

## RESEARCH

# User Models for Culture-Aware Music Recommendation: Fusing Acoustic and Cultural Cues

Eva Zangerle\*, Martin Pichl\* and Markus Schedl†

Integrating information about the listener’s cultural background when building music recommender systems has recently been identified as a means to improve recommendation quality. In this article, we, therefore, propose a novel approach to jointly model users by their *musical preferences* and *cultural backgrounds*. We describe the musical preferences of users by the acoustic features of the songs the users have listened to and characterize the cultural background of users by culture-related socio-economic features that we infer from the user’s country. To evaluate the impact of the proposed user model on recommendation quality, we integrate the model into a *culture-aware recommender system*. By analyzing a dataset comprising approximately 400 million listening events of about 55,000 users from 36 countries, we show that incorporating both acoustic information of the tracks a user has listened to as well as the cultural background of users in the form of a *music-cultural user model* contributes to improved recommendation performance. Furthermore, we provide a systematic analysis of the influence of different features on the quality of the provided culture-aware track recommendations. We find that considering acoustic features that model the characteristics of tracks and a user’s musical preferences have the highest impact on recommendation performance. However, adding socio-economic features allows further improving the recommendation quality. In addition, we identify interesting correlations between acoustic characteristics of music preferences and cultural features of populations at the country level.

**Keywords:** user modeling; context-aware recommender systems; music-cultural user model; culture-aware recommender systems

## 1 Introduction

Recent advances in recommender systems and music information retrieval have shown that contextual information is vital for highly personalized results (e.g., Wang et al. (2012a); Braunhofer et al. (2013); Pichl and Zangerle (2018)). In this scope, context can be defined as “conditions or circumstances which affect some thing” (Kaminskas and Ricci, 2012; Adomavicius and Tuzhilin, 2011), where, e.g., environment-related contextual information may include location, time or weather (Kaminskas et al., 2012). Consequently, the user’s listening context can be defined as the user’s context during listening to music. To this end, the geographic location of a user is often exploited as one basic notion of context. Leveraging GPS coordinates to model *similarity* between listeners, which is key to building recommender systems, results in location-aware systems, which are however agnostic to cultural characteristics and the cultural background of users. In the scope of this article, we define the *cultural background* of users as a

set of attributes that allow for describing the culture the user is embedded in, including social or economic aspects, as well as, e.g., cultural practices, values, and behavior. However, location alone does not necessarily serve as a good indicator for the cultural background of a user, as geographically close users might have a very different cultural background. A user’s cultural background may also not coincide with political borders (Pichl et al., 2017). Notably, the cultural background of a user was identified already by Schedl and Schnitzer (2014) as a possibly relevant aspect to improve recommender systems. We hence argue that modeling users based on musical properties of the songs they listen to (approximating their musical preference) on the one hand and the user’s cultural background on the other contributes to capturing *music-cultural listening patterns*. These patterns particularly describe the complex interrelation between users, their cultural background, and the characteristics of the music they listen to. In this article, we propose a novel *music-cultural user modeling* approach to exploit such listening patterns for recommender systems by integrating information about (i) the acoustic qualities of the music users have listened to and (ii) culture-specific information derived from the users’ location/country to describe the user’s likely cultural background.

\* University of Innsbruck, AT

† Johannes Kepler University (JKU) Linz and Linz Institute of Technology (LIT) AI Lab, AT

Corresponding author: Eva Zangerle ([eva.zangerle@uibk.ac.at](mailto:eva.zangerle@uibk.ac.at))

Leveraging a standardized collection of almost one billion user-generated listening events, we evaluate the proposed user model.<sup>1</sup> By exploiting music-cultural listening patterns captured by the proposed user model in a recommender system, we show that the resulting culture-aware music recommendations are more accurate than those provided by a recommender agnostic to cultural information. Particularly, we find that capturing a user's individual music taste by the high-level audio features of the tracks the user has listened to and adding Hofstede's cultural dimensions (Hofstede et al., 1991) as well as data from the World Happiness Report (WHR) (Helliwell et al., 2016) as a description of the cultural (and socio-economic) background of the user provides the best recommendation results, in terms of accuracy and error measures.

