



Interactive Music Genre Exploration with Visualization and Mood Control

Yu Liang

Jheronimus Academy of Data Science
5211 DA 's-Hertogenbosch, The Netherlands
y.liang1@tue.nl

ABSTRACT

Recommender systems can be used to help users discover novel items and explore new tastes, for example in music genre exploration. However, little work has studied how to improve users' understandability and acceptance of the novel items as well as support users to explore a new domain. In this paper, we investigate how two different visualizations and mood control influence the perceived control, informativeness and understandability of a music genre exploration tool, and further to improve the helpfulness for new music genre exploration. Specifically, we compare a bar chart visualization used by earlier work to a contour plot which allows users to compare their musical preferences with both the recommended tracks as well as the new genre. Mood control is implemented with two sliders to set a preferred mood on energy and valence features (that correlate with psychological mood dimensions). In the online user study, mood control was manipulated between subjects, and the visualizations were compared within subjects. During the study ($N=102$), we measured users' subjective perceptions, experiences and the interactions with the system. Our results show that the contour plot visualization is perceived more helpful to explore new genres than the bar chart visualization, as the contour plot is perceived to be more informative and understandable. Users spent significantly more time and used the mood control more in the contour plot than in the bar chart visualization. Overall, our results show that the contour plot visualization combined with mood control serves as the most helpful way for new music genre exploration, because the mood control is easier to understand and use when made transparent via an informative visualization.

CCS CONCEPTS

- Human-centered computing → User studies; Information visualization; User interface design;
- Information systems → Recommender systems; Personalization.



This work is licensed under a Creative Commons Attribution-NoDerivs International 4.0 License.

IUI '21, April 14–17, 2021, College Station, TX, USA

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8017-1/21/04.

<https://doi.org/10.1145/3397481.3450700>

Martijn C. Willemse

Eindhoven University of Technology
5600 MB Eindhoven
Jheronimus Academy of Data Science
5211 DA 's-Hertogenbosch, The Netherlands
m.c.willemsen@tue.nl

KEYWORDS

recommender system, interactive design, visualization, exploration, music, mood, user study

ACM Reference Format:

Yi Liang and Martijn C. Willemse. 2021. Interactive Music Genre Exploration with Visualization and Mood Control. In *26th International Conference on Intelligent User Interfaces (IUI '21), April 14–17, 2021, College Station, TX, USA*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3397481.3450700>

1 INTRODUCTION

The purpose of a recommender system is not necessarily limited to help users find relevant items that fit only their long-term preferences [8]. Recently, more and more recommender systems have also been proposed to help users to discover novel items, such as novel music tracks [1, 12], or support them to explore new tastes [19, 34]. From the perspective of users, this type of recommender systems can serve as a way to fulfill their internal desires for seeking out new preferences [21]. From the recommender system perspective, supporting new preferences exploration could also keep users away [26] from the so called "filter bubble" issue [28], and encourage them to explore their blind spot [35].

However, little work has investigated the support users need during new preferences exploration and development. For example, recommendations from a new category/domain are usually a bit far from users' current preferences, thus making it hard for users to understand and relate the recommendations to their current preferences, which would subsequently discourage them to delve more into the new domain and demotivate them to explore. To fill this research gap, we look into a specific scenario where users are asked to explore an unfamiliar music genre, and investigate how to better support users to explore a new music genre, extending our existing genre exploration tool [19] with visualizations and user control.

Earlier work has suggested that visualizations with user control can serve a way to justify the recommendations [2, 22] as well as support users to explore the information space [17, 18]. Through providing justification, transparency, and certain level of user control, interactive visualizations have been shown to be useful to help users understand the generated recommendations and the recommendation process better, and further increase their satisfaction [7]. Additionally, in the context of music recommendation, mood is also an important factor to consider due to its strong relation with the

musical preferences [9, 33]. Mood-based interactive music recommender systems have shown to improve users' understanding of the recommendations as well as the music space [1].

In this work, we investigate how interactive design can support users to explore new preferences in the context of music genre exploration. Based on the existing music genre exploration system [19], we investigate how different types of visualizations as well as mood control would improve users' understanding of the recommendations from an unfamiliar new music genre, and support them to explore more. Specifically, we look into two visual explanation ways: a bar chart adapted from previous work [22], and a contour plot visualization which not only shows how the recommendations are related to the user's preferences but also to the new music genre. Additionally, we explore the benefits of mood control on two musical features related to mood (energy and valence). Because mood has been considered an important context that can influence users' musical preferences [33], we expect that the addition of mood control would improve the helpfulness of the exploration tool. Furthermore, we also explore the interactions between the two visualizations and mood control to see whether the addition of mood control is more beneficial for one or the other type of visualization. Overall, our two main research questions are:

RQ1: How do different types of visualization (bar charts or contour plots) influence the perceived helpfulness for new music genre exploration?

RQ2: How does mood control improve the perceived helpfulness for new music genre exploration?

2 RELATED WORK

2.1 Towards new tastes

The so-called "filter bubble" effects [28] have been discussed for a while, the main concern of which is that filtering mechanism might lead to too much personalized content and thus cause undesirable results, such as some extent of ideological polarization [6]. In contrast to the concerns, other studies found no empirical evidence that algorithms would result in filter bubbles [24, 37]. For instance, in the study by Nguyen et al. [27], users who followed the recommendations actually narrowed their content diversity more slowly than those who did not take the recommendations. While there has been no final agreement on whether the "filter bubble issue" should make us worried, there is a common consensus that recommender systems should not only recommend the most relevant items, but also help users to discover novel items and explore potential new tastes. From the perspective of users, new item discovery could also be one of their basic needs [11], and providing such support would be extremely helpful when they would like to proactively develop new preferences [5].

In this section, we briefly summarize research that was designed to support new tastes exploration, from raising the awareness of users to explore [26, 35] to supporting the exploration process towards the new tastes. Visualization can serve as a way to raise users' awareness towards exploring new tastes. Through visualizations, users are able to identify their blind spots [35] or understand the underlying filtering mechanisms [26], and be further motivated to explore the under-explored item space [16]. More recently, by visualizing users' profile inside an overview of the entire news domain

with an interactive tree-map, the interactive news recommender system (NewsViz) proposed by Kunkel et al. [18] provides users with a way to spot the over-represented news sources and raise their awareness towards potential bias.

With self-reflection [13] or the increasing awareness for exploration, people gradually step into the *action stage* of the behavioral change process [29], willing to take actions to explore new tastes (compared to their current preferences). In music domain, to help users to discover songs in a more active way, Kamalzadeh et al. proposed a novel interface called Tagflip [12], which allows users to specify the social tags of songs they would like to play next. The authors found that comparing to the Spotify interface, the proposed Tagflip was perceived more transparent and controllable. However, this approach was not personalized and did not take into account users' current preferences. To support exploration from their current preferences, Taramigkou et al. [34] proposed a music explorer which allows users to explore the latent genre space by taking a gradual path from their current preferences to their desired target genres and discover new tracks along the way.

