ContextPlay: Evaluating User Control for Context-Aware Music Recommendation

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

Nyi Nyi Htun
Department of Computer Science, KU Leuven
Leuven, Belgium
nyinyi.htun@cs.kuleuven.be

Nava Tintarev
Faculty of Engineering, Mathematics and Computer Science, Delft University of Technology
Delft, The Netherlands
n.tintarev@tudelft.nl

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

ABSTRACT

Music preferences are likely to depend on contextual characteristics such as location and activity. However, most recommender systems do not allow users to adapt recommendations to their current context. We therefore built ContextPlay, a context-aware music recommender that enables user control for both contextual characteristics and music preferences. By conducting a mixed-design study (N=114) with four typical scenarios of music listening, we investigate the effect of controlling contextual characteristics in a music recommender system on four aspects: *perceived quality, diversity, effectiveness, and cognitive load.* Compared to our baseline which only allows to specify music preferences, having additional control for context leads to higher perceived quality and does not increase cognitive load. We also find that the contexts of *mood, weather*, and *location* tend to influence user perception of the system. Moreover, we found that users are more likely to modify contexts and their profile during relaxing activities.

CCS CONCEPTS

• **Information systems → Personalization**; **Recommender systems**; • **Human-centered computing** → *User studies*; Empirical studies in HCI.

KEYWORDS

context-aware recommendation, user control, music recommendation

ACM Reference Format:
Yucheng Jin, Nyi Nyi Htun, Nava Tintarev, and Katrien Verbert. 2019. ContextPlay: Evaluating User Control for Context-Aware Music Recommendation. In *27th Conference on User Modeling, Adaptation and Personalization (UMAP '19), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3320435.3320445

UMAP '19, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6021-0/19/06…$15.00
https://doi.org/10.1145/3320435.3320445

1 INTRODUCTION

Researchers have started to notice the benefits of incorporating contextual information into the recommendation process [1–3]. Contextual information has been applied in various recommendation domains, including music [5, 7, 50], film [39], social [51], tourism [47] and learning [48].

In this work, we focus on context-aware music recommendation. Existing research suggests that the influence of *contextual characteristics* on the functions of music listening outweighs the influence of personal characteristics [15, 16]. A number of researchers have focused on incorporating individual context factors such as mood [5], daily activity [50], and time of day [7], but did not investigate how a combination of various contextual characteristics may influence the way people interact with music recommendations. In addition, some research has investigated the influence of personal traits such as visual literacy, locus of control, musical sophistication, etc. [20, 31, 32], but not the influence of contextual characteristics. To address this gap, we designed a music recommender system where users are allowed to indicate the importance (i.e., weight) of six contextual characteristics that were identified in a literature review of contextual characteristics that influence music listening behavior. These include mood, location, weather, social aspects, current activity and time of the day. An online experiment was conducted with 114 participants recruited from Amazon MTurk. Two different versions of the system (one with and one without controlling contextual information) were deployed using a mixed design. The following research questions are addressed in this paper:

RQ1: How does adding control of contextual characteristics influence user perceptions of the system?

RQ2: How do different contextual characteristics such as mood, weather, time of day (time), and presence of other people (social) [26] influence user perceptions of the system?

RQ3: How do different scenarios (location, activity) influence user requirements for user control of music recommendations?

The results of our study show that providing control over contextual information increases perceived recommendation quality without increasing cognitive load. *Mood* seems to be the most significant contextual characteristic. In addition, we found that users are more likely to interact with context parameters and their own profile during relaxing activities.

2 RELATED WORK

We first review previous work on user control in music recommender systems. We then discuss which contextual characteristics may be suitable to consider for different types of control.

2.1 User Control in Recommender Systems

Many recommender systems are opaque to users, and more user control can increase the perceived quality of recommendations [35]. In addition, users tend to be more satisfied when they have control over how recommender systems produce suggestions for them [27]. Controllability often allows users to steer the recommendation process to obtain suggestions that are better suited to them [17]. This, in turn, promotes trust in the system [10, 13], and hence improves the acceptance of recommendations. Controlling the recommendation process may range from providing ratings for an item to adjusting algorithm parameters.

Controlling recommendation results can be as simple as sorting or liking/disliking recommended items. For instance, a system may allow users to iteratively select relevant items from the recommended list to indicate his/her desired outcomes. This approach does not require users to specify the exact features of the desired outcomes, thus the demand for domain expertise is low [30]. MusiCube [40] is an example music recommender system that allows users to fine-tune recommendations by evaluating the results.

Controlling user profiles implies allowing users to modify their profile in order to match their evolving preferences, which has been shown to improve acceptance of recommendations [6, 19, 42]. Millecamp et al. [32] implemented a Spotify-based music recommender system that allows users to fine-tune recommendations by modifying their favorite artists and musical attributes, such as danceability, energy, and valence. Schaffer et al. [42] implemented a movie recommender system and investigated the impact of profile manipulation. Their experiment results showed that users were able to identify sources of bad recommendations and remove them.

Controlling algorithm parameters includes adjusting the weight of a parameter which is usually invisible to users [37]. A prominent example of a music recommender system that allows such control is TasteWeights [8]: users are able to change the weights of their favorite artists. MoodPlay [5] is a music recommender system that allows users to indicate the intensity of their mood. Similar to TasteWeights and MoodPlay, we allowed users to modify the weight of algorithm parameters. Instead of focusing on one contextual characteristic, we took into account a number of characteristics that may be relevant to music listening experiences. These characteristics are reviewed in the next section.