The remainder of the article is organized as follows. Section 2 briefly reviews related work on context- and culture-aware music recommendation. The dataset we use, a processed version of the LFM-1b dataset (Schedl, 2016), is presented in Section 3. Section 4 provides details on (i) our methods for user modeling according to musical preferences and cultural aspects, and (ii) our proposed culture-aware recommender system. The experiments we conducted to evaluate the user models and recommender system approaches are explained in Section 5. We present and discuss the results obtained in Section 6. To gain more insights into the overall and country-specific patterns of acoustic music preferences, Section 7 presents results of an additional study on differences in acoustic preferences between countries and on correlations between cultural and musical features. The paper is rounded off by a summary and outlook to follow-up research in Section 8.

## 2 Related Work

In music recommender systems, unlike for instance in movie recommendation, content-based approaches have been the dominant focus of research for a long time (Knees and Schedl, 2016). Music content is, in this case, either incorporated into the recommendation algorithm in the form of hand-crafted acoustic features or—more recently—by automatic feature extraction from the raw audio signal using deep neural networks. Examples of the former include a rich set of features that have been proposed in the past two decades of music information retrieval research, and range from Mel frequency cepstral coefficients (MFCCs), e.g., Logan (2002), to semantic descriptors of acoustic properties, e.g., Miotto et al. (2010); Turn-bull et al. (2008). For an overview, consider, for instance, Casey et al. (2008); Knees and Schedl (2013). Deep learning-based approaches to automatic feature learning for content-based music recommendation include convolutional neural networks (CNN) and recurrent neural networks (RNN), in particular their variants long short-term memory (LSTM) and gated recurrent units (GRU). For a more detailed review of deep learning approaches in music recommendation, please consider Schedl (2019).

Nowadays, it has become widely accepted that incorporating contextual information into recommender systems contributes to improved recommendations (Adomavicius and Tuzhilin, 2011). Particularly for music recommender

systems, studies showed that users often seek for music that matches their current situation, and hence context (i.e., occasion, event or emotional states) (Kim and Belkin, 2002; Lee and Downie, 2004). In the scope of music recommender systems, Kaminskas and Ricci (2012) distinguish environment-related context (location, time or weather), user-related context (activity, demographic information or emotional state of the user), and multimedia context (text or pictures the user is currently reading or looking at). For our study, the environment-related context of a user is of particular relevance as we aim to leverage both the musical preferences and cultural background of users for improving track recommendations.

Schedl and Schnitzer (2013) performed a study on the contribution of geospatial information to the performance of artist recommender systems. They conclude that if users listen to various different artists, the integration of geospatial information is beneficial. Schedl et al. (2014) approximate the cultural distance of users by the country or continent a user is located in and show that this is beneficial for users particularly in the U.S. and Russia. Furthermore, there are several approaches that exploit places of interest as contextual information, where the idea is to recommend music that suits the environment—in an emotional or cultural sense (Kaminskas et al., 2013; Braunhofer et al., 2011). Rich sensory devices such as smart phones allow mapping a certain location to a certain activity that can be exploited for personalized location-based music recommendations, depending on the user's inferred activity (Wang et al., 2012b). Baltrunas et al. (2011a) propose a context-aware music recommender system for car drivers, where a set of diverse contextual factors are incorporated (e.g., driving style, traffic conditions, weather or road type). Ankolekar and Sandholm (2011) propose the Foxtrot system, which allows users to tag music with geolocations. Based on this information, users can be provided with location-specific music recommendations. Cheng and Shen (2014) model the listener's short-term music needs, their location, and the music's overall popularity to create personalized music recommendations. Hu and Ogiwara (2011) propose a music recommender system that integrates track genre, release year, freshness, and temporal aspects.