Preference exploration might not only start from what users currently like, but also from a new category or domain that they are not familiar with, such as an unfamiliar music genre. Recently, we proposed a personalized music genre exploration tool to support users' exploration from a new music genre [19]. However, one drawback of this exploration approach was that it was not transparent enough for users to understand why certain recommendations were generated, which might influence the perceived acceptance of the recommendations and helpfulness for exploration, especially when the recommendations are from a new category or different from what users usually like.

2.2 Mood and musical preferences

The musical preferences of users can be influenced by mood [33], an affective state, which is more diffuse and long-lived than emotions [3]. There are in general two different models to characterize mood: dimensional and categorical models. The dimensional model can organize different moods in a simple but effective way with two dimensions, which allows for direct comparison of different moods. However, some researchers argued that a dimensional model cannot fully capture the emotional space expressed by music [1, 4] and might obscure important aspects of moods [36]. Categorical models describe moods with a rich set of distinct words [33], however the number of distinct words might be limited compared to what can be perceived by human [36]. In this work, we choose Russell's circumplex model [31], one of the most well-known dimensional models, to represent users' mood in music recommendations due to its simplicity and effectiveness. The Russell's circumplex model is represented by two dimensions: arousal and valence. In this study, we use energy and valence to represent the two dimensions, in which energy measures the intensity of tracks and valence measures the positiveness of tracks. Both features are retrieved from Spotify API¹.

¹<https://developer.spotify.com/documentation/web-api/>

2.3 Interactive visual design in music recommendation

Visualization can serve as a way to explain recommendations [22], while user control, as a part of interaction [10], has also shown to benefit many aspects of recommender systems, such as improving the perceived accuracy and satisfaction of the recommendations [10] as well as helping users better understand the recommendation process [2]. *MusicCube* [32] allows users to select preferred tunes from 2D scatter plots in an interactive way. In *MusicCube*, users could critique the suggested tunes either positively or negatively until they get a proper tune. They could also justify the recommendations by relating recommendations with both positively rated items and the negatively rated items in the 2D scatter plots. *TasteWeights* [2] allows users to adjust weights of their liked music and different sources with a slider. The system visualizes the link between a user's liked music, different context sources and the recommendations, thereby allows the user to justify a recommended song by its connection with the selected context sources and the liked music. The authors found that explaining a hybrid recommendation process through a visual user interface can improve user satisfaction. To investigate the effect of personal characteristics on different visual explanation techniques, Millecamp et al. [22, 23] designed an interactive visual explanation interface which enables users to adjust their preferences for each musical attribute with a slider. Users could also justify the recommendations by comparing the recommended tracks with their personal profile in the attribute space with bar charts and scatter plots. In both of the studies, they found that personal characteristics influence the way users interact with the system and the explanations needed. One relevant work that combines both visualization and mood control in music domain is *Moodplay* by Andjelkovic et al. [1]. *Moodplay* allows users to interactively explore artists in a 2D latent affective space (built upon a categorical emotion model). Users can use an avatar to navigate through the mood space with trails, which can keep track of their preferences along the way of exploration with a decay function. In their between-subjects study [1], the authors manipulated mood space visibility, availability of control over avatar and availability of trails. They found that their proposed visualization of mood improved users' understanding of the space and acceptance of the recommendations.

2.4 This paper: interactive visual design in music genre exploration

We extend the usage of interactive visual design to the context of music genre exploration. We compare two visualizations, a contour plot and a bar chart baseline (see Sec. 3.2 for details of the two visualization approaches), to investigate which visualization can better support users to explore new music genres. The bar chart visualization (adapted from the previous work [22]) is selected as the baseline because of its effectiveness for explanation, which allows users to compare the recommended tracks with their average preferences directly. Due to its straightforward representation, users might find it easy to grasp information from the bar chart. The more complex contour plot visualization is a combination of one scatter plot with two contour plots (for simplicity, we use the term *contour plot* to refer to the whole visualization). The scatter

plot is used to present the recommended items, as it is one of the most popular technique for showing multi-dimensional data [32]. Additionally, the contour plot is chosen to present both the user and genre profile in a clear representation of their feature distributions. Additionally, we also investigate how mood control can benefit the genre exploration process. First, user control can provide users with a feeling of control and means of exploration that would help them to better understand the recommendation process by relating the recommendations to the slider inputs [7], which could then benefit the helpfulness for new genre exploration. Secondly, as users' musical preferences are likely to be influenced by mood [33], the mood control slider would allow users to tailor recommendations based on their current mood [1].

3 SYSTEM DESIGN

3.1 Recommendation algorithms

The recommendation algorithm is adapted from the music genre exploration tool [19]. The system uses Spotify API to extract a user's top listened tracks and corresponding audio features of the tracks. To model a user's musical preferences, a Gaussian Mixture Model is built from the user's top tracks for each audio feature (*acousticness*, *danceability*, *energy*, *valence*, *liveness* and *speechiness*). The recommendations are generated in a content-based way with three different methods: a personalized algorithm, a non-personalized algorithm and a mixed method. The personalized algorithm calculates a track's recommendation score in each feature dimension based on the Gaussian mixture model; tracks with the highest scores averaged over all feature dimensions are returned as the recommendations. The non-personalized algorithm calculates the recommendation scores based on the prototypical tastes of the genre. In this work, the *mixed method* (which produces recommendations by combining the recommendation scores of the two above algorithms) is used for recommendation since it was shown to be more helpful for users to explore a genre by balancing personalized and prototypical tastes based on previous work. The dataset used for recommendation has 10 different music genres: avant-garde, blues, classical, country, electronic, folk, jazz, new-age, rap and r&b, and each genre has around 3300 tracks (we recommend readers to the refer to the original paper [19] for details).

3.2 Visualization techniques

In this section, we describe the two visualization techniques used in the study: a bar chart visualization and a contour plot visualization over energy and valence features. We chose energy and valence for visualization because first and foremost these two features are the best fit for mood control. Additionally, energy and valence are uniformly distributed along the entire range for all genres, as shown in Figure 1. While for most of the genres, the distributions of speechiness and liveness have an extremely high peak around zero. In other words, genre differences on these two features are relatively small. Similarly, acousticness is low for most of the genres (high peak around zero) except for a few genres that are highly acoustic (high peak around one), such as classical music. Therefore, energy and valence are the two most important features on which most genres differ, and thus can provide the most control to the user.