2.2 Contextual Characteristics

Context-aware recommender systems assume that recommendations based on users and items alone may be insufficient and that they can be improved by integrating contextual information into the recommendation process [1, 3]. For music recommender systems, mood has been shown to have a great importance for providing better recommendations [3, 5]. In addition to MoodPlay, there are a number of other music recommender systems that employed differing types of contextual information, such as daily activity [50], time of the day [7], and trending topics [8]. In our work, we are investigating how a combination of contextual characteristics may influence the way people interact with music recommendations. To motivate which contextual characteristics to investigate, we reviewed which characteristics have been observed to influence music listening behavior, as well as the literature on context-aware recommendation. We divide these into two broad categories: environment-, and user-based contextual characteristics.

2.2.1 Environment-based Contextual Characteristics.

Location. North et al. [33] studied the influence of location on the preference of music listening, and showed that the preference of musical descriptors varies depending on the location: certain preferences were predominately reported while being in a particular location. For instance, locations that can be considered arousing (e.g., party) are associated with a preference for musical descriptors that further increase arousal (e.g., invigorating, exciting/festive, loud). On the other hand, locations that represent a low degree of arousal (e.g., yoga room) is associated with a preference for musical descriptors that would further reduce arousal (e.g., relaxing/peaceful or quiet) [33]. Similarly, Krause et al. [28] found that the intensity of music being listened to varies across locations.

Time of day (time). A number of studies have found significant effects of the influence of time of day on music listening behavior. North et al. [34] showed that a great percentage of listening activities occurs in the evening and at the weekend. However, Krause et al. [28] argued that the ubiquitousness of mobile phones, computers and other listening devices allow people to listen to music in any location they find themselves in the daytime, such as the workplace. Similarly, Rana and North [38] found that people are more likely to use music to help them concentrate or think during the working hours than during the evenings. Baltrunas and Amatriain [7] introduced a time-aware music recommender system that suggests albums based on the time of the day. Our approach uses time to recommend music on a different level of granularity (i.e., tracks).

Presence of other people (social). Tarrant et al. [45] found that youths may listen to music for the purpose of fulfilling emotional needs (e.g. when being lonely). Among adults, Egermann et al. [12] observed that music listening was more arousing alone. On the other hand, Liljestrom et al. [29] found that participants have intense emotional responses together with a close friend or partner.

Weather. It has been shown that weather has a significant influence on our mood [41]. Researchers found that pleasant weather is related to a positive mood, while hotter weather in summer may lead to a negative mood [23]. A recent experiment conducted by Spotify in collaboration with AccuWeather also revealed that weather has a significant influence on valence. Specifically, users listen to higher-energy and happier-sounding music on sunnier days, and the opposite on rainy days [46].

2.2.2 User-based Contextual Characteristics.

Activity. Research has consistently found that music listening occurs frequently during personal maintenance (e.g., housework, cooking), active leisure activities (e.g., exercise, socializing), and travelling (e.g., driving, walking) [14, 22, 34, 43, 49]. Volokhin and Agichtein [49] showed the variability of activities people engage in while listening to music, but certain activities such as driving, housework, exercise, and cooking remain at the top across age groups and cultures. Heye and Lamont [18] found that people who frequently listen to music while on the move mainly listen for enjoyment, passing time, and enhancing emotional states. Wang et al. [50] introduced a music recommender system based on users' daily activities. Their experiments showed improved recommendations even in the absence of pre-existing ratings or annotations.

Mood. Previous studies investigating the effects of music therapy have shown that emotion and music are strongly related [25]. Thus, emotion such as valence has been used widely to define attributes of a song. MoodPlay is a music recommender system that employs mood as contextual information to provide better recommendations [5]. It allows users to indicate the intensity of their mood on a linear scale from strong to weak. Dhahri et al. introduced a system that relies on users' mood and implicit feedback to recommend music without any prior knowledge about the user preferences [11]. CoFeel is a user interface that uses emotions for social interaction in a group recommender [9]. CoFeel has been demonstrated as a group music recommender system, GroupFun, which suggests a common playlist for a group by taking into account the music taste of all group members. Bardo uses speech recognition to transform what players say during a role-playing game session into emotion tags and subsequently recommends background music [36].

2.3 Research gap

We reviewed existing work that incorporates contextual information in music recommender systems. Although interesting results have been obtained, the conducted research so far is rather *ad hoc*: a specific contextual characteristic is incorporated and effects on music recommender effectiveness are discussed. In this paper, we research in a more systematic way the influence of different contextual characteristics on interactive music recommendations. Based on a review of both environmental and user based contextual characteristics that have been shown to influence music listening behavior, we selected a broad range of contextual characteristics that may influence interactive music recommendations. These contextual characteristics can be modified in an interactive music recommender interface. We present a user study to systematically assess the influence of these different contextual characteristics on a variety of aspects of music recommender systems.

3 SYSTEM DESIGN

We implemented a context-aware music recommender system, ContextPlay, to evaluate whether and how users control information about contextual characteristics to tailor recommendations. This section explains the recommendation algorithm and the user interface simulating a mobile device for different contextual scenarios.

3.1 Algorithm

We implemented a context-aware recommendation algorithm based on the Spotify recommendation API[1]. Our system creates playlists for the different contextual characteristics identified in the music listening literature. The algorithm consists of two steps: 1) recommendation generation and 2) recommendation ranking.

[1]https://developer.spotify.com/documentation/web-api/reference/browse/get-recommendations

First, we use the Spotify personalization API[2] to retrieve top listened artists, tracks, and genres as the **user profile** for generating recommendations. The maximum number of recommended songs is 100 per call. At the beginning, our algorithm chooses a top artist, a top track, and a genre as input seeds to generate 100 songs.