As for cultural aspects in the broader field of music information retrieval, Ferwerda and Schedl (2016) found that a user's cultural background (modeled by Hofstede's cultural dimensions (Hofstede et al., 1991)) influences how diverse the musical preferences of users are. Particularly, they found that highly individualist countries and countries that are flexible, pragmatic, and eager to adapt to changes listen to more diverse genres. Schedl et al. (2017) also performed a study on whether cultural similarity between countries (described by Hofstede's cultural dimensions and the Quality of Government (QoG) dataset) is reflected in music taste (described by tags annotating music tracks). They found medium correlations of music taste and several cultural and socio-economic factors. Notably, this evaluation is based on the LFM-1b dataset, which is also utilized in the experiments conducted in this study. Furthermore, Liu et al. (2018,

2017) have uncovered similarities between countries based on cultural and socio-economic aspects on the artist level and on the album level.

Pichl et al. (2017) clustered users based on their individual musical preferences and their cultural characteristics. Relying on density-based spatial clustering, they find nine clusters that describe similar users regarding both their musical preference and cultural background. The cultural background of users was described by the World Happiness Report (Helliwell et al., 2016) and the authors found that incorporating cultural information allows for more precise user descriptions compared to relying on geographic information only. However, this evaluation did not target recommender systems and was done on a substantially smaller dataset.

We are not aware of any work exploiting the cultural background of users for the computation of context-aware music recommendations and hence locate a research gap here. In this paper, we show that utilizing the cultural background of users together with their general musical preference contributes to improved recommendation quality.

### 3 Data

In this section, we present the data utilized for performing our analyses and experiments.

For our analyses, we require a dataset that contains a substantial number of listening histories of users as well as country information about these users. There are indeed a number of datasets containing listening histories: the Million Musical Tweets Dataset (Hauger et al., 2013) and the MusicMicro dataset (Schedl, 2013) come with contextual information related to time and location. The musical listening histories dataset (Viglienconi and Fujinaga, 2017), the Yahoo! Music ratings dataset (Dror et al., 2012) and the #nowplaying dataset (Zangerle et al., 2014) contain a substantial number of users, items also including timestamps of LEs; however, no contextual information regarding the user's country is given. Hence, we base our investigations on the LFM-1b dataset (Schedl, 2016), which contains more than one billion listening events created by users of the online music platform Last.fm,<sup>2</sup> where music listeners can share information about their listening behavior. The LFM-1b dataset has been created in the following way using various endpoints of the Last.fm API (Schedl, 2017): first, the top artists labeled by any of the 250 top user-generated tags used on Last.fm were retrieved. Then, the top fans of these artists were fetched, resulting in about 465,000 users. Listening histories (i.e., each user's set of listening events) of a randomly chosen subset of 120,322 users were subsequently downloaded. The creation times of the listening events cover the time span between January 2005 and August 2014.

Since we aim to model music-cultural preferences jointly by individual musical preference and the cultural background of users, we require the data to contain information about the location of the user. For 45.87% of all users within the LFM-1b dataset, country information about the user is available. Therefore, we constrain the dataset to those users (and their tracks) for whom we are

able to obtain country information. This provides us with a dataset comprising 55,191 users, who have listened to a total of 26,022,625 distinct tracks, which are captured by a total of 807,890,921 listening events.

