recommended songs according to their preference	
	Q4: The list of recommendations was appealing
	Q5: The list of recommendations matched my preferences
	Q6: I did not like any of the recommendations in the list
Perceived diversity: the similarity among the recommended songs	
	Q7: Several songs in the list of recommended songs were very different from each other
	Q8: The list of recommended songs covered many genres
	Q9: Most songs were from the same type
	Q10: No two songs in the list seemed alike
Satisfaction: participants' satisfaction about their chosen recommendations	
	Q11: I like the items I have chosen
	Q12: I enjoyed listening to my chosen items
	Q13: The chosen play-list fits my preferences
Choice difficulty: difficulty of choosing a recommended song	
	Q14: The task of making a decision was overwhelming
	Q15: Selecting the best songs was very easy
	Q16: Comparing the recommended songs was very easy
Effectiveness: usefulness of recommendations generated from systems	
	Q17: The music recommender has no real benefit for me
	Q18: I would recommend the music recommender to other
	Q19: I can save time using the music recommender

Q10 is a new added item for perceived diversity in Experiments 2 and 3, because Q7–9 did not show convergent validity to measure diversity in Experiment 1. Besides, we combine the items of perceived quality and perceived accuracy because of a large correlation between these two concepts

- **H2:** The more sophisticated UI (user control/visualizations) will lead to higher perceived diversity.
- **H3:** The more sophisticated UI (user control/visualizations) will lead to higher cognitive load.
- **H4:** Users with higher *Musical sophistication (MS)* are more likely to accept more recommended songs.
- **H5:** Users with higher *Musical sophistication (MS)* are more likely to perceive higher diversity.
- **H6:** Users with higher *Visual memory (VM)* are more likely to have less cognitive load.
- **H7:** *Higher Musical sophistication (MS)* tends to strengthen the effect of user interface on acceptance of recommendations.

Table 4 Demographics for participants in the three experiments

Experiment	N (rejected)	Age mean (SD)	Gender (female %)
Experiment 1: User controls	271 (31)	28.0 (7.1)	55.4
Experiment 2: Visualizations	120 (24)	31.2 (8.0)	41.7
Experiment 3: Controls + Vis.	180 (22)	30.3 (7.6)	41.1

- **H8:** *Higher Musical sophistication (MS)* tends to strengthen the effect of user interface on perceived diversity.
- **H9:** *Higher visual memory (VM)* tends to strengthen the effect of user interface on cognitive load.

3.2.9 Participants

The participants for all three experiments were recruited on Amazon Mechanical Turk (mTurk). The participants were compensated by 2 USD based on the expected completion time. Table 4 shows the number of participants and demographics for the three experiments. These are the participants whose data are valid for the analyses. The participants were required to have a minimum approval rating of 90%. We recorded the unique worker IDs of participants who completed the experiment and rejected the repeated participation. Moreover, to control the quality of study, we rejected the workers who gave contradicting answers to the questions of a measured aspect. For example, if a participant strongly agreed with Q8 and Q9, we will think they gave contradicting answers to the questions of perceived diversity. Note, to ensure the sample size we republished the rejected work on mTurk, and the number of rejected participants is shown in the bracket after the number of participants.

4 Experiment 1: effects of user control

As shown in our framework of music recommender interfaces (see Fig. 1), UI widgets are a crucial part which realize user control over three common recommender components, *recommendations, user profiles, and algorithm parameters*. We define three control levels (low, middle, and high) toward each recommender component. The initial experiment aims to systematically study the impact of different levels of control on recommendations and investigate the impact of personal characteristics on the effectiveness of user controls.

4.1 Setup

We used the Spotify API[10] to design a music recommender system and to present the user controls for three distinct recommender components. We leveraged the algorithm

[10] https://developer.spotify.com/web-api, accessed June 2018.

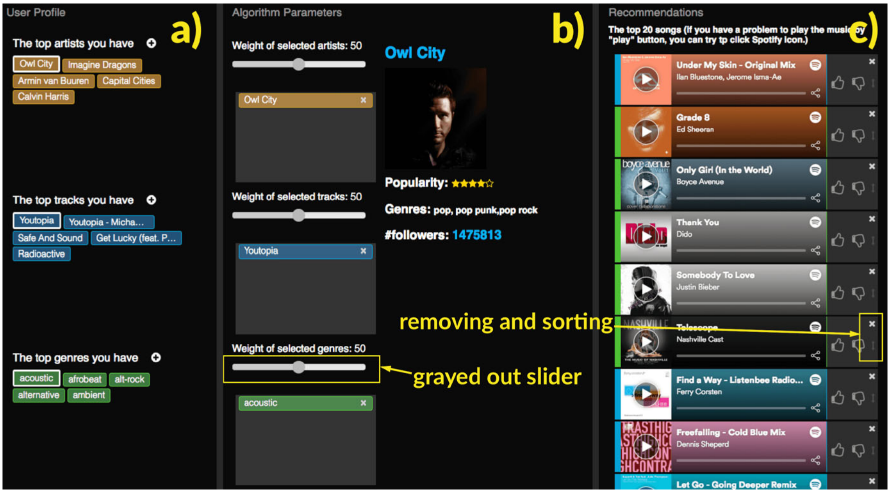

Fig. 7 **a** The recommendation source shows available top artists, tracks and genre tags. **b** The recommendation processor enables users to adjust the weight of the input data type and individual data items. **c** Playlist style recommendations. Some UI controls are disabled in specific settings of user control, e.g., the sliders in **b** are grayed out in the setting 5: REC*PRO

introduced in Sect. 3.2.3 to control the recommendation process as well as our user interface and user controls.

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 the 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.

4.1.1 User interface

The user interface of the recommender is featured with "drag-and-drop" interactions. The interface consists of three parts, as presented in Fig. 7.

(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 playlist style.

As presented in Fig. 7, we use three distinct colors to represent the recommendation source data as visual cues: brown for artists, green for tracks, and blue for genres.

Table 5 Three types of user control employed in our study

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

Additional source data for a particular type are loaded by clicking the **"+"** icon next to the title of the 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.

4.1.2 User controls

Based on our framework described in Sect. 3.1, we defined three user control components in our study: (1) user profile (PRO), (2) algorithm parameters (PAR), (3) recommendations (REC) (see Table 5).

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 (Fig. 7b).

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 (Fig. 7a). 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 playlist whose name contains the selected genre tag.

Table 6 Experimental settings: a cell filled by "*" indicates this control feature is available in the corresponding setting

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

Setting 1 is a baseline

Control for recommendations *(REC)* This type of control influences the recommended songs directly. Since the order of items in a list may affect the experience of recommendations (Zhao et al. 2017), manipulations on recommendations include reordering tracks in a playlist. It also allows users to remove an unwanted track from a playlist. 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 playlist, the rating data are not used to update the user's preference for music recommendations. Therefore, user ratings are not considered as a user control for the purposes of this study.

4.2 Evaluation

4.2.1 Evaluation methods

We employed a between-subjects study to investigate the effects of interactions among different user control on acceptance, perceived diversity, and cognitive load. We consider each of three user control components as a variable. By following the $2 \times 2 \times 2$ factorial design, we created eight experimental settings (Table 6), which allows us to analyze three main effects, three two-way interactions, and one three-way interaction. We also investigate which specific *personal characteristics* (musical sophistication, visual memory capacity) influence acceptance and perceived diversity. 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.

As presented in Sect. 3.2.6, we employed Knijnenburg et al.'s framework (Knijnenburg et al. 2012) to measure the six subjective factors, perceived quality, perceived diversity, perceived accuracy, effectiveness, satisfaction, and choice difficulty (Knijnenburg et al. 2012). In addition, we measured cognitive load by using a classic cognitive load testing questionnaire, the NASA-TLX.[11]

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

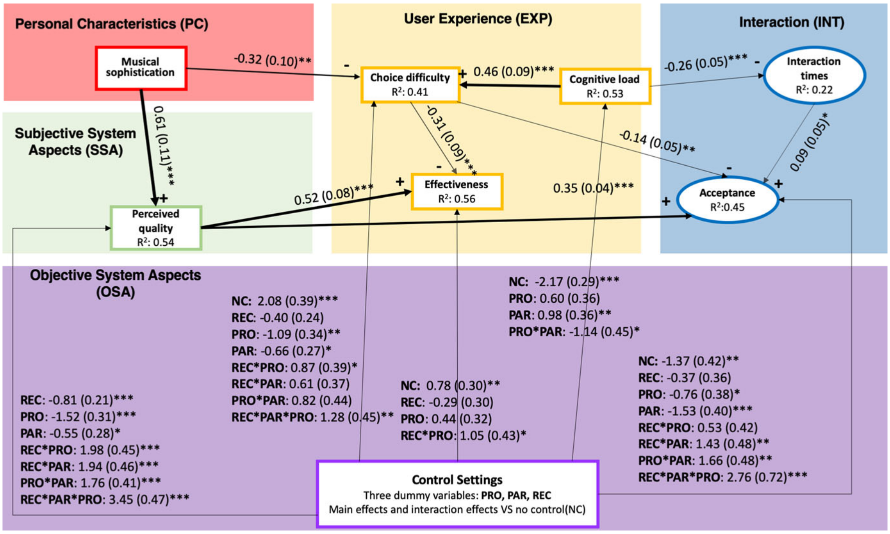

Fig. 8 The structured equation modeling (SEM) results. The number (thickness) on the arrows represents the β coefficients and standard error of the effect. Significance: $***p < .001$, $**p < .01$, $*p < .05$. R^2 is the proportion of variance explained by the model. Factors are scaled to have an SD of 1

4.2.2 Procedure

The procedure follows the design outlined in the general methodology (c.f., Sect. 3.2.1). The *experimental task* is to compose a playlist for the chosen scenario by interacting with the recommender system. Participants were presented with playlist style recommendations (Fig. 7c). 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.