Then, we rank recommendations based on the following context model: $\mathcal{M}$ = {*Location*, *Activity*, *Mood*, *Weather*, *Time*, *Social*}.

Each dimension contains several values; *Location* = {*outdoor*, *highways*, *home*, *office*}, *Activity* = {*driving*, *jogging*, *working*, *relaxing*}, *Mood* = {*happy*, *sad*}, *Weather* = {*sunny*, *rainy*}, *Time* = {*morning*, *evening*}, and *Social* = {*alone*, *party*}. In total, we defined 16 context tags based on these instances. This list is non-exhaustive, but incorporates the different contextual characteristics that were identified in earlier research as relevant for music listening.

We use the Spotify search API[3] to query the playlists that match the specified context tags. We collected 400 songs for each of the 16 context tags. We considered these songs as annotated data and trained a classifier C for the 16 context tags using the SVM algorithm (RBF kernel). We trained this model based on 12 musical features[4] available in the Spofity API. For an input song s, this classifier yields probabilities p for 16 context tags. The sum of all probabilities is equal to 1. Based on the weights w (0-100) assigned for the tags, we can calculate a score $C(s)$ to measure the fit of the song s for the current context.

$$C(s) = \sum_{i=1}^{16} p_i * w_i \tag{1}$$

Of note, all of the tags of each contextual characteristic have the same weight. In the end, we ranked the 100 recommended songs $\mathcal{S}$ by their score. The higher the score for a song, the better it matches the context.

In sum, we first generate personalized recommendations based on the user profile, and then ranked the candidate recommendations based on the contextual information.

3.2 User Interface

Figure 1 shows the user interface designed for our experiments. The interface was designed to fulfill two requirements: 1) it should stimulate listening to music on mobile phones for the presented scenarios, and 2) it should allow users to modify the recommendation algorithm by modifying the contextual information and their own profiles. The interface used in the experiment consists of the following components: depiction of context scenarios, the ContextPlay prototype, and experimental instructions. Below we describe each component in detail.

3.2.1 Scenarios. The background of the interface is constantly adjusted to represent a given scenario of music listening. In addition to the background image, we also try to stimulate users by a short passage describing the *location* and *activity* in the mobile screen before the task. Figure 1, for instance, shows the scenario "Home-Relaxing: imagine you are resting on a sofa at home. Now you want to create a playlist to help you relax.", with a matching background

[2]https://developer.spotify.com/documentation/web-api/reference/personalization
[3]https://developer.spotify.com/documentation/web-api/reference/search/search
[4]popularity, danceability, energy, speechiness, acousticness, instrumentalness, liveness, valence, tempo, loudness, key, and mode

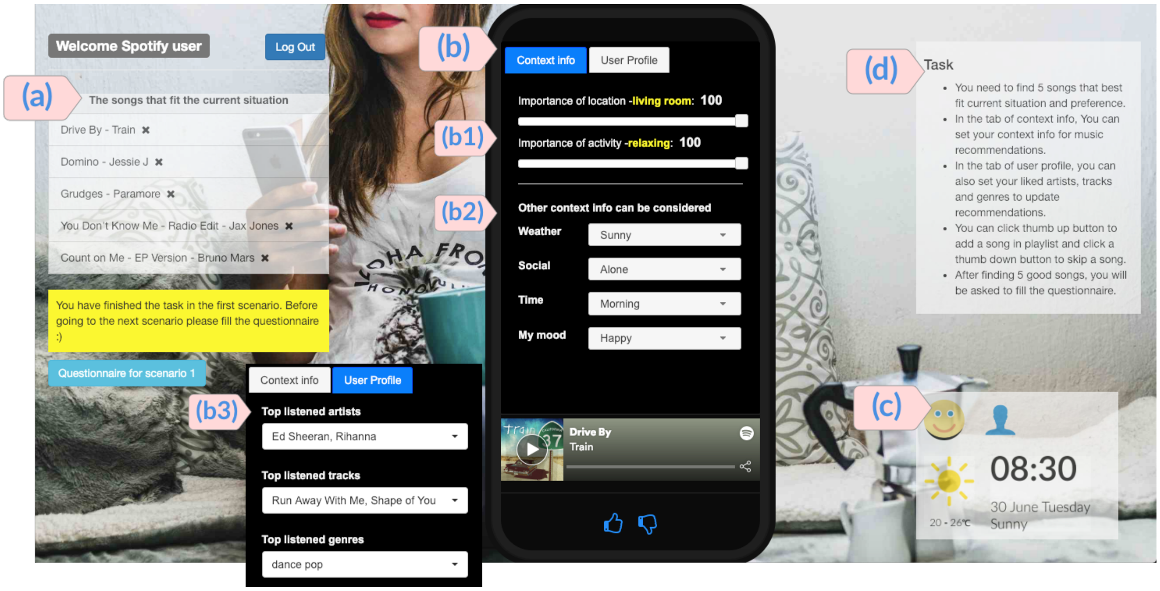

Figure 1: User interface highlighting the scenario of "Home-relaxing"; a) a playlist for showing a user's liked songs), b) the simulated mobile interface of ContextPlay, c) a panel to inform participants the current context information, d) a set of instructions showing the task of the experiment.

figure depicting this scenario. In the experiment, we used four typical scenarios of music listening, namely working in the office, driving on the highway, resting at home, and jogging on the street. The scenarios are further justified in Section 3.3.