Besides the information contained in the LFM-1b dataset, we also require information about the tracks the users listened to (cf. Section 4.1). Particularly, we are interested in content features that are able to describe a given track. Therefore, we rely on the Spotify API to gather content-based audio features, as described in Section 4.1, for each track. For all listening events of users for whom we can obtain country information, we search for the <track, artist, album> triples extracted from the LFM-1b dataset using the Spotify search API<sup>3</sup> to gather the Spotify URI of each track (i.e., we provide all three parts in a conjunctive query). This URI is subsequently used to query the audio features API,<sup>4</sup> which returns the set of audio features describing the contents of a given track (cf. Section 4.1), which allowed gathering 4,326,809 Spotify URIs. For the remainder of the tracks, the Spotify API is not able to correctly resolve the triples to a track. We attribute this to two factors: either the searched track is not provided by Spotify or the track, artist, and album information cannot be matched to a Spotify track URI unambiguously. Also, the Spotify API does not provide all features for all tracks and hence, we remove those tracks for which the API does not provide a full set of audio features from the dataset. Employing this procedure, we are able to acquire the full set of audio features for a total of 3,478,399 tracks. Notably, these 13.36% of the distinct tracks for which we can obtain audio features are able to capture 48.89% of all listening events (i.e., the tracks listened to by users).

The remaining tracks and respective listening events are excluded from the dataset. This eventually results in a dataset of 55,149 users, 394,944,868 listening events and 3,478,399 distinct tracks. **Table 1** depicts the main characteristics of the dataset underlying our analyses.<sup>5</sup> As can be seen, the average number of listening events per user is 7,161, which we consider a substantial number that is able to capture a user's individual musical preferences well. Furthermore, the average number of

**Table 1:** Statistics of the dataset utilized (LE = listening event).

Item	Value
Listening events	394,944,868
Users	55,149
Distinct tracks	3,478,399
Min. LE per user	1
$Q_1$ LE per user	1,442
Median LE per user	5,667
$Q_3$ LE per user	9,738
Max. LE per user	399,210
Avg. LE per User	7,161.41 ( $\pm 10,326.91$ )
Avg. Users per Country	1,155.93 ( $\pm 1,894.96$ )

users per country is 1,156. Along the lines of Ferwerda and Schedl (2016), we constrain the dataset to countries with more than 200 users to ensure that countries are well-characterized and results are valid and representative (at least of a typical music streaming community such as the one at Last.fm). **Table 2** depicts the number of users

**Table 2:** Number of users per country for countries with more than 200 users. We use ISO 3166 2-digit country codes to abbreviate country names.

Abbrev.	Country	Users
US	United States	10,251
RU	Russian Federation	5,021
DE	Germany	4,576
UK	United Kingdom	4,533
PL	Poland	4,403
BR	Brazil	3,882
FI	Finland	1,409
NL	Netherlands	1,375
ES	Spain	1,242
SE	Sweden	1,230
UA	Ukraine	1,140
CA	Canada	1,077
FR	France	1,055
AU	Australia	976
IT	Italy	973
JP	Japan	798
NO	Norway	750
MX	Mexico	705
CZ	Czechia	632
BY	Belarus	558
BE	Belgium	513
ID	Indonesia	484
TR	Turkey	478
CL	Chile	425
HR	Croatia	372
PT	Portugal	291
AR	Argentina	282
CH	Switzerland	277
AT	Austria	276
HU	Hungary	272
DK	Denmark	271
RS	Serbia	253
RO	Romania	237
BG	Bulgaria	236
IE	Ireland	219
LT	Lithuania	202

per country for all countries with more than 200 users within our dataset. In total, the cleaned dataset features users from 36 different countries. Note that countries in this article are abbreviated using their ISO 3166 2-digit country code.<sup>6</sup>

## 4 Methods

In the following, we detail the proposed approach for leveraging individual and cultural listening patterns for the computation of track recommendations based on the underlying dataset (as described in Section 3). We first present our user modeling approach (for individual and cultural listening patterns) and secondly present the proposed music-cultural user model. Subsequently, we show how we leverage this model for the computation of track recommendations.