3.2.1 Bar chart visualization. The bar chart visualization, as shown in Figure 2 (d), is adapted from the work by Millecamp et al. [22] and serve as the baseline in the study. Each bar chart visualizes the energy and valence of a recommended track, as well as the users' average preferences over energy and valence (averaged over their top listened tracks).

3.2.2 Contour plot visualization. The contour plot visualization consists of one scatter plot and two contour plots, in which the recommended tracks are visualized with a scatter plot, and the user's top tracks and genre profile are visualized using contour plots, as shown by Figure 2 (c).

3.2.3 Compare the two visualizations. The contour plot allows users to compare their musical preferences, the typical tastes of a music genre and the recommended tracks within one single graph, while the bar charts visualization only allows users to compare how recommended tracks are related to their own preferences. Moreover, the contour plot visualization presents the density over the users' musical feature preferences in the 2D mood space, while the bar charts only shows users' average preferences over energy or valence. This makes the contour plot more informative, but also more complex and perhaps harder to understand for some users.

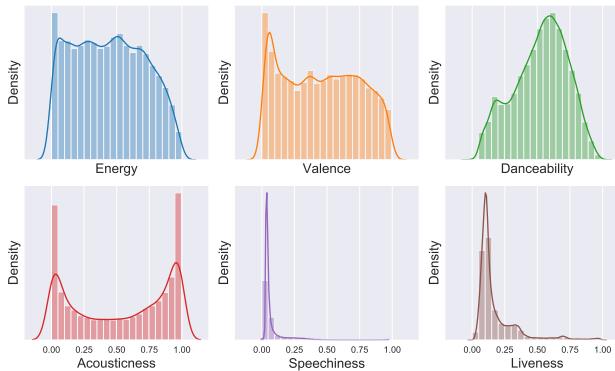


Figure 1: Audio feature distribution for the genre dataset

3.3 Mood control

The mood control sliders allow users to adjust recommendations within the top 300 recommended tracks returned by the recommendation algorithm. Specifically, users can adjust the recommendations with two sliders, one for energy and one for valence, in line with the 2D mood space, as shown in Figure 2 (b). Both sliders span the entire feature range of Spotify², but we constrain the adjustable range (i.e., the colored region on the slider) to the minimal and maximal feature value of the top 300 tracks used for the recommendations. Whenever a user moves the slider, the top 300 tracks will be re-ranked accordingly and a new top 10 tracks will be presented. The first step of the re-ranking is to calculate the Euclidean distance between the tracks and the slider values set by the user. Let (e_u, v_u) represents the energy and valence value set by

²According to the documentation of Spotify API, both energy and valence have a range of 0.0 to 1.0

the user, and (e_i, v_i) represents the the energy and valence value of track i . The Euclidean distance between the track and the sliders is calculated as:

$$d(i) = \sqrt{(e_u - e_i)^2 + (v_u - v_i)^2} \quad (1)$$

The final recommendation score for a track is calculated by averaging the reverse ranking of the track in the original recommendation list after user control. Specifically, let $r_{original}(i)$ represents the ranking of track i in the original recommendation list, $r_d(i)$ represents the ranking of track i based on the distance score (smaller distance means higher rank), $|R|$ denotes the length of the recommendation list for re-ranking (set to 300 in our study). The recommendation score is calculated as:

$$score = 0.5 \times (|R| - r_{original}(i) + 1) + 0.5 \times (|R| - r_d(i) + 1) \quad (2)$$

3.4 User interface

Figure 2 presents two example recommendation interfaces. The interface always presents the top 10 recommended tracks presented in a playlist style (a). Based on the conditions (see Sec. 4.2) the interface also contains other components: (b) two mood sliders for adjusting valence and energy, (c) a contour plot, and (d) bar charts. When hovering over the recommended tracks in (a), the corresponding track would be highlighted in the visualization (c) or (d), and vice versa. The explanations of the interface components are also available when hovering over the question mark at the top right corner of the mood control or visualization components. At the bottom of the display, users can save the playlist to their Spotify profile.

3.5 Measurements and expectations

Our overall goal is to investigate how the two different visualizations (RQ1) as well as mood control (RQ2) would influence the perceived helpfulness for new music genre exploration. During the study, we measure users' perceived informativeness of the visualizations, perceived control over the system, perceived understandability of the recommendations and helpfulness of the recommendations to explore a new genre. Additionally, we also capture several user interactions, such as slider usage, users' interaction time with the recommended tracks, and overall time spent within the system as objective measures.

Overall, we expect the contour plot to be more helpful than the bar chart (RQ1). Comparing to the bar chart, the contour plot could be perceived more informative as it allows users to inspect the relations between the new genre, the recommended songs, and the current preferences in a more intuitive way, by visualizing the trade-off the recommendation algorithm is making to bridge users' current preferences and the chosen new music genre. Additionally, the contour plot could be more informative in a sense that it allows users to see the distribution of their feature preferences—for example, where the strongest preferences are—rather than the average preferences as in the bar chart. The increased informativeness would then improve users' understanding of the recommendations, which would further improve the perceived helpfulness of the recommendations for new genre exploration. However contour plots might be a bit harder to process for users, due to their complexity,