3.2.2 ContextPlay. We designed the interface to allow modifying contextual parameters on a simulated mobile phone (see Figure 1, *b*). This simulated mobile interface has two tabs to allow users to navigate between the *user profile*, and the *contextual information*.

User profile. Users are able to specify their musical preferences in all conditions by selecting artists, tracks and genres to be seeded from. Figure 1, *b3* shows the view of the user profile containing three dropdown items with multi-select options, in which users can select multiple artists, tracks and genres to be considered for generating recommendations. However, due to the restriction of the Spotify API, we only allowed the users to select up to two artists, two tracks, and one genre.

Contextual information. The "context info" component consists of two sections (Figure 1, *b1 and b2*). In the first section (*b1*), two sliders enable controlling the weights for location and activity. The values of location and activity are fixed: users are not allowed to modify the location and activity, because they are determined by the scenario depicted in the background image, but they can modify the weight or relative importance of these two factors with a slider. The second section (*b2*) has four drop-down menus for modifying additional contextual characteristics, such as mood, weather, time, and social aspects. By default, these contextual characteristics are empty. Leaving these contextual characteristics optional allows us to study which characteristics are important for users. Of note, the order of the context field was counterbalanced among the participants using a Latin square design.

When users update these additional contextual characteristics (mood, weather, time, and social), related icons will appear in the interface. The four icons shown in the panel (Figure 1, *c*) represent, for instance, that the user is listening to *music with other people, has a happy mood, and that it is a sunny morning*. When users change the context settings, the system also updates the recommendations.

The recommended items are presented one by one in an embedded Spotify player. A pair of thumb rating icons at the bottom allows the user to rate songs. A positive rating saves the song, while a negative rating skips the song and shows the next recommended song in the ranking.

3.2.3 Experiments. The user interface also contains some widgets to support the experiment. A playlist (Figure 1, *a*) shows the songs liked by a user. When the list contains five songs a questionnaire link will appear below the playlist. A panel on the right side (Figure 1, *d*) lists all of the required steps of the experimental task.

3.3 Scenarios

Location and activity are the most influencing contextual characteristics to music listening [15]. Therefore, in our experiment, we define listening scenarios by considering different activity levels and locations. Figure 2 shows a 2x2 taxonomy for the music listening context: location on one axis, which specifies whether the user is listening to music indoor or outdoor, and on the other axis the activity, either focused or relaxing.

For our experiment, we selected four scenarios which correspond to a quadrant of this taxonomy. The locations and activities chosen are selected to be diverse with regard to the type and amount of

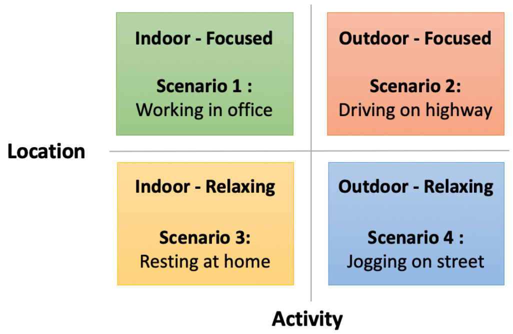

Figure 2: A taxonomy for scenarios based on location and activity.

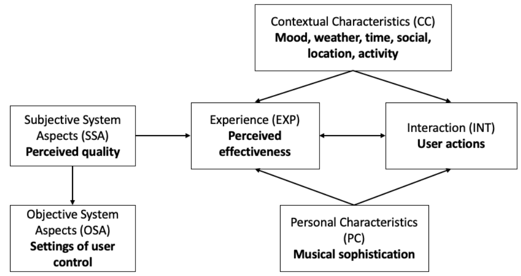

Figure 3: The user-centric evaluation framework of recommenders as used in our experiment.

user control that would be expected to be optimal (e.g., driving is largely a hand-free type of activity).

4 EXPERIMENTAL DESIGN

We conducted a mixed-design user study to answer our research questions.

Using a between-subjects approach, we introduce two experimental settings (one with, and one without, controls for contextual information) and each setting was evaluated by a group of participants (N=70). To avoid the influence of UI on the experiment, the two settings have the same UI layout and visual style. In the baseline condition, user control of context information is hidden, but the participant can still inspect the context information.

Using a within-subjects approach, we investigate how two of four scenarios (counter-balanced using a Latin Squares design) influence the actual control behaviour and the requirements of user control on context-aware recommendation. We measure different aspects of the user experience based on the user-centric evaluation framework for recommender systems [24].

4.1 Hypotheses

In this study, we evaluated a context-aware music recommender system with two settings of user control to investigate how the control of context information (RQ1) and contextual characteristics (RQ2) affect user perception of system, and how the scenarios influence the user requirements on control for music recommendations (RQ3). Therefore, we have the following hypotheses:

H1: The *control of context* significantly increases user perception of system.

H2: The *contextual characteristics* will influence user perception of system.

H3: The *scenarios* will influences user requirement on control for the music recommender.

4.2 Participants

We recruited 140 participants (Age: Mean = 29.92, SD = 9.12; Gender: Female = 45.86%, Male = 54.41%) with Amazon Mechanical Turk (mTurk), and paid $1 USD for an estimated completion time of 30 minutes. The participants were required to have a minimum approval rating of 90%. We recorded the unique worker IDs of participants who completed the experiment to avoid repeated participation. After rejecting the participants with contradictory and low quality responses, we kept 114 participants with valid responses.