### 4.1 User Modeling: Musical Preferences

As for modeling individual musical preferences, we gather content-based audio features for each of the tracks in the dataset by querying the Spotify API<sup>7</sup>—following the lines of, e.g., Pichl et al. (2016); Andersen (2014); McVicar et al. (2011). We make use of these Spotify high-level features for a number of reasons: first, the LFM-1b dataset does not contain audio data that we could use to extract audio features from. Second, our analyses aim at investigating the general suitability of merging acoustic and cultural cues for music recommendation rather than low-level feature engineering and hence, we rely on Spotify’s audio features as a compact characterization of tracks. These content features are extracted from the audio signal of a track and comprise:

1. *Danceability* describes how suitable a track is for dancing and is based “on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.”
2. *Energy* measures the perceived intensity and activity of a track. This feature is based on the dynamic range, perceived loudness, timbre, onset rate and general entropy of a track.
3. *Speechiness* detects presence of spoken words in a track. High speechiness values indicate a high degree of spoken words (talk shows, audio book, etc.), whereas medium to high values indicate e.g., rap music.
4. *Acousticness* measures the probability that the given track contains only acoustic instruments.
5. *Instrumentalness* measures the probability that a track contains no vocals (i.e., it is instrumental).
6. *Tempo* quantifies the rate of the beat in beats per minute.
7. *Valence* measures the “emotional positiveness” conveyed by a track (i.e., cheerful and euphoric tracks reach high valence values).
8. *Liveness* captures the probability that the track was performed live (i.e., whether an audience is present in the recording).

### 4.2 User Modeling: Cultural Aspects

As for the cultural dimension, we propose to model cultural aspects on a country level and make use of two different resources: Hofstede’s cultural dimensions

(Hofstede, 1980; Hofstede et al., 1991)<sup>8</sup> and the World Happiness Report<sup>9</sup> of 2016 (Helliwell et al., 2016), which we describe in the following.

A widely accepted instrument to describe cultures is Hofstede's cultural dimensions (HOF). This framework describes a nation's culture and values by the following six dimensions:

1. *Power distance* (PD) is defined as "the extent to which the less powerful members of organizations and institutions (like the family) accept and expect that power is distributed unequally" (Helliwell et al., 2016).
2. *Individualism* (IDV) captures the extent to which people are integrated into groups. Societies with high scores possess only loose ties and the individual is considered more important than the collective group.
3. *Masculinity* (MAS) assesses a preference in society for achievement, heroism, assertiveness and material rewards for success. Low masculinity (femininity) signals a preference for cooperation, modesty, caring for the weak and quality of life.
4. *Uncertainty avoidance* (UA) measures to which degree members of a society tolerate ambiguity. Countries with a high score tend to rely on stiff codes, guidelines, and laws. In contrast, lower scoring countries show more tolerance and acceptance of differing thoughts.
5. *Long-term orientation* (LTO) measures the connection of the past with current and future actions or challenges. Low-scoring societies tend to keep traditions and norms and are suspicious of societal change, while high-scoring societies encourage thrift and adaptation.
6. *Indulgence* (IND) captures the happiness of a country and "relatively free gratification of basic and natural human drives related to enjoying life and having fun". In countries with low indulgence scores, gratification of needs is suppressed and regulated by strict social norms.

In addition to Hofstede's cultural dimensions, we complement our model with socio-economic characteristics of countries. We capture these by figures extracted from the World Happiness Report (WHR) (Helliwell et al., 2016). Schimmack et al. (2002) showed that cultural factors are directly influenced by the subjective well-being of people. Therefore, we rely on the WHR as it captures people's cognitive and affective evaluations of their daily life and thus, their subjective well-being (Diener, 2000) on a country level. The WHR provides the following set of measures capturing the perceived happiness of countries:

1. *Freedom* measures the perceived freedom to make life choices.
2. *Healthy life expectancy* captures the healthy life expectancy at birth in a given country.
3. *Generosity* specifies whether people in a country are willing to spend money on a charity.
4. *Social support* states if people have people helping them if they need support (i.e., relatives or friends).

5. *Trust* measures the publicly perceived absence of corruption in government and business.
6. *Happiness* quantifies the subjective and perceived happiness.
7. *GDP* is the real gross domestic product per capita.