4.3 Results

This experiment aims to investigate how *personal characteristics* (RQ1) and *user controls* (RQ2) influence user perception (diversity, acceptance, and cognitive load), and how personal characteristics moderate the effect of user control on user perception (RQ3).

4.3.1 Analytical approaches

We employ three validated questions to measure each subjective factor in a questionnaire. To establish the validity of these question items, we perform a confirmatory factor analysis (CFA) before evaluation. We eliminated the factors *perceived diversity* from the model because of low AVE[12] (0.41), which are lower than the recommended value 0.5. We also removed the factor *satisfaction* based on the modification indices, because all the items of satisfaction load on perceived quality are very large. As a result, we refine the answers to our questions and establish the validity of the factors in our study.

After the iterative trimming steps, Fig. 8 shows our fitted SEM model which consists of eight experimental conditions and four subjective factors: perceived quality, effectiveness, choice difficulty, and cognitive load. Objective system aspects (OSA) are represented by experimental conditions. Based on previous studies (Knijnenburg et al. 2012), we chose two factors for subjective system aspects (SSA): perceived accuracy and perceived quality. In addition, we define three factors: effectiveness, choice difficulty, and cognitive load for user experience (EXP). In interaction (INT), we count the number of liked songs in the playlist and the total interaction times with control components. Moreover, this model takes cognitive load as a component, given that we expect a difference across the control settings.

The fit of our SEM model is adequate: $\chi^2_{130} = 196.116$, $p < .001$; root-mean-squared error of approximation (RMSEA) $= 0.046$; comparative fit index (CFI) $=$ 0.971; Turker–Lewis Index (TLI) $= 0.954$.

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.[13] All answers to the questions are modeled as ordinal variables. We introduce three 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.

4.3.2 General results

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

Recommendation acceptance 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 shows significantly negative effects on acceptance. In contrast, the two-way interaction effects and three-way interaction effects show significantly positive effects on acceptance. (Figure 9a 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 9f shows the marginal effects of control settings and their interaction on perceived quality.) Moreover, quality positively influences acceptance. Thus, from the perspective of user

[12] 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.

[13] http://lavaan.ugent.be/, accessed August 2019.

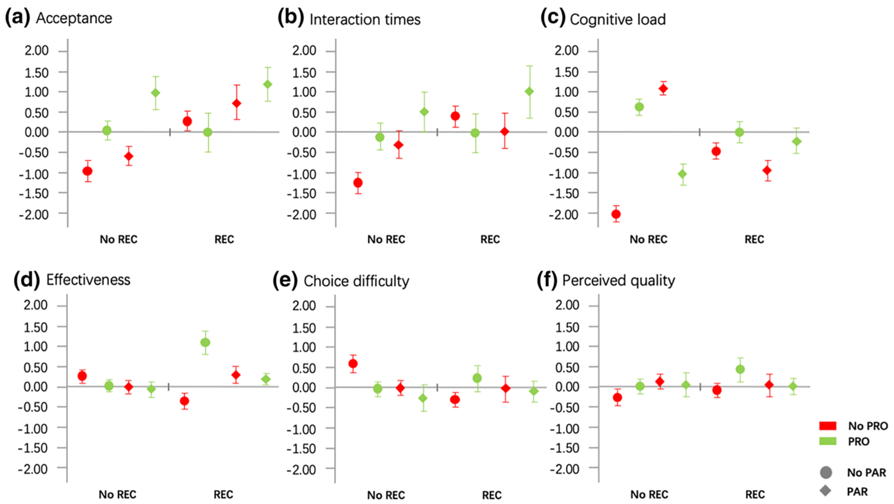

Fig. 9 Marginal effects for three control components (REC, PRO, and PAR) on user interactions: **a** acceptance and **b** interaction time; user experience: **c** cognitive load, **d** effectiveness, and **e** choice difficulty; and subjective aspect: **f** perceived quality. Legend given for PRO and PAR

control, this result supports hypothesis **H1**: More sophisticated UI (user control) will increase recommendation acceptance.

Perceived diversity According to the result of CFA, the question items of diversity measured in our questionnaire do not have a convergent validity due to low AVE value. Thus, we are not able to measure how user controls and personal characteristics influence perceived diversity effectively. Unfortunately, in *Experiment 1*, we cannot validate the hypotheses (H2, H5, and H8) related to perceived diversity.

Cognitive load The results of SEM (see Fig. 8) show that *the control settings directly affect cognitive load*. More specifically, individual control on PRO or PAR tends to increase cognitive load, while the interaction effect of PRO*PAR has a significantly negative effect on cognitive load (also see Fig. 9c). In turn, the increased cognitive load also increases the choice difficulty and decreases the interaction times. Thus, we cannot accept the hypothesis **H3**: More sophisticated UI (user control) will increase cognitive load.

Other interactions Additionally, the results of SEM (see Fig. 8) show that the settings of control (OSA) significantly correlate with all the measured factors of subjective system aspects (SSA) and user experience (EXP) directly.

4.3.3 Personal characteristics

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

Visual memory The SEM did not show a significant effect of visual memory on cognitive load or acceptance (INT), and is not depicted in Fig. 8 (PC). This suggests that users' visual memory does not correlate with cognitive load or acceptance. Therefore, we remove the visual memory in our model. Thus, the result does not support hypothesis **H6:** Users with higher visual memory (VM) are more likely to have less cognitive load. Besides, we did not find a moderation effect of VM on the significant effects of user control on cognitive load. Thus, we cannot accept the hypothesis **H9:** Higher visual memory (VM) tends to strengthen the effect of user interface on cognitive load.

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, decreasing choice difficulty leads to high effectiveness and high acceptance (PC → SSA → EXP → INT). Thus, choice difficulty acts as a mediator. The result supports the hypothesis **H4:** Users with higher musical sophistication (MS) are more likely to accept more recommended songs. However, musical sophistication does not have significant moderation effect on the impact of user control on acceptance. Thus, we cannot accept the hypothesis **H7:** Higher musical sophistication (MS) tends to strengthen the effect of user interface on acceptance of recommendations.

4.3.4 User actions

In Experiment 1, we mainly record how many times the user tweak the control components. In specific, we capture the times the algorithm parameters were modified (*parChange*), the times the seeds were added and removed (*proChange*), and the times the recommended items were removed and sorted (*recChange*). The SEM model shows that higher cognitive load will decrease the total interaction times, which in turn decrease the recommendation acceptance.

4.4 Discussion of Experiment 1 results

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 $2 \times 2 \times 2$ factorial design.

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

4.4.1 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 Fig. 9c). The increased cognitive load, in turn, leads to lower interaction times, thereby decreasing recommendation acceptance. Compared 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 control the underlying mechanism of recommendations.

4.4.2 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 for (PRO, PAR), which shows a benefit of combining user profile and algorithm parameters in user control. Moreover, combining 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.

4.4.3 Three-way interaction effects

The interaction of PRO*PAR*REC tends to increase acceptance (see Fig. 9a), and it does not lead to higher cognitive load (see Fig. 9c). Moreover, it also tends to increase interaction times. Therefore, we may consider having three control components in a system.

Consequently, we answer the research question. **RQ3**: *How does the complexity of the user interface (user controls) influence user perception of recommendations?* It seems that more complex user controls (combining PAR with a second control component or combining three control components) tend to increase acceptance significantly.

To keep the UI layout consistent in all experiment settings, we decided to disable the unsupported control functions by graying out the UI components rather than hiding them. However, this design decision may also influence user perception especially for cognitive load, and, to some extent, explains why more complex user controls do not increase cognitive load significantly.

4.4.4 Effects of personal characteristics

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

Visual memory (VM) The SEM model suggests that visual memory is not a significant factor that affects the cognitive load of controlling recommender systems or moderates the effect of user interface on cognitive load. 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. Finally, the question items for diversity did not converge in our model, so we are not able to make a conclusion about the influence of visual working memory on diversity.

Musical sophistication (MS) 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 leads to low choice difficulty, which in turn increases acceptance. Since the total indirect effects is significant (Est. $= 0.19$, SE $= 0.04$, $p < .001$), we can conclude MS has a positive effect on recommendation acceptance. Although we did not find a moderation effect of MS on the impact of user control on recommendation acceptance, the result implies 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. Finally, the question items for diversity did not converge in our model, so we are not able to make a conclusion about the influence of musical sophistication on diversity in Experiment 1.

4.4.5 Discussion on experimental design

When we compared different conditions of user control, we disabled the UI components that do not function in the current experimental condition for two reasons as opposed to hiding these components. First, this design decision was made to minimize the effect of UI variation on user perception of music recommendations: hiding a part of the UI would mean a very large difference between control conditions, with highly differing amounts of information as well as space on the screen.