4.3 Measurements

Recent related work found that musical sophistication (MS) can influence how users interact with, and benefit from, control in music recommender systems [20, 21]. In order to control for how MS influences the way of controlling context information, we measure MS in the pre-study questionnaire using ten questions from the sub-scale "general MSI" of Goldsmiths Musical Sophistication Index (Gold-MSI)[5]. Since the setting of user control may influence user cognitive load [21], we also use NASA-TLX[6], a widely used questionnaire to assess cognitive load after finishing each task.

Our experiment uses a user-centric evaluation framework for recommender systems [24] to gauge how the setting of user control (objective factor) and contextual characteristics influence the user experience and interaction. This framework (Figure 3) shows the potential *interactions* between components such as Objective System Aspects (OSA), Subjective System Aspects (SSA), User Experience (EXP), Interaction (INT), Contextual Characteristics (CC), and Personal Characteristics (PC). We employed this framework to analyze how user control of contextual information (OSA) influences perceived quality and diversity (SSA), perceived effectiveness (EXP), and user interaction (INT). Moreover, we consider the effects of musical sophistication (PC), and contextual characteristics (CC), on EXP and INT. All question items are measured on a seven-point Likert scale from Completely disagree to Completely agree.

We also record user actions in a log file, including the completion time of task, the specific context info and user profile elements modified by users, the number of times users modify context and user profile elements, and the number of listened songs and skipped songs.

4.4 Procedure

The **experimental task** is to find five good songs that best match the presented scenario and user's music preference. Each participant needed to perform this task for two scenarios. The procedure contains the following steps:

[5] http://www.gold.ac.uk/music-mind-brain/gold-msi/
[6] https://humansystems.arc.nasa.gov/groups/tlx

(1) *Tutorial of study* - Participants were asked to read the description of the user study and watch a video tutorial introducing the main features of systems. Only the available features are shown for a particular setting in the video. The "Start" button of the study is only activated after the end of the tutorial. After authorizing their Spotify accounts to the system, participants are redirected to a pre-study questionnaire.
(2) *Pre-study questionnaire* - This questionnaire asks the user's age and gender, and measures musical sophistication.
(3) *Manipulating recommender and rating songs* - Users can modify the context and user profile elements as explained above. The system presents recommended songs one by one in an embedded player. The play buttons allow users to listen to 30-second excerpts. The system will reset the context information and user profile before starting the second scenario. To ensure sufficient time spent on exploring recommendations, the rating widget only appears 20 seconds after showing the recommended song. Similarly, the link for the questionnaire only appears after finishing the experimental task.
(4) *Post-study questionnaire* - Participants fill a post-study questionnaire after finishing the task in both scenarios. According to Knijnenburg's framework [24], this questionnaire asks five questions for perceived quality, diversity, and effectiveness, and six questions for gauging cognitive load. In the end, users are able to provide free text comments.

5 RESULTS

5.1 Analytical approaches

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[7]. All answers to the questions are modeled as ordinal variables. SEM is able to analyze the effects in an integrative structure where we can associate all the detected effects.

We employ several validated questions items [24] to measure each subjective factor in a questionnaire such as perceived quality, perceived diversity, and effectiveness.

To establish the validity of these question items, we perform a Confirmatory Factor Analysis (CFA) before evaluation. As a result, we refine the answers to our questions (Table 1) and establish the validity of the factors in our study.

The resulting SEM model (Figure 4) shows how control for context (OSA), musical sophistication (PC), and contextual characteristics (CC) influences perceived diversity and quality (SSA), effectiveness (EXP), and user interaction (INT).

The fit of our SEM model is adequate: χ^2_{114} = 154.90, p <.01; root mean squared error of approximation (RMSEA)= 0.040; 90% CI: [0.022, 0.055], Comparative Fit Index (CFI) = 0.996; Turker-Lewis Index (TLI) = 0.995.

5.2 Effects of control settings

The results of the SEM (see Figure 4) show that *the control settings* (OSA) *have a directly correlation with perceived quality*. Compared to the baseline, control of contextual characteristics increases perceived quality (p<0.01). In turn, the increased quality positively affects perceived effectiveness, and the increased effectiveness allows users to finish the task (find five good songs) with fewer attempts. However, the control setting has no significant effect on perceived diversity and cognitive load.

Thus, we can accept the hypothesis **H1**: *The control of context significantly increases user perception of system.*

5.3 Effects of contextual characteristics

In total, we investigated six contextual characteristics (CC). We removed three items (activity, social, and time) from the SEM model because they do not have significant effects on any other factors. We find that *mood* positively correlates with perceived quality (p<.05) and diversity (p<.05), which further positively influences the perceived effectiveness. Moreover, both *weather* and *location* directly influence effectiveness: weather tends to negatively influence the effectiveness (p<.05) and location tends to positively influence the effectiveness (p<.05). Note, since the value of location is preset, encoding "0/1" represents for different locations (indoor/outdoor) rather than considering location or not. Therefore, we can interpret this result as outdoor tends to increase the effectiveness. In addition, cognitive load negatively influences effectiveness (p<.01). Thus, we can accept hypothesis **H2**: *The contextual characteristics will influence user perception of system.*

5.4 User interaction

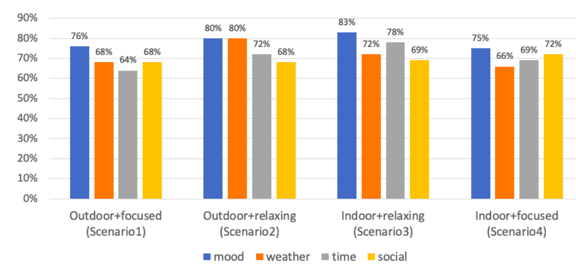

Figure 5: The percentage of users who controlled each context information in four scenarios.