#### 4.3 Music-Cultural User Model

Based on the features we leverage to capture a user's musical preferences (Section 4.1) and a user's cultural background (Section 4.2), we propose the following music-cultural user model for computing culture-aware recommendations.

Generally, we characterize a user's individual musical preferences and cultural background in a single feature vector. As for capturing a user's *individual musical preferences* based on the tracks listened to, we leverage the audio features of tracks as presented in Section 4.1. Except for tempo, all of these features are given in the range of [0,1]. For tempo, we apply a linear min-max scaling to also represent it in the range of [0,1]. To exclude tracks with audio features that distort a user's aggregated musical features, we remove outlier tracks from the user's listening history by applying the median absolute deviation (MAD) outlier detection method (Leys et al., 2013). We consider a feature value an outlier if it is not within  $M \pm a \cdot MAD$ , where  $M$  is the median of this particular feature across all tracks of a user and  $MAD$  is the median absolute deviation of these values. As for the choice of  $a$ , we set a strongly conservative threshold  $a = 3$  as proposed by Leys et al. (2013). Hence, a value is considered an outlier if it is not within three  $MAD$ s around the median. Lastly, a track is considered as an outlier in the list of tracks of a particular user if one of its features is considered an outlier and consequently removed from the user's listening history. For each of the features, we compute the average feature value and the standard deviation across all tracks in the user's listening history and add these average and standard deviation (SD) values to the user's feature vector. We chose to add the standard deviation of each of these features to mitigate the effects of averaging a large number of features that potentially differ substantially.

For the approximation of the *cultural background* of users (or rather, the country they live in) by socio-economic aspects, we rely on the variables of Hofstede's cultural dimensions and the World Happiness Report and extract these based on the user's country information. We add these variables to the feature vector to find *cultural* listening patterns that reflect cultural similarity better than the geographic distance. For each of these variables, we perform a linear min-max scaling such that all elements of the vectors are within [0, 1] and concatenate it with the user vector.

#### 4.4 Recommendation Computation

We model the computation of context-aware music recommendations based on the proposed user model as a learning task for rating prediction, where we aim to learn the probability  $P$  that a given user  $u$  has listened to a given track  $t$ . To learn these probabilities  $P(u, t)$  for all users and tracks, we rely on Gradient Boosting Decision Trees.

Particularly, we utilize the popular XGBoost system (Chen and Guestrin, 2016), a scalable end-to-end tree boosting approach which has been shown to be highly suited for recommendation tasks (Pacuk et al., 2016; Ayaki et al., 2017; Tran, 2016). Using XGBoost, we set the learning objective to logistic regression for binary classification, which provides us with the desired probabilities. For the training phase, we set the training objective to be the binary classification error rate (i.e., the number of wrongly classified tracks in relation to all tracks classified, where tracks with a prediction value larger than 0.5 are classified as relevant for the given user, and all other tracks are considered irrelevant for the user).

Please note that we deliberately chose a classification-based recommendation approach and refrained from utilizing more elaborate recommender approaches such as context-aware matrix factorization (Baltrunas et al., 2011b) or tensor-based factorization approaches (Karatzoglou et al., 2010) as we aim to focus on user modeling aspects in this paper. Hence, we chose to compare different user models based on a simple classification-based recommendation approach which also allows us to get a deeper understanding of the contribution of individual features of the user model (cf. Section 6).