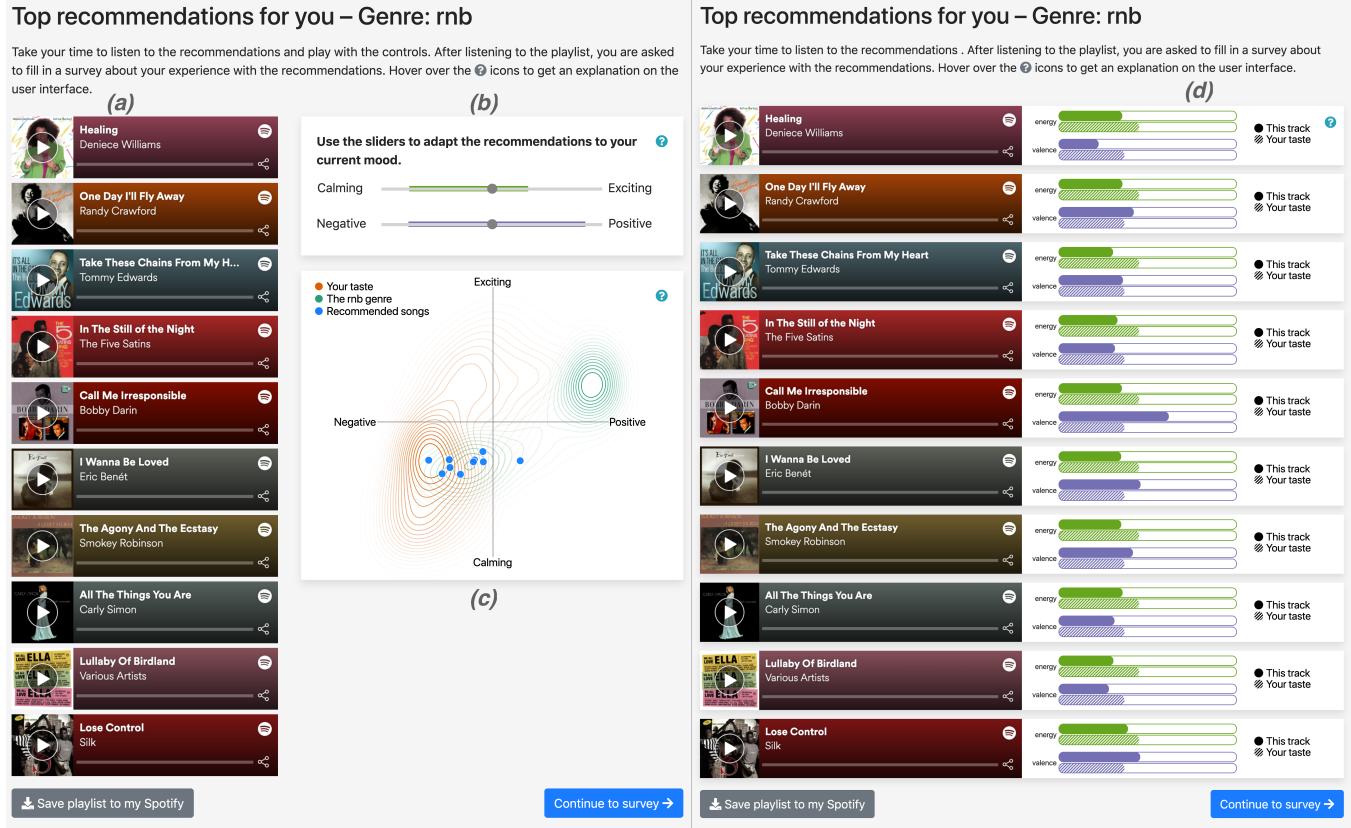


Figure 2: The recommendation interface components: (a) the top 10 recommended tracks, (b) the two mood sliders for adjusting preferred energy (calming–exciting) and valence (negative–positive), (c) the contour plot, and (d) the bar charts.

which might affect the perceived informativeness for some users. We also expect that the addition of mood control would increase the helpfulness for new genre exploration (RQ2). When provided with mood control, we expect that users would perceive the system to be more controllable, as they are able to adjust the recommendations based on the two mood features (energy and valence). Furthermore, the addition of mood control would also allow users to justify the change of the recommendation output through visualizations, thus help them to understand the recommendation process better, as also indicated by some earlier work [7]. The increased controllability would then improve users' understandability of the recommendations, and at last positively influence their perceived helpfulness of the recommendations for new genre exploration.

Importantly, the question is whether mood control in the bar chart visualization is as effective as in the contour plot visualization. We expect that mood control would be helpful in both visualizations. In the bar chart visualization, the mapping is direct: users can directly observe the result of the mood control on the individual recommended track. We expect, however, that mood control would be even more helpful in the contour plot visualization, as the contour plot can provide users with an overall picture of how the recommendations relate to both the user profile and the new genre. Therefore, we expect a positive interaction effect between the contour plot and mood control on user perceptions, showing that mood

control is more beneficial when combined with the contour plot than with the bar chart, potentially overcoming the complexity of the contour plot. Moreover, given that we expect mood control would be more helpful in the contour plot condition, we also expect users to interact differently with the system, for example they might use the sliders more, show longer interaction times, or saved playlists more with the contour plot than with the bar charts.

4 STUDY DESIGN

4.1 Participants

Most of the participants of the study were recruited from the JF Schouten participant database of Eindhoven University of Technology, with the rest recruited via convenience sampling. They were invited for the online study by email or instant messages and informed about the basic procedures of the study. All participants were required to have an active Spotify account. Furthermore, they should not have any visual or hearing impairment. As reimbursement, we raffled 20 euros for every five participants using a lottery (expected value per person = 4 euros).

4.2 Mixed factorial design

A 2x2 mixed factorial design was conducted to investigate the effect of different visualizations and mood control on the system for

music genre exploration. In the study, mood control was manipulated between-subjects and visualizations were manipulated within-subjects. During the study, participants were randomly assigned to the conditions in which mood control was either available for genre exploration or not. Participants were required to go through two phases, in which different visualizations were shown (the order of the two visualizations was counterbalanced) and after each visualization they filled in a questionnaire measuring their experience with that visualization.

4.3 Study procedure

At the beginning of the study, participants first needed to agree with the informed consent which informed them about the study procedure, asked for permission to access their Spotify data and described how their data would be used and stored anonymously. They were then asked to log in with their Spotify account and redirected to the Musical Sophistication Index survey [25], which measured their musical expertise in active engagement and emotional engagement. After filling in the survey, participants moved to the next part in which they got two recommendations phases for the two visualizations. In each phase, they started at the genre selection page to select a genre to explore. Next, they were redirected to the recommendation page (see Figure 2 for the interface displays) based on the study conditions as discussed in Section 4.2). Specifically, one of the visualizations (a contour plot or bar charts) was presented. Moreover, in the condition with mood control, users were allowed to adjust the presented recommendations with the two mood control sliders. Upon arriving the recommendation page, participants first received a walk-through, which explained the interface elements. In the task instruction, participants were asked to take their time and listen to the recommendations [and test the mood control if mood control was available]. Participants could also save the playlist to their profile. After finishing the exploration and listening to the recommendations, participants were asked to fill in a post-task questionnaire (see Table 1) about their experience with the exploration tool. Next, participants started with their second phase with the same mood control setting but with a different visualization (as visualization was measured within-subjects). During the genre selection stage, they were asked to select a different genre to explore. Note that the already selected genre in the first phase was greyed out to prevent participants to receive the same recommendations. After the second phase, they again received a post-questionnaire and were thanked for the participation.

4.4 Subjective and objective measures

Following the user centric framework by Knijnenburg et al. [15], we measured several subjective system aspects (SSA) or user experience (EXP) with a post-task questionnaire (see Table 1). The measured SSA are (1) perceived control (2) perceived informativeness, (3) perceived accuracy, and (4) perceived understandability, and the measured EXP is (5) perceived helpfulness. We adapted questions for perceived control from Pu et al. [30], perceived accuracy from Knijnenburg et al. [15], perceived helpfulness from [19], and constructed new items for the two additional factors: perceived informativeness and understandability. In addition to SSA/EXP, several user interaction (INT) behaviors with the system were also

measured. The measured INT are (1) user interaction time with the system (2) track interactions (3) slider interactions and (4) playlist saving behavior (whether the generated playlist is saved the user's Spotify). Moreover, for each participant, we also measured the musical sophistication (Personal Characteristic) on active engagement (MSAE) and emotional engagement (MSE) based on the Goldsmith Musical Sophistication Index [25].