Second, hiding the disabled control components may negatively affect the user understanding of the recommendation. Our user interfaces do not only support user control, but also illustrate how the recommended songs are selected.

However, as disabling the unsupported UI control components may cause a mismatch between user expectation and actual operation, we have tried to mitigate the issue of expectations by explicitly explaining which functions are available in the current condition in our video tutorial.

Therefore, in our design of the experiment, we make a trade-off between the consistency of UI and user mental model.

5 Experiment 2: effects of visualizations

To investigate whether the validated hypotheses still hold for another part of UI framework, *visualizations*, we implemented two versions of bubble charts having different levels of complexity. We are interested in seeing how personal characteristics and visualizations influence perceived diversity, acceptance, and cognitive load, as well as how the personal characteristics influence the impact of visualizations on user perception.

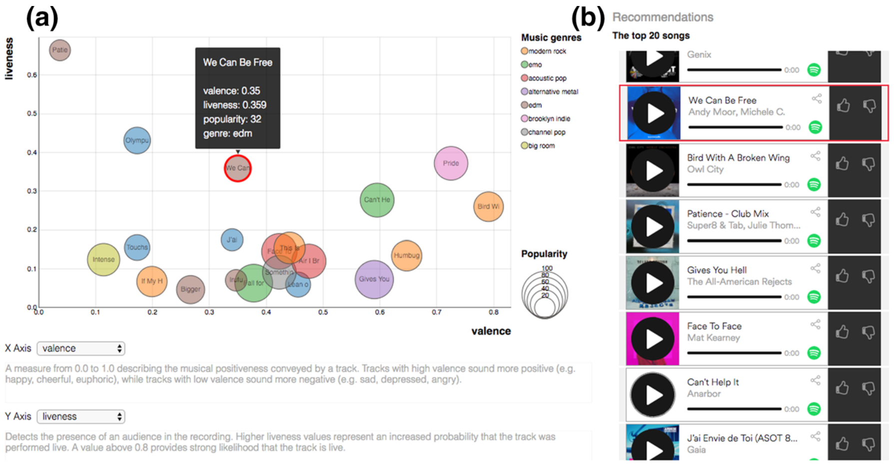

Fig. 10 Design of the user interface for a music recommender, section **a** a visualization view of the diversity of recommendations (*ComBub*); section **b** a list view of recommendations

5.1 Setup

5.1.1 User interfaces

Figure 10 illustrates the design of a user interface which consists of two sections: *section (a)* a visualization view shows an interactive visualization which allows users to explore songs by visualized attributes; *section (b)* a list view shows all items in a list, and each of them is associated with a particular circle in the visualization view. When the user clicks on a circle, the corresponding item in the list will be highlighted (red border) and vice versa. Each item in the list has a play icon and a thumb rating widget.

We hypothesize that visualizing additional meta-data of music such as audio features may result in higher perceived diversity (H2). Therefore, we designed the interfaces with two requirements. First, the visualizations should present multiple data dimensions effectively: in our case, we show two common attributes *genres and popularity*, and seven *additional audio features*. Second, the visualization should represent coverage by a particular attribute to reflect diversity, e.g., how the items are distributed by genres. Based on the above considerations, the bubble chart is selected as our primary visualization due to its good ability to present multidimensional data (Kim et al. 2016). Moreover, to test our assumption, we also need to compare this relatively complex bubble chart (*ComBub*) with a baseline visualization. We consider a simple bubble chart (*SimBub*) as a good candidate since it meets the first requirement and uses almost the same visual presentation as *ComBub*. The visualizations were implemented with the D3.js library.[14]

ComBub Section (a) of Fig. 10 shows the design of *ComBub* that encodes the recommendations results in three ways. First, it uses a circle to represent each recommended

[14] https://d3js.org/, accessed June 2018.

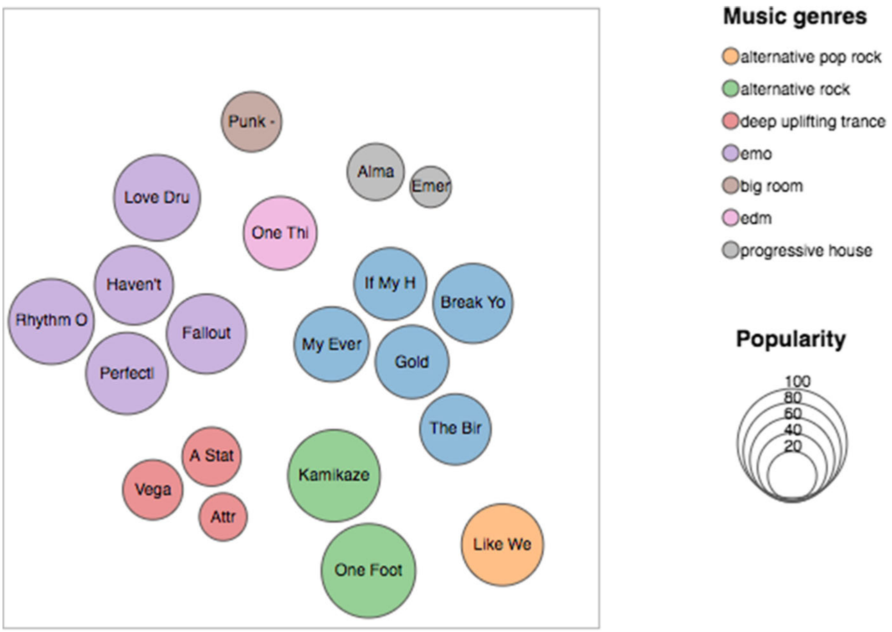

Fig. 11 Design of the baseline *SimBub* visualization for enhancing perceived diversity of recommendations

song: the X-axis and Y-axis are used to present two specified audio features. Second, the circle is color-coded for music genres, which allows users to distinguish song genres by their color. Third, the circle size (radius) is determined by the popularity score (from 1 to 100) which has been transformed by a visual square-root function. The function is defined as:

$$R(p) = 6 * \sqrt{\frac{p}{\pi}} \tag{1}$$

where p is an item's popularity score.

This encoding allows the user to inspect multiple dimension of the song simultaneously. The interface can be used to support advanced exploration for users such as popular pop songs with high danceability and high valence (happy, cheerful).

Common interactions such as zooming and panning are supported. The details of a particular item will appear in a tooltip window when the mouse hovers over it. By clicking on a circle, its associated item will be highlighted in the list synchronously. Below the plot, two drop-down menus are used to select audio features to visualize songs on the bubble chart. The scale of all audio features ranges from 0.0 to 1.0.

In summary, *ComBub* allows users to specify two audio features to plot recommendations in two dimensions and inspect the details and distribution of genres and popularity as they wish. As explained above, the visualization is able to explain the diversity of recommendations from various aspects.

SimBub Figure 11 illustrates the design of *SimBub*. To save space, the figure omits the recommendation list associated with the visualization that is identical to the one in section (b) of Fig. 10. We designed the simplest form of a bubble chart as a baseline for two reasons. First, this bubble chart represents items by labeled circles, which is a popular visualization among 13 common visualizations evaluated for visualizations at Internet scale (Viegas et al. 2007). Second, it can be seen as a variation of *ComBub* without presenting audio features. Thus, it is easier for us to investigate the effects of the additional visualized audio features in *ComBub*. Compared to *ComBub*, this chart may be easier and sufficient for casual users to interpret and perceive diversity. In this sense, our study answers the question whether showing the additional audio features can lead to added value in terms of diversity and other investigated metrics of recommendations.

5.2 Evaluation

To address our research questions, we conducted a user study to evaluate two visualizations in terms of recommendation acceptance, perceived diversity, and cognitive load.

To separate the effects of algorithm, we control the actual diversity of recommendations to stay at a compared and moderate level. The actual diversity was measured by intra-list similarity (ILS) (Ziegler et al. 2005) on music genres. We measure the similarity $C_o\,(b_k, b_e)$ between items b_k, b_e based on the Jaccard similarity coefficient. Intra-list similarity for a_i's list P_{w_i} is defined as follows:

$$\mathrm{ILS}(P_{w_i}) = \frac{\sum_{b_k \in \Im P_{w_i}} \sum_{b_e \in \Im P_{w_i}, b_k \neq b_e} C_o\,(b_k, b_e)}{2} \tag{2}$$

The Jaccard similarity is the number of common features for two sets A and B divided by the total number of features in the two sets.

$$C_o\,(b_k, b_e) = \frac{\left|b_k \bigcap b_e\right|}{\left|b_k \bigcup b_e\right|} \tag{3}$$

For all participants, recommendations shown in the two visualizations have a similar actual diversity calculated by ILS score (*ComBub*: Mean $= 21.39$, SD $=$ 1.32, *SimBub* : Mean $= 20.87$, SD $= 1.65$). Lower scores obtained denote higher diversity.

5.2.1 Evaluation methods

We conducted a between-subjects study where participants evaluated two user interfaces (*ComBub* vs. *SimBub*).

The *independent variable* of the study is the type of visualization. We employ the questionnaire described in Table 3.

5.2.2 Procedure

Same as in Experiment 1, we first asked users to read a brief description of the study task and to watch a one-minute video that shows all the functions and interactions supported by each visualization. They then fill the same pre-study questionnaire before starting study.