For the classification task carried out, we require a rating for each track that allows us to define whether a given track was listened to (and thus, considered relevant) for a given user. Hence, we add a binary factor (rating) to the processed dataset: for each unique <user, track> combination, the rating  $r_{ij}$  is 1 if the user  $u_i$  has listened to track  $t_j$  at least once. Please note that users and tracks may be represented by different models as described in Section 5.1.2. Due to a lack of publicly available data, our dataset does not contain any implicit feedback of users (i.e., skipping behavior, session durations, or dwell times during browsing the catalog). This is why we cannot estimate any preference towards an item a user has not listened to as proposed by Hu et al. (2008). Thus, we assume tracks the user has not listened to (in the case of implicit data, all non-observed tracks) as negative examples (Hu et al., 2008). Even though there is a certain bias towards negative values as some missing values might be positive, Pan et al. (2008) found that this method for rating estimation works well. The rating  $r_{ij}$  for a given user  $u_i$  and given track  $t_j$  can now be defined as stated in Equation 1.

$$r_{i,j} = \begin{cases} 1 & \text{if } u_i \text{ listened to } t_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

We train an XGBoost model that performs a binary classification on the relevance of tracks for the given users. We extract the probabilities underlying the classification decision, which can be used to (i) perform a ranking of tracks by their probability of relevance in the recommendation task which allows us to conduct a ranking-based evaluation of the proposed models, and (ii) evaluate the predictive performance of the proposed models by computing error metrics.

## 5 Experiment Design

This section reports on the experiments conducted for evaluating the previously described culture-aware recommender system.

### 5.1 Experimental Setup

In the following, we first present the user models evaluated and describe the evaluation method utilized for capturing the recommendation performance of the proposed user model.

#### 5.1.1 Evaluation Strategy

To evaluate the performance of the proposed contextual user modeling in regard to recommendation quality, we perform a per-user evaluation. Therefore, we use each user's listening history and perform a *leave-k-out* evaluation per user (also referred to as hold-out evaluation) (Cremonesi et al., 2008; Breese et al., 1998; Cremonesi et al., 2008), where we set  $k$  to 50 (as described later in this section).

The underlying dataset only provides items with positive feedback (Hu et al., 2008) (i.e., items that have been listened to by the user) gathered via users' listening histories. As the recommendation task is transformed into a rating prediction task, we require the dataset to also include negative examples. Therefore (and as described previously in Section 4.4), for each user, we randomly add tracks the user did not interact with (i.e., tracks  $t_j$  with  $r_{ij} = 0$  for the given user  $u_i$ ) to the dataset until the listening history of each user in both the training and test sets are filled with 50% relevant and 50% non-relevant items for the user. We chose to oversample the positive class to avoid class imbalance and hence, a bias towards the negative class (the number of tracks not listened to is much larger than the number of tracks listened to, for all users).

As we aim to evaluate the benefit of adding cultural aspects in a track recommendation scenario, we also need to characterize tracks. For our proposed model, we rely on the acoustic features of each track and add these to the track vector. However, we also need to assign cultural features to tracks to be able to match users of a certain culture with tracks that are listened to by users with a similar cultural background. This is particularly relevant for tracks in the negative class. Preliminary experiments showed that we cannot assign randomly computed cultural features or the cultural features of the current user to tracks as this causes the XGBoost model to learn that all tracks with the user's culture assigned belong to the positive class, whereas all tracks from any other culture (i.e., culture information that is consistent across a number of users or purely random culture information) belong to the negative class. Therefore, we propose to assign the cultural features of the country in which the track is most popular to each track. We argue that the track is most characteristic and representative for the country in which the track is most popular. Therefore, we first compute the playcounts of each track in each country within the dataset. Next, we normalize the playcount ( $PC$ ) of each track  $t \in T$  (i.e., the universe of tracks in the dataset) in each country  $c$  by the total amount of listening events

of the country (i.e., we compute  $\frac{PC(c,t)}{\sum_{j \in T} PC(c,j)}$  for each country  $c$  and for each track  $t$ ). This allows us to infer the country in which it accounts for the highest share of listening events and hence, is most popular. We subsequently assign the culture of this country to the track. For obtaining negative samples (tracks), we randomly select a track from the dataset that the current user has not listened to and again assign this track the cultural features of the country where the track is most popular.