To understand the effects of scenario on user interactions, we recorded which control components' values were changed by users in each scenario. Thereby, we know which contextual characteristics and user profile elements are more likely to be modified in a particular scenario. Since two experimental settings have different control components, we only analyze interaction data (N=61) in the setting of full user control.

Figure 5 shows the percentage of participants who modified each **contextual characteristic** in four scenarios. The figure does not include the data of location and activity because their values were preset in the experiment. In general, *mood* is the most modified CC in all scenarios, and weather was modified as much as mood in the outdoor+relaxing scenario. Additionally, in general, users seem to modify contexts slightly more in *relaxing* scenarios (Scenarios 2 and 3) than *focused* scenarios (Scenarios 1 and 4)

[7]http://lavaan.ugent.be/

Construct	Question items	R^2	AVE
Perceived quality	I liked the songs recommended by the system.	0.77	0.86
	The recommended songs fitted my preference.	0.87	
	The recommended songs were well-chosen.	0.93	
	The recommended songs were relevant.	0.88	
Effectiveness	I would recommend the music recommender to others.	0.89	0.72
	The music recommender has no real benefit for me.	0.49	
	I can save time using the music recommender.	0.78	
Perceived diversity	The recommendations contained a lot of variety.	0.94	0.75
	The recommendations covered many music genres.	0.63	
	Most songs were from the same genre.	0.68	

Table 1: Results of Confirmatory Factor Analysis (CFA). Question items with R2 values lower than 0.5 and large modification indices are removed in the refined results. Both the convergent validity and the discriminant validity of our model hold.

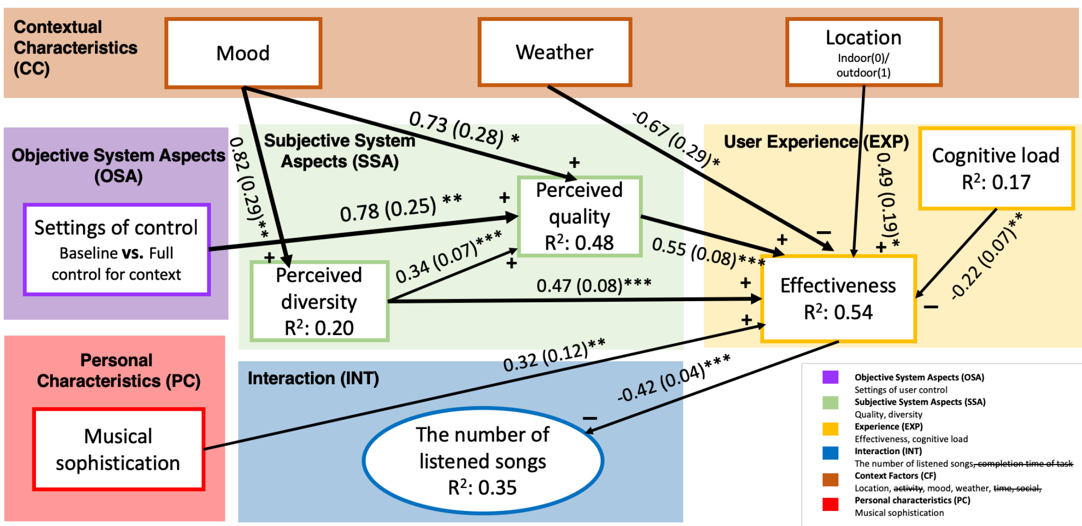

Figure 4: The structured equation modeling (SEM) results. The number (thickness) on the arrows represents the β coefficients and standard error of the effect. Significance: *** p < .001, ** p < .01, * p < .05. R^2 is the proportion of variance explained by the model. Factors are scaled to have a SD of 1.

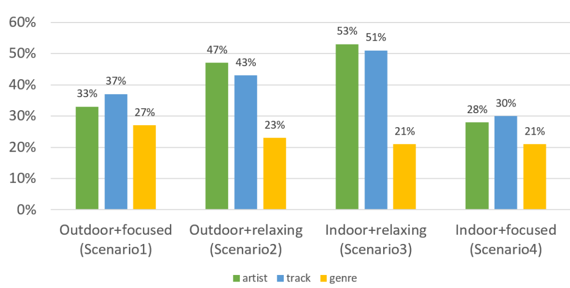

Figure 6: The percentage of users who controlled each user profile element in four scenarios.

Figure 6 shows the percentage of participants who manipulated **user profile** elements. The most manipulated components are the liked *artists* and *tracks* in four scenarios. Only a quarter of participants modified *genres* in all of the scenarios. The *relaxing* scenarios seem to engage more users to modify their profile compared to the *focused* scenarios.

Thus, we can accept hypothesis **H3**: *Scenarios will influence user requirement on control for the music recommender.*

5.5 Musical sophistication

In line with earlier results [21], the personal characteristic musical sophistication has a positive effect on perceived effectiveness, which in turn increases the efficiency for performing the task.

5.6 Subjective feedback

We asked participants which contextual characteristics they would like to consider for music recommendations. Overall, participants commented positively about controlling context, e.g., *"this is useful for creating playlists for myself for different situations.". Mood* is the most mentioned contextual characteristic, mentioned by 92% participants, and two participants mentioned they would like to have a fine grained control for mood. This was followed by activity (75%) and social (56%). Only 42% of participants mentioned location and weather. Two participants also mentioned other characteristics they would like to be taken into account: season and temperature.