Participants were given the same *task* while testing the two visualizations: each participant needs to listen to and rate all songs in the list with the possibility to explore recommendations through the interface.

Despite the same algorithm and input seeds, the recommendations generated by Spotify vary between different requests. Thus, the potential influence of users' familiarity with recommendation data is avoided. After using each visualization, the user was asked to fill out a post-study questionnaire.

5.3 Results

This experiment aims to investigate how *personal characteristics* (RQ1) and *visualizations* (RQ2) influence user perception (diversity, acceptance, and cognitive load), and how personal characteristics moderate the effect of visualizations on user perception (RQ3).

5.3.1 Analytical approaches

Similar to the analytical approaches in Experiment 1, we first perform a confirmatory factor analysis (CFA) to establish the validity of question items before evaluation. We eliminated the factors *satisfaction* from the model based on the modification indices, because two question items (Q12, Q13) of satisfaction load on perceived quality are very large. As a result, we refine the answers to our questions and establish the validity of the factors in our study.

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.[15] All answers to the questions are modeled as ordinal variables.

After the iterative trimming steps, Fig. 12 shows our fitted SEM model which consists of two experimental conditions and five subjective factors: perceived quality, perceived diversity, effectiveness, choice difficulty, and cognitive load. Objective system aspects (OSAs) are represented by experimental conditions. Subjective system aspects (SSAs) refer to perceived quality. In addition, we define three factors: effectiveness, choice difficulty, and cognitive load for user experience (EXP). We count the number of liked songs in the playlist and the number of interactions (clicking and hovering) with the items in visualizations for interaction (INT).

The fit of our SEM model is adequate: $\chi^2_{83} = 111.852$, $p = .019$; root-mean-squared error of approximation (RMSEA) $= 0.060$; comparative fit index (CFI) $=$ 0.981; Turker–Lewis Index (TLI) $= 0.982$.

[15] http://lavaan.ugent.be/, accessed February 2019.

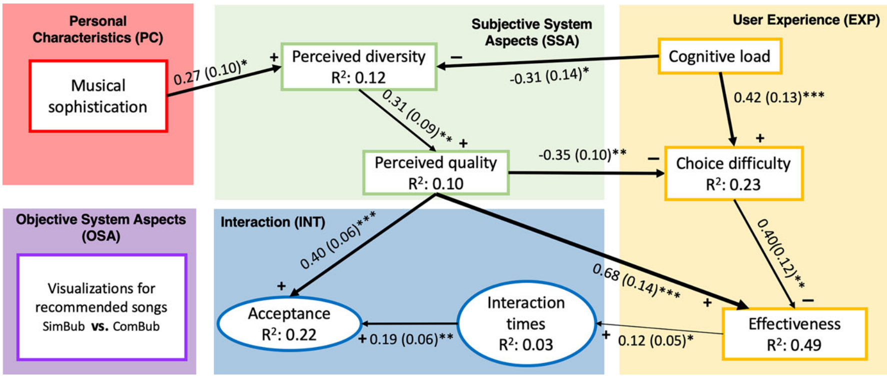

Fig. 12 Structured equation modeling (SEM) results of Experiment 2. The number (thickness) on the arrows represents the β coefficients and standard error of the effect. Significance: $***p < .001$, $**\ p < .01$, $*p < .05$. R^2 is the proportion of variance explained by the model. Factors are scaled to have an SD of 1

5.3.2 General results

Recommendation acceptance We calculated recommendation acceptance by the percentage of liked songs in the playlist. Figure 12 does not show a significant effect of visualizations on the acceptance of the resulting recommendations. Thus, the result does not support hypothesis **H1**: More sophisticated UI (visualizations) will increase recommendation acceptance.

Perceived Diversity The results of SEM does not show a significant effect of visualizations on perceived diversity, which does not support the hypothesis **H2**: More sophisticated UI (visualizations) will increase recommendation diversity.

Cognitive load We did not find a significant effect of visualizations on cognitive load from the results of SEM. Thus, we cannot accept the hypothesis **H3**: More sophisticated UI (visualizations) will increase recommendation cognitive load.

5.3.3 Personal characteristics

We then check whether the two personal characteristics musical sophistication (MS) and visual memory (VM) significantly influence user perception and the impact of visualization on user perception.

Visual Memory (VM) We do not find any significant effect of VM on the measured variables: recommendation acceptance, perceived diversity, and cognitive load. Thus, no result supports the hypothesis **H6**: Users with higher VM are more likely to have less cognitive load. We also do not find a moderation effect of VM on the impact of visualization on user perception. Thus, we cannot accept the hypothesis **H9**: Higher VM tends to strengthen the effect of user interface on cognitive load.

Musical Sophistication (MS) The SEM results show that MS positively influences perceived diversity directly, which in turn positively influences acceptance of recommendations via the mediator perceived quality. Thus, we can accept hypothesis **H4**: Users with higher MS are more likely to accept more recommended songs, and hypothesis **H5**: Users with higher MS are more likely to perceive higher diversity.

We also perform a moderation analysis to better understand how MS influences the relation between visualization and user perception (perceived diversity, acceptance). The results of the moderation analysis does not show a significant effect. Thus, we cannot accept either hypothesis **H7**: Higher musical sophistication (MS) tends to strengthen the effect of user interface on acceptance of recommendations, nor **H8**: Higher musical sophistication (MS) tends to strengthen the effect of user interface on perceived diversity.

5.3.4 User actions

In this experiment, we capture how many times the users interact (clicking and hovering) with the items presented in visualizations. The model shows that higher MS will increase the perceived diversity, which in turn leads to more user interactions via the mediators perceived quality and effectiveness (PC $\rightarrow$ SSA $\rightarrow$ EXP $\rightarrow$ INT).

5.4 Discussion of Experiment 2

Overall, no significant difference was found between the two visualizations in terms of users' acceptance of recommendations, perceived diversity, or cognitive load.

Visualizing the audio features of music has a limited impact on acceptance and perceived diversity Compared to *SimBub*, additional audio features of songs visualized in *ComBub* do not have significant added value for increasing acceptance and perceived diversity if we disregard the effect of the personal characteristics MS and VM.

Although visualizing some additional features may increase the *acceptance and understanding* (Andjelkovic et al. 2016), we do not find such a benefit for *ComBub*. *ComBub* does not allow users to modify audio feature data to update recommendations. Thus, we speculate that the lack of controllability hinders the value of visualizing additional data for acceptance. Moreover, we think that the understandability of what the features refer to could be a problem that hindered many people from profiting from the visualization of audio features. We further speculate that the results may indicate that the visualization in contrast is good at helping users find how items are different in terms of their audio features.

MS positively influences acceptance in an indirect way Similar to the results of Experiment 1, we also find that MS positively influences acceptance of recommendations. We speculate that in general users with high MS are able to make better use of visualizations to inspect the recommendations.

MS positively influences perceived diversity The SEM results show a significantly positive correlation between MS and the perceived diversity. Since we did not find significant difference of user perception between two visualizations, we cannot investigate the moderation effect of MS on the impact of visualization on user perception. This result implies that in general users with higher MS tend to perceive higher diversity through visualizations.

6 Experiment 3: effects of combining user control and visualizations

In the previous two studies, we investigated the effects of user control and visualizations separately. Experiment 3 investigates the effects of combining user control and visualizations in terms of acceptance, perceived diversity, and cognitive load, as well as the influence of personal characteristics.

6.1 Setup

The recommender algorithm is the same as the one used in Experiment 1 (user controls), based on the seed-based recommender engine provided by Spotify.

6.1.1 User interfaces

In this experiment, we have three conditions as described below: Control, SimBub+Control, and ComBub+Control.

Full Control Identically to the setting of full user control REC*PRO*PAR in Experiment 1, users are allowed to manipulate three recommender components, recommendations, user profile, and algorithm parameters.

Visualizations We further implement two new user interfaces (*ComBub + Full Control*, and *SimBub + Full Control*) that combine *Control* with either the ComBub or SimBub visualization (as introduced in Experiment 2). Consequently, the only difference between the two interfaces is the visualization: one has a simple bubble chart (SimBub), whereas the other one has a relatively complex bubble chart (ComBub) as shown in Fig. 13. Participants are able to explore recommendation items through either visualization. After users tune the recommendation parameters, the visualization also updates based on the newly generated recommendations.

6.2 Evaluation

In this experiment, we want to evaluate the user interfaces combining full control and (one of either two) visualization(s), against the baseline user interface only having full user control. We conducted a between-subjects experiment where participants evaluated one of the three interfaces.