5 RESULTS

5.1 Descriptive statistics

We ran the experiment in June 2019 and received 102 valid responses (four participants were removed from the analysis because they only interacted with the system very shortly). The mean age of the participants was 23.9 ($SD = 7.0$; range from 18 – 70), with 60 males and 44 females. Figure 3 presents the genres that participants chose to explore.

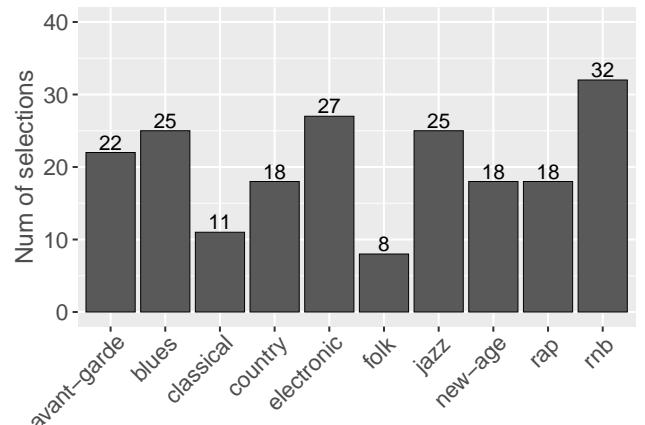


Figure 3: Genre selection frequencies indicating how often users choose to explore each genre

5.2 Interaction data

In this section, we discuss the effects of mood control and the two visualizations on the measured interaction behaviors to see whether users used the system differently in different conditions.

5.2.1 Log interaction time. A multilevel linear regression was applied to test the effects of control, different visualizations and recommendation phase on the log transformed interaction time. We found that users interacted with the system shorter in the second phase (i.e. the second time they used the system with the other visualization) than in the first phase ($\beta = -0.21, se = 0.05, p < .001$) regardless of the conditions. Additionally, users interacted longer with the contour plot than bar charts ($\beta = 0.11, se = 0.05, p < .05$), as also shown in Figure 4(a). Interestingly, users with mood control did not interact with the system longer than those without (differences are insignificant).

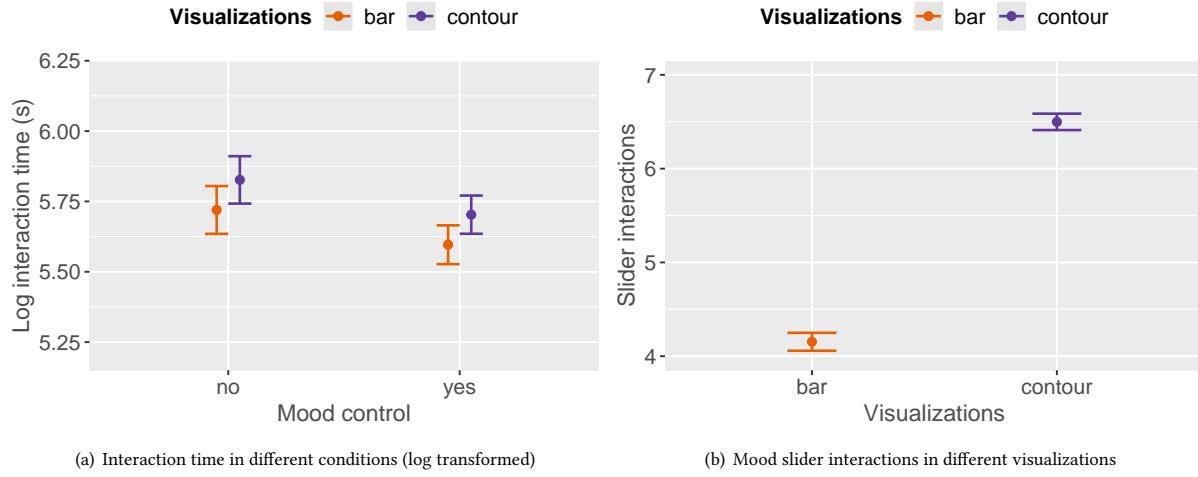


Figure 4: Estimated effects of mood control and visualizations on user interaction behaviors (error bar indicates the standard error)

5.2.2 Track interactions. We applied a multilevel Poisson regression (Poisson regression was used because interactions are count data) to study the effects of different factors on track interactions. Similar to the log interaction time, users interacted with the recommended tracks slightly less in the second phase than in the first phase ($\beta^3 = -0.06$, $se = 0.02$, $p < .05$). However, other than this effect, no further effects of visualization or mood control on track interactions were found.

5.2.3 Slider interactions in the mood control condition. Figure 4(b) shows the effect of different visualizations on user slider usage. Note that the use of sliders usage is only available in the mood control condition. We observe that users indeed used the slider much more with the contour plot than with the bar charts ($\beta = 0.41$, $se = 0.08$, $p < .001$ according to the multilevel Poisson regression). However, the slider usage did not differ between experimental phases, despite the observed reduction in interactions for the second phase.

5.2.4 Playlist saved or not. We observe no effect of mood control, different visualizations or experimental phase on users' playlist saving behaviors to their Spotify account. On average, users saved the playlist 14.2% of the time, regardless of the conditions or the experimental phase.

5.3 Structure Equation Model

A Structure Equation Model (SEM) was fitted to study the relation between the experiment conditions (OSA), the measured subjective constructs (SSA and EXP), user interaction with the system (INT), and user musical sophistication (PC) [15]. Before fitting the SEM, we first checked the validity of the question items and the subjective constructs from the post-task questionnaire (see Table 1) with Confirmatory Factor Analysis (CFA). The CFA was performed to test the

quality of the question items for each subjective construct by treating all answers as ordinal. Based on the modification indices, we removed the construct perceived accuracy because several items of other constructs (i.e., helpfulness and understandability) also loaded high on accuracy. The average variance extracted (AVE) of the rest four subjective constructs is higher than the recommended value 0.5, indicating that the convergent validity the constructs holds [14]. Cronbach's alphas indicate good internal consistency of each construct. The CFA was then expanded into a SEM. When constructing the SEM, we checked the effect of the experimental phase and the presentation order of visualizations, but none of the effects were significant so we excluded them from the final SEM. Table 1 presents the question items with corresponding SEM coefficients. The fitted SEM is presented in Figure 5. The SEM has an adequate fit: $\chi^2(114) = 141.910$, $p < .05$, $RMSEA = .035$ (90%CI = [.008, .052]), $CFI = 0.988$, $TLI = 0.986$.