Participants also indicated that more control over genres was desired: *"include more genres for the variety".*

6 DISCUSSION

In this section, we discuss our findings on control of contextual characteristics, and its influence on perception and interaction.

6.1 Control of context

Earlier work has shown that combining control over the user profile and algorithm parameters leads to better perceptions of recommendations than only controlling a single component [21]. Similarly, our results suggest that *users tend to perceive higher quality of recommendation when they can control the context*, which positively influences effectiveness and user interaction. Compared to the control over the user profile in the baseline setting, control of context seems to be very effective in searching for songs for a particular situation. Therefore, we could confirm the hypotheses **H1** and answer research question **RQ1**. Some user comments also reflect this merit. In addition, control of context does not lead to higher cognitive load, which is in line with our earlier findings [21] that combining two levels of control does not lead to higher cognitive load compared to one level of control.

We suggest that recommender designers include control of context to increase perceived recommendation quality.

6.2 Contextual characteristics

Our results show that *mood*, *weather*, and *location* tend to directly influence the users' perception of music. The most influential contextual factor is mood as it positively influences perceived recommendation quality and diversity. This is in line with previous research [4, 5, 44] which has found a strong correlation between music and mood. Numerous participants commented that mood plays an important role in music listening and they would like to control mood for music recommendation. Surprisingly, we found that weather negatively influences the (perceived) effectiveness. This may be influenced by the quality of the trained models for weather tags such as rainy and sunny.

In addition, users tend to perceive higher effectiveness in an outdoor environment. A potential reason may be that users are more likely to associate music with the current context while doing outdoor activities. Therefore, we could confirm the hypotheses **H2**, and answer research question **RQ2**.

Our recommendation to system designers is to offer a fine-grained control for mood in order to increase perceived recommendation quality and diversity.

6.3 User interaction

The SEM model only shows a significant effect for the number of listened songs for user interaction. In addition, the log file also reveals which control components were more likely to be manipulated by users in each scenario. Overall, *mood* was the most modified factor, which might be explained by the high relevance between mood and music mentioned in subjective feedback.

As for the user profile, we see users are more likely to modify artists and tracks rather than genres. Due to the limitation of Spotify API, we show genres based on user top listened artists, and hence most genres should be familiar to participants. However, the subjective feedback indicates users would like to see more genres to increase recommendation diversity. Moreover, users are more likely to modify both contextual characteristics and user profile elements in the relaxing scenarios than in the focused scenarios because of the higher attention required by the focused activities. Thus, this addresses research question **RQ3**, and we can accept hypothesis **H3**.

Recommender practitioners may consider adapting control settings to a particular scenario, such as fewer control options when the user is performing a focused task.

7 LIMITATIONS

Firstly, for the scenarios we asked participants to consider the recommendations in a simulated environment using background images and the context icons. While this helped participants to imagine the hypothetical scenarios, it may be still difficult for participants to accurately assess their behaviors or preferences.

Secondly, we trained a classification model for context tags using the user created playlists on Spotify. However, for a specific context users may have different music preferences, which may influence the effectiveness of the classification model.

Thirdly, although we asked users to manipulate the system based on the contextual environment and their music preferences, some users may still click on some components out of curiosity. Thus, the recorded actions may not perfectly reflect user intentions.

8 CONCLUSION

We have described ContextPlay, a context-aware music recommender system that empowers users to inspect and control the contexts considered for music recommendations. Unlike previous systems that focused on individual contextual characteristics, ContextPlay took into account six characteristics that are related to music listening. We then conducted an online user experiment to investigate how controlling contextual characteristics influences user perception and interaction with recommendations. We found that the control of contexts, and in particular mood, has positive effects on perceived recommendation quality and diversity. Moreover, users tend to modify context parameters and their own profile in the relaxing scenarios regardless of locations. Overall, compared to existing research on context-aware music recommenders, our research considered six contexts and sheds light on the contexts that users might want to control and the potential advantages of controlling contexts in such a music recommender. In the future, we plan to extend our research findings by implementing a mobile app and experimenting in the wild.

ACKNOWLEDGMENTS

Part of this research has been supported by the KU Leuven Research Council (grant agreement C24/16/017).

REFERENCES

[1] Gediminas Adomavicius and Francesco Ricci. 2009. RecSys' 09 workshop 3: workshop on context-aware recommender systems (CARS-2009). In *Proceedings of the third ACM conference on Recommender systems*. ACM, 423–424.

[2] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering* 6 (2005), 734–749.

[3] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 217–253.

[4] Judy I Alpert and Mark I Alpert. 1990. Music influences on mood and purchase intentions. *Psychology & Marketing* 7, 2 (1990), 109–133.

[5] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: Interactive Mood-based Music Discovery and Recommendation. In *Proc. of UMAP '16*. ACM, 275–279.

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

[7] Linas Baltrunas and Xavier Amatriain. 2009. Towards time-dependant recommendation based on implicit feedback. In *Workshop on context-aware recommender systems (CARS'09)*. Citeseer.

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

[9] Yu Chen and Pearl Pu. 2012. CoFeel: Using Emotions for Social Interaction in Group Recommender Systems. In *First International Workshop on Recommendation Technologies for Lifestyle Change (LIFESTYLE 2012)*. 48.

[10] Pieter Walter de Vries. 2004. *Trust in systems: effects of direct and indirect information.* Technische Universiteit Eindhoven.

[11] Chaima Dhahri, Kazunori Matsumoto, and Keiichiro Hoashi. 2018. Mood-Aware Music Recommendation via Adaptive Song Embedding. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 135–138.