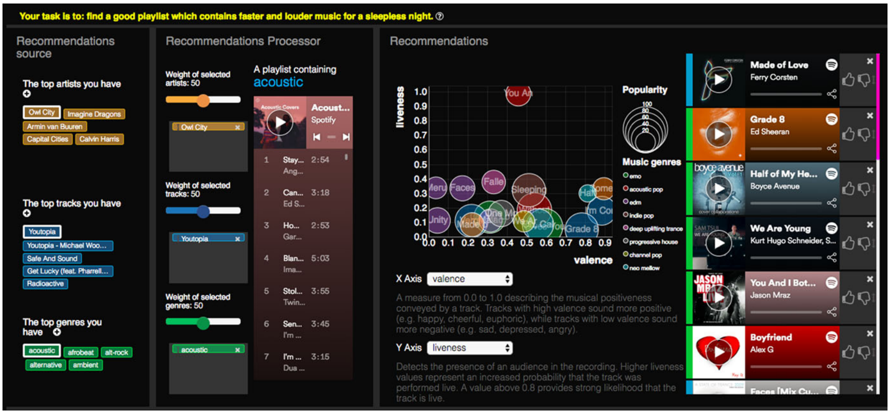

Fig. 13 This user interface combines a visualization (ComBub) and the full user control widgets

6.2.1 Procedure

The participants follow the same study procedure to evaluate the three user interfaces. In the end, they were asked to fill the same questionnaire used in the previous experiments.

6.3 Results

This experiment aims to investigate how *personal characteristics* (RQ1) and *user control* + *visualizations* (RQ2) influence user perception (diversity, acceptance, and cognitive load), and how personal characteristics moderate the effect of user control + visualizations on user perception (RQ3).

6.3.1 Analytical approaches

Similar to the analytical approaches in the previous two experiments, we first perform a confirmatory factor analysis (CFA) to establish the validity of the question items before evaluation. We eliminated the factor *satisfaction* from the model based on the modification indices, because the items of satisfaction (Q12, Q13) load on effectiveness are very large. As a result, we refine the answers to our questions and establish the validity of the factors in our study.

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.[16] All answers to the questions are modeled as ordinal variables. After the iterative trimming steps, Fig. 14 shows our fitted SEM model which consists of three experimental conditions and five subjective factors: perceived quality, perceived diversity, effectiveness, choice difficulty, and cognitive load. The fit of our SEM model is adequate: $\chi^2_{137} = 198.463$, $p < .001$; root-mean-squared error of

[16] http://lavaan.ugent.be/, accessed February 2019.

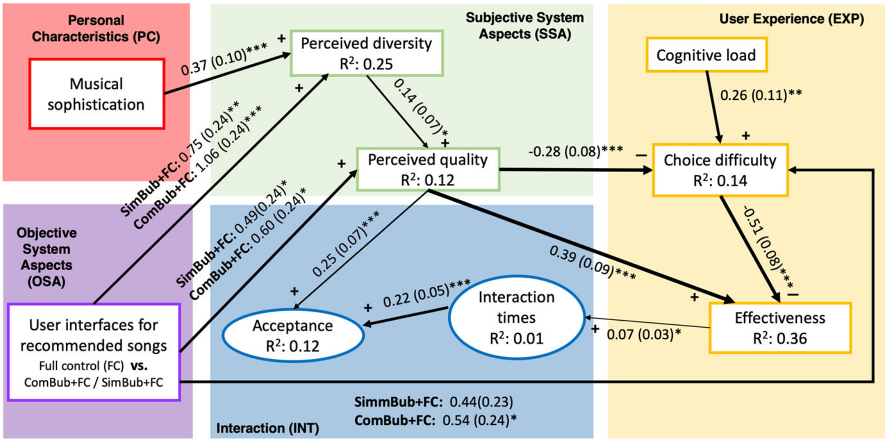

Fig. 14 The structured equation modeling (SEM) results of Experiment 3. The number (thickness) on the arrows represents the β coefficients and standard error of the effect. Significance: $***p < .001$, $**p < .01$, $*p < .05$. R^2 is the proportion of variance explained by the model. Factors are scaled to have an SD of 1

approximation (RMSEA) = 0.053; comparative fit index (CFI) = 0.978; Turker–Lewis Index (TLI) = 0.983.

6.3.2 General results

Recommendation acceptance Although the results of our SEM do not show a significant *direct* effect of user interfaces on recommendation acceptance of recommendations (see Fig. 14), we do find *indirect* influences of the user interface on acceptance. More specifically, both ComBub + Full Control and SimBub + Full Control tend to increase both perceived diversity and quality (Fig. 15a, b shows the marginal effects of user interfaces on diversity and quality, respectively), which in turn leads to higher recommendation acceptance.

However, we also find that ComBub + Control significantly increases the choice difficulty , which results in lower acceptance. The marginal effects of UI on acceptance (Fig. 15f) show that these two opposite indirect effects probably cancel out. In sum, these results do not support hypothesis **H1**: The more sophisticated UI will increase recommendation acceptance.

Perceived diversity. The results of the SEM show that the complex UI has a positive effect on perceived diversity. In addition, musical sophistication also positively influences the diversity. The marginal effects of UI on diversity (Fig. 15a) indicate that the diversity in ComBub + Control and SimBub + Control is higher than the baseline condition (Control). Thus, the results support the hypothesis **H2**: The more sophisticated UI will increase perceived diversity.

Cognitive load Although the marginal effects of UI on cognitive load (Fig. 15e) show that the values in both ComBub + Control and SimBub + Control are slightly

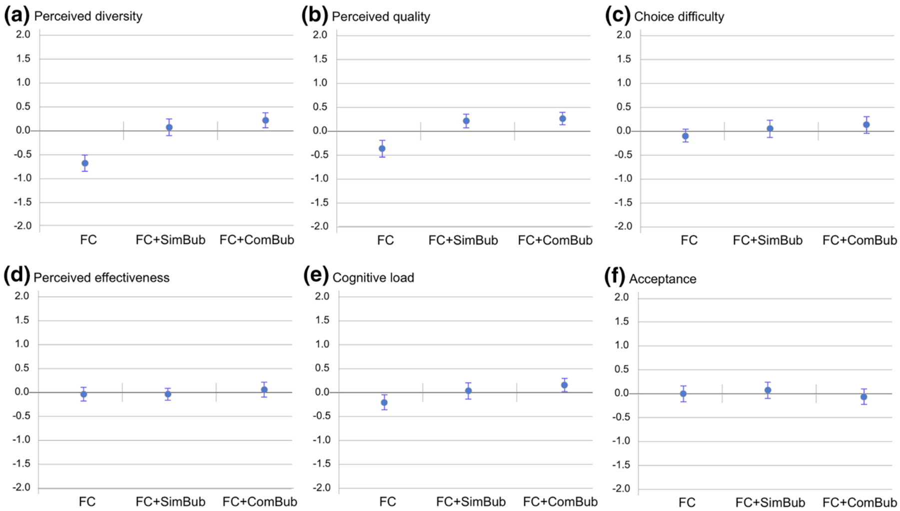

Fig. 15 Marginal effects for two visualizations (Control + SimBub and Control + ComBub) on three factors that are significantly influenced by user interface: **a** perceived diversity and **b** perceived quality, **c** choice difficulty, **d** perceived effectiveness, **e** cognitive load, and **f** acceptance. The effects of the baseline "full control" condition are set to zero

higher than the baseline condition (Control), the results of our SEM do not show a significant effect of the UI on cognitive load. Thus, we cannot accept hypothesis **H3**: The more sophisticated UI will increase cognitive load.

6.3.3 Personal characteristics

This section will present the main effect of personal characteristics on user perception as well as their moderation effect on the relation between the user interface and user perception.

Visual Memory (VM) The SEM does not show a significant effect of visual memory on any of the measured factors. Therefore, we remove the visual memory in our model. Thus, the result does not support hypothesis **H6**: Users with higher VM are more likely to have less cognitive load. In addition, VM does not moderate the effect of user interface on cognitive load, and we therefore cannot accept support hypothesis **H9**: Higher VM tends to strengthen the effect of user interface on cognitive load.

Musical Sophistication (MS) The model shows that MS positively influences perceived diversity, which in turn leads to higher perceived quality. Furthermore, the increased quality positively influenced acceptance. Thus, we can accept hypotheses **H4**: Users with higher MS are more likely to accept more recommended songs, and **H5**: Users with higher MS are more likely to perceive higher diversity.

Additionally, we also perform a moderation analysis to investigate whether MS moderates the significant impact of user interface on perceived diversity. The result shows a significant moderation effect of MS on the relation between FC+ComBub and perceived diversity ($b = 0.20$, SE $= 0.09$, $p < .05$). To better illustrate this

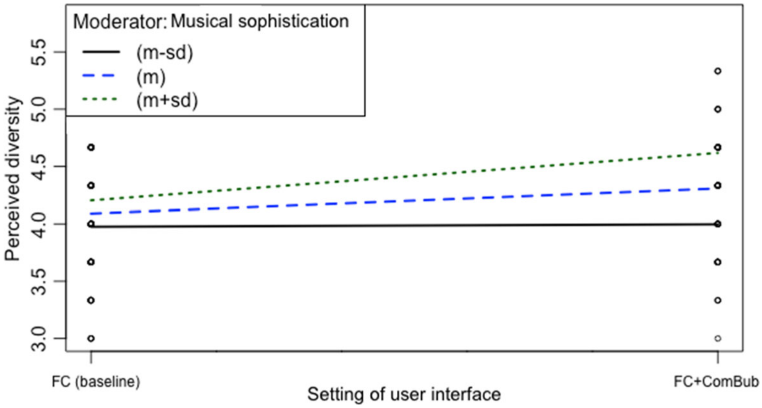

Fig. 16 Simple slopes (1 SD above and 1 SD below the mean) of the moderating effect of musical sophistication (MS) on the relation between user interface and perceived diversity

interaction, we use the rockchalk[17] function to automatically plot the simple slopes (1 SD above and 1 SD below the mean) for analyzing the moderating effect (Judd et al. 2001; Dawson 2014; Blair 2019) (Fig. 16). This figure shows that those users who have lower MS (the black solid line) perceived almost same diversity in the baseline UI (FC) and more complex UI (FC+ComBub), and perceive lower diversity overall than average (the blue dashed line). Those users who have higher MS (the green dotted line) perceive higher diversity when they have a more sophisticated UI (FC+ComBub) as well and perceive higher diversity than average. The difference in the slopes for those who have higher or lower MS shows that MS moderates the relationship between UI and perceived diversity. Therefore, the results support hypothesis **H8**: Higher MS tends to strengthen the effect of user interface on perceived diversity, but cannot accept hypothesis **H7**: Higher MS tends to strengthen the effect of user interface on acceptance of recommendations.