5.4 SEM Results

5.4.1 Perceived informativeness and control. The left part of the SEM in Figure 5 shows the effects of conditions on informativeness and perceived control. Consistent with our expectations, the results show that users perceived the contour plot to be much more informative than the bar charts, and the perceived control indeed is higher when users are provided with mood control. There is no direct effect of the contour plot visualization on perceived control, but there is a strong indirect effect via perceived informativeness ($\beta = .332$, $se = .057$, $p < .001$) showing that the contour plot is perceived more controllable than the bar chart as it is perceived to be more informative. We also observe a marginal significant interaction effect of contour plot and mood control on perceived informativeness, suggesting that with control, the contour plot becomes even more informative. Moreover the same interaction has a marginal significant indirect effect on perceived control, via perceived informativeness ($\beta = .212$, $se = .109$, $p = .052$). One possible

³Here, β is the coefficient of the Poisson regression. This also applies to the coefficient in slider interactions

Table 1: Post-task questionnaire: factor scores for each item of the subjective constructs in the SEM model. AVE=Average Variance Extracted for this construct by the items, Alpha = Cronbach's alpha indicating the internal consistency of the construct

Subjective constructs	Question item	SEM coef.
Perceived control	I feel in control of modifying the recommendations.	0.894
AVE=0.617	The recommender allows only limited control to modify the recommendations.	-0.368
Alpha=0.78	I found it easy to modify the recommendations in the recommender.	0.672
Informativeness	The {contour chart is, bar charts are} informative.	0.895
AVE=0.729	The {contour chart has, bar charts have} no real benefit to me.	-0.861
Alpha=0.88	I can find better recommendations using the {contour chart, bar charts}	0.753
Helpfulness	The playlist supports me in getting to know the new genre.	0.705
AVE=0.644	The playlist motivates me to more delve into the new genre.	0.493
Alpha=0.83	The playlist is useful in exploring a new genre.	0.672
Understandability	I understand how the recommended songs relate to my musical taste.	0.763
AVE=0.701	It is easy to grasp why I received these recommended songs.	0.771
Alpha=0.85	The recommendation process is clear to me.	0.593
Accuracy(dropped)	The recommended songs fitted my preference.	
	I would give the recommended songs a high rating.	
	I liked each of the recommended songs provided by the system.	

(1) All questions are answered on a 7-point Likert scale. (2) All factor loadings are significant $p < .001$.

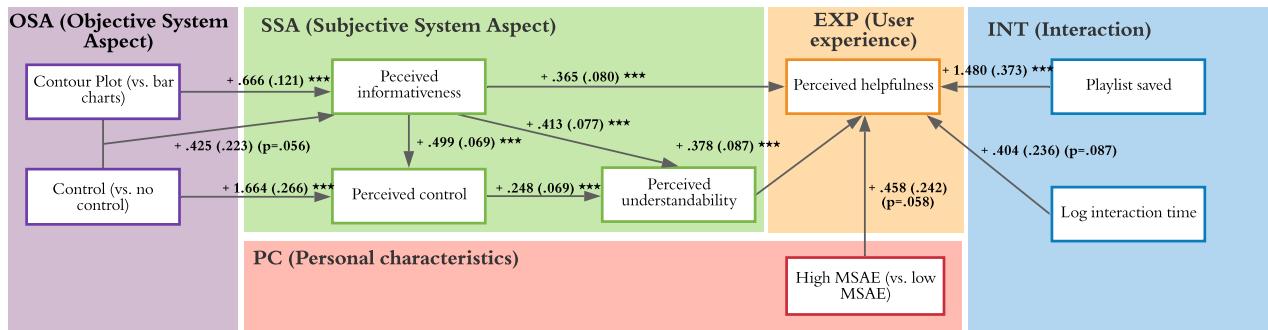


Figure 5: Structural Equation Model. Significant levels: * $p < .001$; ** $p < .01$; * $p < .05$. Numbers on the arrows indicate standardized coefficients and standard error (in brackets) of the effect**

explanation for this is that the contour plot could make it easier for users to track the change of recommendations by simply comparing the current recommendation position with previous recommendation position (retrieved from their visual working memory [20]). In summary, combining contour plot with mood control has the highest impact on informativeness and perceived control, which also provides a reason for why users are making use of the slider more in the contour plot condition (see also Sec. 5.2.3).

5.4.2 Understandability. As expected, perceived understandability is positively affected by both informativeness and perceived control. This suggests that recommendations from a new music genre are perceived more understandable when the visualization is perceived more informative and the recommendations are perceived more controllable.

We observe no direct effects of different conditions on understandability, instead all effects are mediated by perceived control and informativeness. Looking at the indirect effect of the visualization conditions on understandability, we observe that the contour plot is perceived more understandable than bar charts (total (indirect) effect is $\beta = .358, se = .067, p < .001$) through two pathways. The stronger pathway is via perceived informativeness, which indicates that users perceived the contour plot to be more understandable because it is more informative ($\beta = .275, se = .062, p < .001$). One possible explanation for this is, comparing to the bar chart visualization, the contour plot allows users to compare their profile, genre profile and recommended tracks in one single graph, which could better help them to understand the relation among the three. The second, weaker, pathway is via both perceived informativeness and perceived control ($\beta = .083, se = .028, p = .003$). The contour plot is perceived to be more informative, thus more controllable,

and the improved controllability positively influence the perceived understandability.

Looking at the indirect effect of mood control on understandability ($\beta = .491, se = .159, p = .002$), we find that the strongest pathway goes via perceived control ($\beta = .413, se = .135, p = .002$). The results indicate that users perceive the system with mood control to be more understandable because it is perceived to be more controllable. Moreover, the indirect interaction effect of mood control and the contour plot on understandability via perceived control seems to suggest that, users perceive the system to be somewhat more understandable when mood control is combined with the contour plot rather than with the bar chart visualization. However, this effect is only marginally significant ($\beta = .229, se = .118, p = 0.052$) and much weaker than the main effect of the perceived control.

5.4.3 Helpfulness. Finally, we inspect the two research questions regarding how different visualizations (RQ1) and mood control (RQ2) would influence the perceived helpfulness for new genre exploration. Firstly, the SEM shows that the perceived helpfulness is positively influenced by informativeness and understandability. This indicates that recommendations from a new genre are perceived more helpful for exploration when they are more understandable or presented in a more informative way through visualization. However, no direct effect of perceived control on helpfulness is found. The effect on helpfulness is completely mediated by understandability.