[12] Hauke Egermann, Mary Elizabeth Sutherland, Oliver Grewe, Frederik Nagel, Reinhard Kopiez, and Eckart Altenmüller. 2011. Does music listening in a social context alter experience? A physiological and psychological perspective on emotion. *Musicae Scientiae* 15, 3 (2011), 307–323.

[13] Gavan J Fitzsimons and Donald R Lehmann. 2004. Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science* 23, 1 (2004), 82–94.

[14] Alinka E Greasley and Alexandra Lamont. 2011. Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae* 15, 1 (2011), 45–71.

[15] Fabian Greb, Wolff Schlotz, and Jochen Steffens. 2016. Situational and dispositional influences on the functions of music listening. (2016).

[16] Fabian Greb, Wolff Schlotz, and Jochen Steffens. 2017. Personal and situational influences on the functions of music listening. *Psychology of Music* (2017), 763–794.

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

[18] Andreas Heye and Alexandra Lamont. 2010. Mobile listening situations in everyday life: The use of MP3 players while travelling. *Musicae Scientiae* 14, 1 (2010), 95–120.

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

[20] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of individual traits on diversity-aware music recommender user interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, 291–299.

[21] 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*. ACM, 13–21.

[22] Patrik N Juslin, Simon Liljeström, Daniel Västfjäll, Gonçalo Barradas, and Ana Silva. 2008. An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 8, 5 (2008), 668.

[23] Matthew C Keller, Barbara L Fredrickson, Oscar Ybarra, Stéphane Côté, Kareem Johnson, Joe Mikels, Anne Conway, and Tor Wager. 2005. A warm heart and a clear head: The contingent effects of weather on mood and cognition. *Psychological science* 16, 9 (2005), 724–731.

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

[25] Stefan Koelsch. 2009. A neuroscientific perspective on music therapy. *Annals of the New York Academy of Sciences* 1169, 1 (2009), 374–384.

[26] Sherrie YX Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly* (2006), 941–960.

[27] Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *UMUAI'12* 22, 1 (2012), 101–123.

[28] Amanda E Krause, Adrian C North, and Lauren Y Hewitt. 2016. The role of location in everyday experiences of music. *Psychology of Popular Media Culture* 5, 3 (2016), 232.

[29] Simon Liljeström, Patrik N Juslin, and Daniel Västfjäll. 2013. Experimental evidence of the roles of music choice, social context, and listener personality in emotional reactions to music. *Psychology of Music* 41, 5 (2013), 579–599.

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

[31] 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. *Procs. of IUI*, 1–12.

[32] Martijn Millecamp, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. 2018. Controlling Spotify recommendations: effects of personal characteristics on music recommender user Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, 101–109.

[33] Adrian C North and David J Hargreaves. 1996. Situational influences on reported musical preference. *Psychomusicology: A Journal of Research in Music Cognition* 15, 1-2 (1996), 30.

[34] Adrian C North, David J Hargreaves, and Jon J Hargreaves. 2004. Uses of music in everyday life. *Music Perception: An Interdisciplinary Journal* 22, 1 (2004), 41–77.

[35] John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1085–1088.

[36] Rafael R Padovani, Lucas N Ferreira, and Levi HS Lelis. 2017. Bardo: Emotion-Based Music Recommendation for Tabletop Role-Playing Games. (2017).

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

[38] Shabbir A Rana and Adrian C North. 2007. The role of music in everyday life among Pakistanis. *Music Perception: An Interdisciplinary Journal* 25, 1 (2007), 59–73.

[39] Alan Said, Shlomo Berkovsky, and Ernesto W De Luca. 2010. Putting things in context: Challenge on context-aware movie recommendation. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*. ACM, 2–6.

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

[41] Jeffrey L Sanders and Mary S Brizzolara. 1982. Relationships between weather and mood. *Journal of General Psychology* (1982).

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

[43] John A Sloboda, Susan A O'Neill, and Antonia Ivaldi. 2001. Functions of music in everyday life: An exploratory study using the Experience Sampling Method. *Musicae scientiae* 5, 1 (2001), 9–32.

[44] Valerie N Stratton and Annette H Zalanowski. 1991. The effects of music and cognition on mood. *Psychology of Music* 19, 2 (1991), 121–127.

[45] Mark Tarrant, Adrian C North, and David J Hargreaves. 2000. English and American adolescents' reasons for listening to music. *Psychology of Music* 28, 2 (2000), 166–173.

[46] Eliot Van Buskirk. 2017. Spotify, Accuweather Reveal How Weather Affects Music Listening. https://insights.spotify.com/us/2017/02/07/spotify-accuweather-music-and-weather/

[47] Mark Van Setten, Stanislav Pokraev, and Johan Koolwaaij. 2004. Context-aware recommendations in the mobile tourist application COMPASS. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer, 235–244.

[48] Katrien Verbert, Nikos Manouselis, Xavier Ochoa, Martin Wolpers, Hendrik Drachsler, Ivana Bosnic, and Erik Duval. 2012. Context-aware recommender systems for learning: a survey and future challenges. *IEEE Transactions on Learning Technologies* 5, 4 (2012), 318–335.

[49] Sergey Volokhin and Eugene Agichtein. 2018. Understanding Music Listening Intents During Daily Activities with Implications for Contextual Music Recommendation. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM, 313–316.

[50] Xinxi Wang, David Rosenblum, and Ye Wang. 2012. Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 99–108.

[51] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Zi Huang. 2014. A temporal context-aware model for user behavior modeling in social media systems. In *Procs. of SIGMOD*. ACM, 1543–1554.