6.3.4 User actions

In this experiment, we recorded user interaction with both user control components and visualizations. The SEM model shows that the UI negatively influences the interaction times via the mediators choice difficulty and effectiveness. Moreover, we see that number of interactions (interaction times) positively influences recommendation acceptance.

6.4 Discussion of Experiment 3 results

We observe two opposite indirect effects of UI on recommendation acceptance. Despite the positive effects of a sophisticated UI on diversity and quality, the more sophisticated UI also seems to increase the choice difficulty, thereby decreasing other UX

[17] https://cran.r-project.org/web/packages/rockchalk/index.html.

factors such as effectiveness. As a result, the decreased effectiveness will counterbalance the increased acceptance. Consequently, we argue that the sophistication of UI (visualizations) has a very limited effect on recommendation acceptance.

In addition, compared to the baseline condition *Full Control*, adding visualization to full control increases the perceived diversity and quality significantly. However, we do not find significance in perceived diversity between the two conditions that combine visualizations and user control, which is in line with the result of Experiment 2. Arguably, visualizations add value to the full user control in terms of perceived diversity and quality.

Moreover, as we have found in the previous two experiments, visual memory does not have a significant effect on any measured factors. However, we find that *musical sophistication* significantly influences *perceived diversity* of music recommendations as well. Meanwhile, musical sophistication has a moderating effect that strengthens the positive correlation between (type of) UI and perceived diversity.

7 General discussion and conclusion

We investigated the main effects of two personal characteristics on the perception of the music recommendations with user control-oriented user interfaces by three experiments. We employed the user-centric framework of Knijnenburg et al. (2012) to construct our conceptual model and evaluate the music recommender system mainly from three aspects, acceptance, perceived diversity, and cognitive load. Moreover, the moderation analyses allow us to demonstrate how personal characteristics influence the impact of three user control levels and visualizations on user perception.

In the next section (Sect. 7.1), we present a general discussion of our results. This is followed by limitations (Sect. 7.2) and concluding remarks including suggestions for future work in Sect. 7.3.

7.1 Discussion of results

Personal characteristics Our results suggest that in general *musical sophistication* (MS) positively influences recommendation acceptance and perceived diversity regardless of complexity of user interface. MS allows users to perceive higher recommendation quality, which in turns appears to lead to high item acceptance (in Experiments 1 and 3). A positive effect of MS on perceived diversity was also found when we compared (in Experiment 2) a more sophisticated bubble chart against a simplified bubble chart, as well as when we evaluated (in Experiment 3) the full control setting combined with the more sophisticated bubble chart (Control + ComBub) against the full control combined with a simplified bubble chart (Control + SimBub).

Furthermore, the moderating effect found in Experiment 3 implies that users with high MS tend to perceive higher recommendation diversity through more complex user interface.

The results of three experiments suggest that *visual memory* (VM) may not be a significant factor that affects the cognitive load of controlling recommender systems.

The cognitive load for the type of controls and visualizations used may not be strongly affected by individual differences in visual memory. In other words, controlling the more advanced recommendation components and viewing the sophisticated visualization do not seem to demand a high visual memory.

Control levels In addition, we also investigated the effects of *control levels* on user perception, which show a significant difference among different settings of user control in terms of cognitive load and acceptance.

Compared to the control of algorithm parameters or user profile, the control of recommendations 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 ratings. However, the control of recommendations is a limited operation that does not allow users to understand or control the deep mechanism of recommendations. Adding multiple controls allows us to improve on existing systems w.r.t. control, and does not necessarily result in higher cognitive load. Adding an additional control component to algorithm parameters increases the acceptance of recommended songs significantly. Interestingly, the UIs 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* in controlling both user profile and algorithm parameters than only controlling each of them alone, which shows a benefit of combining user profile and algorithm parameters in user control. Arguably, it is beneficial to add either the control for recommendations, or user profile, to the control for algorithm parameters. Thus, we can conclude that increasing the complexity of user control by incorporating multiple control components leads to higher recommendation acceptance but does not increase cognitive load significantly.

Research Questions Consequently, we are able to answer the research questions outlined in the beginning of the manuscript. We address each question individually:

RQ1: Main effects of personal characteristics *How do personal characteristics influence user perception of recommendations (diversity, acceptance, and cognitive load)?* In Experiment 1, the SEM shows that MS positively influences the perception of recommendation quality and effectiveness, and results in a higher level of song acceptance. Therefore, we infer that users with higher MS have better ability to leverage different control components to explore songs, and this influences their perception of recommendation quality, thereby accepting more songs. Since the AVE value of the factor perceived diversity does not indicate convergent validity, we cannot measure the effects on diversity.

The results of Experiment 2 suggest that users with higher MS tend to perceive higher diversity. In addition, MS also influences recommendation acceptance via the mediators perceived diversity and perceived quality.

In Experiment 3, we found that MS positively influences recommendation acceptance via the mediators perceived diversity and quality, and MS also influences perceived diversity directly, which is in line with the results of Experiment 2.

To sum up, MS positively influence perceived diversity and acceptance, but VM does not affect cognitive load significantly.

RQ2: Moderating effects of personal characteristics *How do personal characteristics moderate the effect of user controls/visualizations on user perception of recommendations (diversity, acceptance, and cognitive load)?* The moderating effect of MS on the relation between UI and diversity implies that users with high MS tend to make better use of a more sophisticated UI such as the UI combining both full user control and the complex bubble chart (ComBub), which in turn leads to higher perceived diversity.

Therefore, we only find musical sophistication moderates the impact of complex UI combining user control and visualization on perceived diversity. Moreover, the results do not show a significant moderating effect of personal characteristics on acceptance and cognitive load.

RQ3: The effect of UI complexity *How does the complexity of user controls/visualizations influence user perception of recommendations (diversity, acceptance, and cognitive load)?* Having multi-level user control does not significantly increase cognitive load. It seems that combining algorithm parameters with a second control component increases acceptance significantly.

Moreover, compared with the simple bubble chart (SimBub), a more complex visualization (ComBub) does not yield higher acceptance and perceived diversity by visualizing additional audio features of music.

We then compared two UIs that combine full user control and visualizations against a baseline UI with only full user control. The results indicate combining full user control and visualizations tends to increase diversity but does not significantly influence acceptance and cognitive load.

As a result, complex user control positively influences acceptance and more complex user interfaces (combining full user control and visualizations) positively influence diversity, but complex visualization does not affect user perception.

7.2 Limitations

First, although we tried our best to minimize the potential harms to evaluation such as filtering workers and avoiding acquiescence bias by introducing contradictory statements, we cannot ignore the potential limitations (Kittur et al. 2008) of using a crowd-sourcing platform like Amazon Mechanical Turk to evaluate a system with relatively complex tasks.

Second, to minimize the effect of UI variation on user perception to music recommenders, we decided to gray out the UI components that do not function in the current experimental condition. Although we have explicitly indicated which control functions will be available in a video tutorial before study, the disabled control components may still evoke some users expectation and therefore lower satisfactions/perceptions (Sect. 4.4.5). However, we think this illustrates a trade-off that we considered when we decided whether to disable or hide the unsupported functions. The alternative, hiding a part of UI also has a very large effect on the UI layout, which may also introduce another variable in our experiment.

Table 7 Effects of the investigated personal characteristics (PC) on music recommender user interfaces in three experiments

PCs	Experiment 1 user controls	Experiment 2 visualizations	Experiment 3 Controls+Vis.
Visual memory (VM)	Acceptance (no) Diversity (no) Cognitive load (no)	Acceptance (no) Diversity (no) Cognitive load (no)	Acceptance (no) Diversity (no) Cognitive load (no)
Musical sophistication (MS)	Acceptance (+) Diversity (no) Cognitive load (no)	Acceptance (+) Diversity (+) Cognitive load (no)	Acceptance (+) Diversity (+ , m) Cognitive load (no)

"+" stands for a significantly positive effect, "no" means no significant effect, and "m" means the moderation effect of PCs on the impact of UI on user perception

Third, 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.

Finally, to ensure enough user engagement in testing two visualizations, we required users to spend at least ten minutes for each visualization and listen and rate all recommended songs. Thus, the recorded actions may not reflect the real user intention for clicking items on visualizations.