Again, we do not observe any direct effects of different conditions on helpfulness. All effects are mediated by the other perceptual measures (informativeness, control and understandability). The overall (indirect) effect of the contour plot on helpfulness is significant: users perceive the contour plot visualization to be more helpful than the bar chart visualization ($\beta = .378, se = .082, p < .001$). The largest pathway of the contour plot on helpfulness is via informativeness ($\beta = .243, se = .066, p < .001$). In other words, users perceive the contour plot to be more helpful than the bar charts because the contour is more informative. A potential explanation for this is, comparing to the bar charts, the contour plot allows users to have an overview of the new genre, while the overview could help users to understand the item space [18] of the new genre and therefore encourage them to explore more. In addition, a smaller pathway is via informativeness and understandability ($\beta = .103, se = .033, p = .002$), and the smallest pathway is via informativeness, perceived control and perceived understandability ($\beta = .031, se = .013, p = .019$). Comparing to bar charts, the contour plot visualization is perceived more informative. The increased informativeness further increases the perceived control. Both the increased informativeness and the improved controllability then positively influence users' understanding of the recommendations, which subsequently improves the perceived helpfulness for new genre exploration.

The overall (indirect) effect of mood control on helpfulness is much smaller. Users seem to perceive the system with mood control more helpful than without, however, the total indirect effect is only marginal significant ($\beta = .238, se = .123, p = .053$). The only significant pathway is via perceived control and understandability ($\beta = .156, se = .065, p = .016$). This indicates that the addition of mood control improves the perceived control, which then increase

users' understandability of the recommendations, and thus makes the system more helpful for users to explore a new genre. All other pathways are very weak and insignificant. Interestingly, although the mood control alone has limited impact on helpfulness, the interaction of the mood control with contour plot revealed a significant indirect effect on helpfulness ($\beta = 0.242, se = 0.123, p = .049$). The most important pathway is via perceived informativeness: the increased informativeness of the contour plot combined with mood control further improves the perceived helpfulness for new genre exploration ($\beta = .155, se = .080, p = .053$). Other pathways via informativeness and understandability ($\beta = .066, se = 0.039, p = .091$) or via informativeness, control and understandability ($\beta = .020, se = 0.013, p = .115$) were much weaker and insignificant.

Summarizing, both the contour plot (compared to the bar chart visualization) and mood control provide users with more helpfulness to explore a new genre. The strongest increase in helpfulness is observed when the contour plot is combined with mood control as shown by the significant indirect interaction effect. This suggests that the addition of mood control indeed shows its benefits in terms of helpfulness when the recommendations are visualized in a more informative way using the contour plot. Looking at the personal characteristics, users with higher MSAE (Musical Sophistication on Active Engagement) perceive the recommendations to be more helpful than users with lower MSAE based on a median split. However, this effect is only marginally significant ($p = .058$) and there is no effect of MSAE on any other measures, despite earlier research has shown that higher MSAE can lead to a higher acceptance of recommendations [22] and users with higher MSAE perceived recommendations to be more helpful [19]. Finally, we observe that users' interactions with the system serve as an indicator to the perceived helpfulness. Specifically, users who saved the recommended playlist to their Spotify account perceived the recommendations to be more helpful than users who did not. Additionally, the interaction time (log transformed) also marginally affects the perceived helpfulness.

6 CONCLUSION

To conclude, we observe that the contour plot is perceived more helpful for new genre exploration than the bar chart visualization. The contour plot visualization provides users with an overview of their current preferences and the new genre, and describes how the recommendations are relate to both the new genre and their preferences, which is more informative and thus more helpful for understanding or getting to know the recommendations as well as the item space of the new genre. With the increased informativeness, users would also understand the recommendations better. The improved understanding towards the recommendations would then make the recommendations more helpful for users as a start point to explore a new genre. Comparing to earlier work that mostly focus on explaining the recommendations based on users' preferences [22, 32], the contour plot used in the study also allows users to compare the recommendations with the item space of the new genre, and seems to serve a way for users to understand how the recommendations bridge their preferences with the new genre. Interestingly, the addition of mood control itself does not make the system more helpful for exploration, but mood control paired with

the contour plot indeed benefits the perceived helpfulness for new genre exploration most. When provided with mood control, users perceive the contour plot to be more controllable than bar charts. The improved controllability can then benefit users' understanding of the whole recommendation process. Given that users also used the slider more with the contour plot, the mood control seems to not only serve as a function for users to adjust recommendations to their desired mood, but also when paired with contour plot, helps users to understand the recommendation process more, which further encourages them to explore more. In other words, control is much effective when accompanied by a visualization that is informative, showing that transparency and interactivity of recommender systems are inherently interconnected aspects [7].

This study also has some limitations. The results might not be easily generalized to the general public as most of our participants were university students. The contour plot might be easier for the university students to understand than for regular users, as university students are often well educated in understanding plots and visualizations. For our future work, we would like to extend the study and recruit more people with different occupations and educational levels, to see whether the benefits of the contour plot paired with mood control of the music exploration tool will still hold on a more general user sample.

ACKNOWLEDGMENTS

We would like to thank Thijs Meeuwisse and Joris Hilberink for programming the study and collecting the data.