7.3 Conclusion

We have presented an in-depth study to investigate the effects of two personal characteristic, *musical sophistication* and *visual memory* on user perception. Based on our nine research hypotheses, we are particularly interested in understating how visual memory influences cognitive load and how musical sophistication influences acceptance and diversity.

Table 7 summarizes the results of our three experiments. It shows the main effects of personal characteristics on acceptance, perceived diversity[18], and cognitive load.

Our results suggest that visual memory does not influence the acceptance of recommendations, perceived diversity, and cognitive load regardless of user control elements or visualizations. In contrast, musical sophistication appears to positively affect the acceptance of recommendations for different levels of user control. Musical sophistication also appears to have an influence on perceived diversity, when visualizations are supplied. Surprisingly, personal characteristics did not interact with cognitive load for different levels of control.

Overall, these findings suggest that the design of both control widgets and visualizations can benefit from tailoring to the personal characteristic of musical sophistication.

Our future work will focus on three directions. First, it is important to extend this model by investigating other potential personal characteristics that may influence

[18] We note that the result for perceived diversity in Experiment 1 was inconclusive as the item did not fit.

cognitive load beyond musical sophistication and visual memory, such as *choice persistence* (Knijnenburg et al. 2011). Second, based on this extended model, we intend to investigate further adaptive strategies (e.g., hiding, color coding) that are suitable to the personal characteristics of users. Finally, we plan to validate our research finding in other application domains such as online learning and exploring articles on debated topics or news.

Acknowledgement This research has been supported by the KU Leuven Research Council (grant agreement C24/16/017).

References

Alba, J.W., Hutchinson, J.W.: Dimensions of consumer expertise. J. Consum. Res. **13**(4), 411–454 (1987)

Aljukhadar, M., Senecal, S., Daoust, C.E.: Using recommendation agents to cope with information overload. Int. J. Electron. Commer. **17**(2), 41–70 (2012)

Al-Maskari, A., Sanderson, M.: The effect of user characteristics on search effectiveness in information retrieval. Inform. Process. Manag. **47**(5), 719–729 (2011)

Andjelkovic, I., Parra, D., O'Donovan, J.: Moodplay: interactive mood-based music discovery and recommendation. In: Proceedings of UMAP'16, pp. 275–279. ACM (2016)

Aykin, N.M., Aykin, T.: Individual differences in human–computer interaction. Comput. Ind. Eng. **20**(3), 373–379 (1991)

Bakalov, F., Meurs, M.J., König-Ries, B., Sateli, B., Witte, R., Butler, G., Tsang, A.: An approach to controlling user models and personalization effects in recommender systems. In: Proceedings of IUI'13, pp. 49–56. ACM (2013)

Blair, A.: Chapter 14: Mediation and moderation. https://ademos.people.uic.edu/Chapter14.html (2019). Accessed 7 Mar 2019

Bogdanov, D., Haro, M., Fuhrmann, F., Xambó, A., Gómez, E., Herrera, P.: Semantic audio content-based music recommendation and visualization based on user preference examples. Inform. Process. Manag. **49**(1), 13–33 (2013)

Bostandjiev, S., O'Donovan, J., Höllerer, T.: Tasteweights: a visual interactive hybrid recommender system. In: Proceedings of RecSys'12, pp. 35–42. ACM (2012)

Bostandjiev, S., O'Donovan, J., Höllerer, T.: LinkedVis: exploring social and semantic career recommendations. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces, pp. 107–116. ACM (2013)

Bradford, G.R.: A relationship study of student satisfaction with learning online and cognitive load: Initial results. Internet High. Educ. **14**(4), 217–226 (2011)

Brusilovsky, P., Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, pp. 3–53. Springer, New York (2007)

Carenini, G., Conati, C., Hoque, E., Steichen, B., Toker, D., Enns, J.: Highlighting interventions and user differences: informing adaptive information visualization support. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 1835–1844. ACM (2014)

Champiri, Z.D., Shahamiri, S.R., Salim, S.S.B.: A systematic review of scholar context-aware recommender systems. Exp. Syst. Appl. **42**(3), 1743–1758 (2015)

Chandler, P., Sweller, J.: Cognitive load theory and the format of instruction. Cognit. Instr. **8**(4), 293–332 (1991)

Chen, L., Pu, P.: Critiquing-based recommenders: survey and emerging trends. UMUAI **22**(1–2), 125–150 (2012)

Chen, P.I., Liu, J.Y., Yang, Y.H.: Personal factors in music preference and similarity: user study on the role of personality traits. In: Proceedings of International Symposium on Computer Music Multidisciplinary Research (CMMR) (2015)

Chen, H., Parra, D., Verbert, K.: Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. Exp. Syst. Appl. **56**, 9–27 (2016)

Chernev, A.: When more is less and less is more: the role of ideal point availability and assortment in consumer choice. J. Consum. Res. **30**(2), 170–183 (2003)

Conati, C., Carenini, G., Hoque, E., Steichen, B., Toker, D.: Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. Comput. Graph. Forum **33**, 371–380 (2014)

Conati, C., Carenini, G., Toker, D., Lallé, S.: Towards user-adaptive information visualization. In: Proceedings of AAAI'15, pp. 4100–4106. AAAI Press (2015)

Dawson, J.F.: Moderation in management research: what, why, when, and how. J. Bus. Psychol. **29**(1), 1–19 (2014)

de Vries, P.W.: Trust in systems: effects of direct and indirect information. Technische Universiteit Eindhoven (2004)

Domik, G.O., Gutkauf, B.: User modeling for adaptive visualization systems. In: Proceedings and IEEE Conference on Visualization, 1994 (Visualization'94), pp. 217–223. IEEE (1994)

Ferwerda, B., Graus, M.: Predicting musical sophistication from music listening behaviors: a preliminary study. arXiv preprint arXiv:180807314 (2018)

Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2241–2246. ACM (2015)

Ferwerda, B., Tkalcic, M., Schedl, M.: Personality traits and music genre preferences: How music taste varies over age groups. In: 1st Workshop on Temporal Reasoning in Recommender Systems (RecTemp) at the 11th ACM Conference on Recommender Systems, Como, August 31, 2017, vol. 1922, pp. 16–20. ACM Digital Library (2017a)

Ferwerda, B., Tkalcic, M., Schedl, M.: Personality traits and music genres: What do people prefer to listen to? In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 285–288. ACM (2017b)

Fitzsimons, G.J., Lehmann, D.R.: Reactance to recommendations: when unsolicited advice yields contrary responses. Market. Sci. **23**(1), 82–94 (2004)

Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, pp. 54–89. Springer, New York (2007)

Gena, C., Brogi, R., Cena, F., Vernero, F.: The impact of rating scales on user's rating behavior. In: Proceedings of UMAP'11, pp. 123–134. Springer, New York (2011)

He, C., Parra, D., Verbert, K.: Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. Exp. Syst. Appl. **56**, 9–27 (2016)

Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp. 241–250. ACM (2000)

Hilliges, O., Holzer, P., Klüber, R., Butz, A.: Audioradar: A metaphorical visualization for the navigation of large music collections. In: International Symposium on Smart Graphics, pp. 82–92. Springer, New York (2006)

Hu, R., Pu, P.: Enhancing recommendation diversity with organization interfaces. In: Proceedings of IUI'11, pp. 347–350. ACM (2011)

Inoue, S., Aoyama, H., Nakata, K.: Cognitive analysis for knowledge modeling in air traffic control work. In: International Conference on Human–Computer Interaction, pp. 341–350. Springer, New York (2011)

Jin, Y., Seipp, K., Duval, E., Verbert, K.: Go with the flow: effects of transparency and user control on targeted advertising using flow charts. In: Proceedings of AVI'16, pp. 68–75. ACM (2016)

Jin, Y., Cardoso, B., Verbert, K.: How do different levels of user control affect cognitive load and acceptance of recommendations? In: Proceedings of IntRS Co-located with RecSys'17, CEUR-WS, pp. 35–42 (2017)

Jin, Y., Tintarev, N., Verbert, K.: Effects of individual traits on diversity-aware music recommender user interfaces. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 291–299 . ACM (2018a)

Jin, Y., Tintarev, N., Verbert, K.: Effects of personal characteristics on music recommender systems with different levels of controllability. In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 13–21 . ACM (2018b)

Judd, C.M., Kenny, D.A., McClelland, G.H.: Estimating and testing mediation and moderation in within-subject designs. Psychol. Methods **6**(2), 115 (2001)

Jugovac, M., Jannach, D., Lerche, L.: Efficient optimization of multiple recommendation quality factors according to individual user tendencies. Exp. Syst. Appl. **81**, 321–331 (2017)

Kamehkhosh, I., Jannach, D.: User perception of next-track music recommendations. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 113–121. ACM (2017)
Kim, H., Choo, J., Park, H., Endert, A.: Interaxis: steering scatterplot axes via observation-level interaction. IEEE TVCG'16 **22**(1), 131–140 (2016)
Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with mechanical Turk. In: Proceedings of CHI'08, pp. 453–456. ACM (2008)
Knees, P., Schedl, M., Pohle, T., Widmer, G.: Exploring music collections in virtual landscapes. IEEE Multimed. **14**(3), 46–54 (2007)
Knijnenburg, B.P., Reijmer, N.J., Willemsen, M.C.: Each to his own: how different users call for different interaction methods in recommender systems. In: Proceedings of RecSys'11, pp. 141–148. ACM (2011)
Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. UMUAI **22**(4–5), 441–504 (2012)
Komiak, S.Y., Benbasat, I.: The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q. **30**(4), 941–960 (2006)
Konstan, J.A., Riedl, J.: Recommender systems: from algorithms to user experience. UMUAI'12 **22**(1), 101–123 (2012)
Kramer, T.: The effect of measurement task transparency on preference construction and evaluations of personalized recommendations. J. Market. Res. **44**(2), 224–233 (2007)
Lallé, S., Conati, C., Carenini, G.: Impact of individual differences on user experience with a visualization interface for public engagement. In: Proceedings of UMAP'17, pp. 247–252. ACM (2017)
Lee, J.H., Kim, Y.S., Hubbles, C.: A look at the cloud from both sides now: an analysis of cloud music service usage. In: ISMIR, pp. 299–305 (2016)
Lekkas, Z., Tsianos, N., Germanakos, P., Mourlas, C., Samaras, G.: The effects of personality type in user-centered appraisal systems. In: International Conference on Human–Computer Interaction, pp. 388–396. Springer, New York (2011)
Mayer, R.E., Moreno, R.: Nine ways to reduce cognitive load in multimedia learning. Educ. Psychol. **38**(1), 43–52 (2003)
McCarthy, K., Salem, Y., Smyth, B.: Experience-based critiquing: Reusing critiquing experiences to improve conversational recommendation. In: Proceedings of ICCBR'10, pp. 480–494. Springer, New York (2010)
Millecamp, M., Htun, N.N., Jin, Y., Verbert, K.: Controlling spotify recommendations: effects of personal characteristics on music recommender user interfaces. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 101–109. ACM (2018)
Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In: Proceedings of the 2019 Conference on Intelligent User Interface, pp. 1–12. ACM (2019)
Müllensiefen, D., Gingras, B., Musil, J., Stewart, L.: The musicality of non-musicians: an index for assessing musical sophistication in the general population. PloS One **9**(2), e89642 (2014)
O'Donovan, J., Smyth, B., Gretarsson, B., Bostandjiev, S., Höllerer, T.: PeerChooser: visual interactive recommendation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1085–1088. ACM (2008)
Pampalk, E., Rauber, A., Merkl, D.: Content-based organization and visualization of music archives. In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 570–579. ACM (2002)
Parra, D., Brusilovsky, P.: User-controllable personalization: a case study with setfusion. IJHCS **78**, 43–67 (2015)
Perera, R.E.: Optimizing human-computer interaction for the electronic commerce environment. J. Electron Commer. Res. **1**(1), 23–44 (2000)
Perik, E., De Ruyter, B., Markopoulos, P., Eggen, B.: The sensitivities of user profile information in music recommender systems. In: Proceedings of Private, Security, Trust, pp. 137–141 (2004)
Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.P., Jonker, C.M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. UMUAI **22**(4–5), 357–397 (2012)
Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of RecSys'11, pp. 157–164. ACM (2011)
Randall, T., Terwiesch, C., Ulrich, K.T.: Research note-user design of customized products. Market. Sci. **26**(2), 268–280 (2007)

Saito, Y., Itoh, T.: Musicube: a visual music recommendation system featuring interactive evolutionary computing. In: Proceedings of VINCI'11, p. 5. ACM (2011)
Schaffer, J., Höllerer, T., O'Donovan, J.: Hypothetical recommendation: a study of interactive profile manipulation behavior for recommender systems. In: FLAIRS Conference, pp. 507–512 (2015)
Schedl, M., Zamani, H., Chen, C.W., Deldjoo, Y., Elahi, M.: Current challenges and visions in music recommender systems research. Int. J. Multimed. Inform. Retr. **7**(2), 95–116 (2018)
Sweller, J.: Cognitive load during problem solving: effects on learning. Cognit. Sci. **12**(2), 257–285 (1988)
Tintarev, N., Dennis, M., Masthoff, J.: Adapting recommendation diversity to openness to experience: a study of human behaviour. In: Proceedings of UMAP'13, pp. 190–202. Springer, New York (2013)
Tintarev, N., Masthoff, J.: Effects of individual differences in working memory on plan presentational choices. Front. Psychol. **7**, 1793 (2016). https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01793/bibTex
Tintarev, N.: Presenting diversity aware recommendations: making challenging news acceptable. In: Proceedings of FATREC'17 (2017)
Tkalcic, M., Kunaver, M., Tasic, J., Košir, A.: Personality based user similarity measure for a collaborative recommender system. In: Proceedings of the 5th Workshop on Emotion in Human–Computer Interaction-Real world challenges, pp. 30–37 (2009)
Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. In: First International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011), p. 106 (2011)
Toker, D., Conati, C., Carenini, G., Haraty, M.: Towards adaptive information visualization: on the influence of user characteristics. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 274–285. Springer, New York (2012)
Torrens, M., Hertzog, P., Arcos, J.L.: Visualizing and exploring personal music libraries. In: ISMIR (2004)
Tsai, C.H., Brusilovsky, P.: Enhancing recommendation diversity through a dual recommendation interface. In: Proceedings of RecSys IntRS'17, p. 10 (2017)
Verbert, K., Parra, D., Brusilovsky, P., Duval, E.: Visualising recommendations to support exploration, transparency and controllability. In: Proceedings of IUI'13, pp. 351–362. ACM (2013)
Viegas, F.B., Wattenberg, M., Van Ham, F., Kriss, J., McKeon, M.: Manyeyes: a site for visualization at internet scale. IEEE TVCG **13**(6), 1121–1128 (2007)
Wong, D., Faridani, S., Bitton, E., Hartmann, B., Goldberg, K.: The diversity donut: enabling participant control over the diversity of recommended responses. In: Proceedings of CHI EA'11, pp. 1471–1476. ACM (2011)
Zhang, X., Chignell, M.: Assessment of the effects of user characteristics on mental models of information retrieval systems. J. Assoc. Inform. Sci. Technol. **52**(6), 445–459 (2001)
Zhao, Q., Adomavicius, G., Harper, F.M., Willemsen, M., Konstan, J.A.: Toward better interactions in recommender systems: cycling and serpentining approaches for top-n item lists. In: Proceedings of CSCW'17, pp. 1444–1453. ACM (2017)
Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings WWW'05, pp. 22–32. ACM (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yucheng Jin is a researcher at the Human-Computer Interaction group of KU Leuven. He obtained his Ph.D. by focusing on transparency and user control of recommender systems in 2019 at KU Leuven. His research interests are recommender system user interfaces, personalization and user modeling, and natural user interfaces. He has published several papers at premier venues of recommender systems such as UMAP, RecSys, CIKM. Besides, he acts as a member of the program committee for the ACM Conference on Intelligent User Interfaces, and the ACM Recommender Systems Conference.

Nava Tintarev is an Assistant Professor and Technology Fellow at Delft University of Technology. She completed her PhD at the University of Aberdeen in 2010, and was previously an Assistant Professor at Bournemouth University (UK). She has published over 60 peer reviewed papers, primarily on improving the transparency in, and decision support for, recommender systems. She in on the management team for Delft Design for Values and an active member of Delft Data Science, where she is contributing to bringing together, integrating, and expanding existing practices and expertise for responsible data science. She acts as a senior member of the program committee for the ACM Conference on Intelligent User Interfaces, the ACM Recommender Systems Conference, and the Conference on User Modeling Adaptation and Personalization. She is serving as a Program Chair for the Intelligent User Interfaces conference 2020 in Cagliari.

Nyi Nyi Htun is a Postdoctoral researcher at KU Leuven working on a Mixed-initiative Recommender Systems project with Prof. Katrien Verbert. To date, his main research interests have been in Human-Computer Interaction, Interactive Information Retrieval & Recommender Systems, Digital Health and Business Information Systems. Nyi-Nyi finished his PhD in 2017 at Glasgow Caledonian University where he addressed the problems with non-uniform information access in Collaborative Information Retrieval. His PhD experiment outcomes have been published at venues such as SIGIR, CHIIR and IP&M. In 2017, Nyi-Nyi also worked at the Interactive and Trustworthy Technologies research group at Heriot-Watt University where he participated in a number of falls and rehabilitation projects funded by Digital Health & Care Institute, Scotland.

Katrien Verbert is an Associate Professor at the HCI research group of KU Leuven. She obtained a doctoral degree in Computer Science in 2008 at KU Leuven, Belgium. She was a post-doctoral researcher of the Research Foundation – Flanders (FWO) at KU Leuven. She was an Assistant Professor at TU Eindhoven, the Netherlands (2013–2014) and Vrije Universiteit Brussel, Belgium (2014–2015). Her research interests include visualization techniques, recommender systems, visual analytics, and digital humanities. She has been involved in several European and Flemish projects on these topics, including the EU ROLE, STELLAR, STELA, ABLE, LALA and BigDataGrapes projects. She is also involved in the organization of several conferences and workshops (general chair IUI 2021, program chair LAK 2020, general chair EC-TEL 2017, program chair EC-TEL 2016, workshop chair EDM 2015, program chair LAK 2013 and program co-chair of the EdRecSys, VISLA and XLA workshop series, DC chair IUI 2017, DC chair LAK 2019).