REFERENCES

- [1] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2019. Moodplay: Interactive music recommendation based on Artists' mood similarity. *International Journal of Human-Computer Studies* 121 (jan 2019), 142–159. <https://doi.org/10.1016/j.ijhcs.2018.04.004>
- [2] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: A Visual Interactive Hybrid Recommender System. In *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin, Ireland) (*RecSys '12*). Association for Computing Machinery, New York, NY, USA, 35–42. <https://doi.org/10.1145/2365952.2365964>
- [3] John T Cacioppo, Gary G Berntson, Jeff T Larsen, Kirsten M Poehlmann, Tiffany A Ito, et al. 2000. The psychophysiology of emotion. *Handbook of emotions* 2 (01 2000), 173–191.
- [4] Geoffrey L. Collier. 2007. Beyond valence and activity in the emotional connotations of music. *Psychology of Music* 35, 1 (jan 2007), 110–131. <https://doi.org/10.1177/0305735607068890>
- [5] Michael D. Ekstrand and Martijn C. Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (*RecSys '16*). Association for Computing Machinery, New York, NY, USA, 221–224. <https://doi.org/10.1145/2959100.2959179>
- [6] Seth Flaxman, Sharad Goel, and Justin M. Rao. 2016. Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320. <https://doi.org/10.1093/poq/nfw006>
- [7] Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications* 56 (sep 2016), 9–27. <https://doi.org/10.1016/j.eswa.2016.02.013>
- [8] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (*RecSys '16*). Association for Computing Machinery, New York, NY, USA, 7–10. <https://doi.org/10.1145/2959100.2959186>
- [9] Joris H. Janssen, Egon L. van den Broek, and Joyce H. D. M. Westerink. 2012. Tune in to your emotions: a robust personalized affective music player. *User Modeling and User-Adapted Interaction* 22, 3 (jul 2012), 255–279. <https://doi.org/10.1007/s11257-011-9107-7>
- [10] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of Personal Characteristics on Music Recommender Systems with Different Levels of Controllability. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) (*RecSys '18*). Association for Computing Machinery, New York, NY, USA, 13–21. <https://doi.org/10.1145/3240323.3240358>
- [11] Mohsen Kamalzadeh, Dominikus Baur, and Torsten Möller. 2012. A survey on music listening and management behaviours. *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, Porto, Portugal (01 2012).
- [12] Mohsen Kamalzadeh, Christoph Kralj, Torsten Möller, and Michael Sedlmair. 2016. TagFlip: Active Mobile Music Discovery with Social Tags. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) (*IUI '16*). Association for Computing Machinery, New York, NY, USA, 19–30. <https://doi.org/10.1145/2856767.2856780>
- [13] Bart P. Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender Systems for Self-Actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (*RecSys '16*). Association for Computing Machinery, New York, NY, USA, 11–14. <https://doi.org/10.1145/2959100.2959189>
- [14] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating Recommender Systems with User Experiments. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, 309–352. https://doi.org/10.1007/978-1-4899-7637-6_9
- [15] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (oct 2012), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- [16] Jayachithra Kumar and Nava Tintarev. 2018. Using visualizations to encourage blind-spot exploration. *CEUR Workshop Proceedings* 2225 (2018), 53–60.
- [17] Johannes Kunkel, Benedikt Loepf, and Jürgen Ziegler. 2017. A 3D Item Space Visualization for Presenting and Manipulating User Preferences in Collaborative Filtering. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) (*IUI '17*). Association for Computing Machinery, New York, NY, USA, 3–15. <https://doi.org/10.1145/3025171.3025189>
- [18] Johannes Kunkel, Claudia Schwenger, and Jürgen Ziegler. 2020. NewsViz: Depicting and Controlling Preference Profiles Using Interactive Treemaps in News Recommender Systems. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) (*UMAP '20*). Association for Computing Machinery, New York, NY, USA, 126–135. <https://doi.org/10.1145/3340631.3394869>
- [19] Yu Liang and Martijn C. Willemsen. 2019. Personalized Recommendations for Music Genre Exploration. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) (*UMAP '19*). Association for Computing Machinery, New York, NY, USA, 276–284. <https://doi.org/10.1145/3320435.3320455>
- [20] Steven J. Luck and Edward K. Vogel. 2013. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences* 17, 8 (aug 2013), 391–400. <https://doi.org/10.1016/j.tics.2013.06.006>
- [21] Leigh McAlister and Edgar Pessemier. 1982. Variety Seeking Behavior: An Interdisciplinary Review. *Journal of Consumer Research* 9, 3 (dec 1982), 311. <https://doi.org/10.1086/208926>
- [22] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To Explain or Not to Explain: The Effects of Personal Characteristics When Explaining Music Recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) (*IUI '19*). Association for Computing Machinery, New York, NY, USA, 397–407. <https://doi.org/10.1145/3301275.3302313>
- [23] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2020. What's in a User? Towards Personalising Transparency for Music Recommender Interfaces. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) (*UMAP '20*). Association for Computing Machinery, New York, NY, USA, 173–182. <https://doi.org/10.1145/3340631.3394844>
- [24] Judith Möller, Damian Trilling, Natali Helberger, and Bram van Es. 2018. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society* 21, 7 (jul 2018), 959–977. <https://doi.org/10.1080/1369118X.2018.1444076>
- [25] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE* 9, 2 (feb 2014), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- [26] Sayooran Nagulendра and Julita Vassileva. 2014. Understanding and Controlling the Filter Bubble through Interactive Visualization: A User Study. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (Santiago, Chile) (*HT '14*). Association for Computing Machinery, New York, NY, USA, 107–115. <https://doi.org/10.1145/2631775.2631811>
- [27] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) (*WWW '14*). Association for Computing Machinery, New York, NY, USA, 677–686. <https://doi.org/10.1145/2566486.2568012>

- [28] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.
- [29] James O. Prochaska and Wayne F. Velicer. 1997. The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion* 12, 1 (sep 1997), 38–48. <https://doi.org/10.4278/0890-1171-12.1.38>
- [30] Pearl Pu, Li Chen, and Rong Hu. 2011. A User-Centric Evaluation Framework for Recommender Systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) (RecSys '11). Association for Computing Machinery, New York, NY, USA, 157–164. <https://doi.org/10.1145/2043932.2043962>
- [31] James Russell. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39 (12 1980), 1161–1178. <https://doi.org/10.1037/h0077714>
- [32] Yuri Saito and Takayuki Itoh. 2011. MusiCube: A Visual Music Recommendation System Featuring Interactive Evolutionary Computing. In *Proceedings of the 2011 Visual Information Communication - International Symposium* (Hong Kong, China) (VINCI '11). Association for Computing Machinery, New York, NY, USA, Article 5, 6 pages. <https://doi.org/10.1145/2016656.2016661>
- [33] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7, 2 (jun 2018), 95–116. <https://doi.org/10.1007/s13735-018-0154-2>
- [34] Maria Taramigkou, Eftimios Bothos, Konstantinos Christidis, Dimitris Apostolou, and Gregoris Mentzas. 2013. Escape the Bubble: Guided Exploration of Music Preferences for Serendipity and Novelty. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong, China) (RecSys '13). Association for Computing Machinery, New York, NY, USA, 335–338. <https://doi.org/10.1145/2507157.2507223>
- [35] Nava Tintarev, Shahin Rostami, and Barry Smyth. 2018. Knowing the unknown. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing - SAC '18*. ACM Press, New York, New York, USA, 1396–1399. <https://doi.org/10.1145/3167132.3167419>
- [36] Yi-Hsuan Yang and Homer H. Chen. 2012. Machine Recognition of Music Emotion. *ACM Transactions on Intelligent Systems and Technology* 3, 3 (may 2012), 1–30. <https://doi.org/10.1145/2168752.2168754>
- [37] Frederik J. Zuiderveen Borgesius, Damian Trilling, Judith Möller, Balázs Bodó, Claes H. de Vreese, and Natali Helberger. 2016. Should we worry about filter bubbles? *Internet Policy Review* 5, 1 (mar 2016), 1–16. <https://doi.org/10.14763/2016.1.401>