Based on the dataset that now contains an equal amount of positive and negative samples for each user, we use a leave- $k$ -out evaluation strategy. Therefore, we have to compute a hold-out set of size  $k$  for each user: along the lines of previous research (He et al., 2017; Elkahky et al., 2015), we randomly select 50 positive samples (tracks that the user has listened to) and 500 negative samples (tracks the user has not listened to). These 550 tracks form the test set for each user, whereas the recommender system is trained on the remainder of the dataset. Subsequently, we compute the predicted ratings for the tracks in the test set as presented in Section 4.4, aiming to rank the 50 positive samples on top, whereas the negative samples should be ranked on the bottom of the ranked list of recommendations.

### 5.1.2 Evaluated Models and Baselines

To assess the performance of each of the proposed user models, variations thereof and two baseline approaches in terms of recommendation quality, we separately evaluate these different user models and compare their performance. An overview of the evaluated modeling approaches is depicted in **Table 3**. The evaluated models describe a user either by the user’s individual music preferences described by the acoustic features of the tracks the user listened to (U\_AF), the user’s cultural/socio-economic background described by Hofstede’s dimensions (U\_HOF) and the World Happiness Report (U\_WHR), or the user identifier (U\_ID). Similarly, we describe tracks by their acoustic features (T\_AF), the culture they are embedded in (T\_HOF and T\_WHR) or by their track identifier (T\_ID). Please note that we include the user and track identifiers in the respective models as this allows us to extend and directly compare the approaches to a baseline model (User + Track), that is only based on these two identifiers. As

**Table 3:** Overview of evaluated models, where features prefixed with U describe a user and features prefixed with T describe a track; the models on two last rows serve as baselines.

Model	User Features	Track Features
Music + Culture	U_ID, U_AF, U_WHR, U_HOF	T_ID, T_AF, T_WHR, T_HOF
Music	U_ID, U_AF	T_ID, T_AF
Culture	U_ID, U_WHR, U_HOF	T_ID, T_HOF, T_WHR
Country	U_ID, U_Country_ID	T_ID, T_Country_ID
User + Track	U_ID	T_ID

can be seen from **Table 3**, we evaluate the music-cultural model (*Music + Culture*) as proposed in Section 4.3. We also individually evaluate the performance of a model solely relying on musical preferences of users and features of tracks (*Music* model), and analogously a model that describes users and tracks by their cultural background (*Culture* model).

Furthermore, we investigate a set of baselines to compare our proposed models to. First, we evaluate an approach that uses each user’s listening history and additionally, utilizes the user’s country code (e.g., US for users from the United States) as contextual information for both the user and the track (*Country* model). Here, we aim to evaluate whether the country code may act as a proxy for cultural factors of users. Furthermore, we evaluate a context-agnostic baseline relying solely on the users’ listening histories and hence, a model that solely relies on the user and track ids for classification (*User + Track*) in a traditional collaborative filtering approach.

### 5.1.3 Evaluation Metrics

We model the context-aware recommendation of tracks as a rating prediction task, therefore we use the root mean squared error (RMSE) and mean absolute error (MAE) to measure the prediction error. We compute the RMSE and MAE for each individual user and consequently compute the average among all users. Furthermore, we are also interested in a decision-based evaluation (Celma, 2010) of our approach and therefore, compute *precision*, *recall*, and the  $F_1$ -measure to assess the top- $n$  accuracy (Cremonesi et al., 2010), where  $n$  is the number of top-ranked track recommendations that is evaluated. Therefore, we require the set of computed recommendations to be ranked. Hence, we rank the track recommendation candidates with respect to the probability that they belong to the positive class in descending order and compute the top- $n$  track recommendations. Next, we have to transform the rating prediction task into a binary classification task (Pan et al., 2008) for deciding whether a given track is relevant or not for a given user. For our experiments, we consider all predicted probabilities  $P(u, i) > 0.5$  as a predicted interaction and thus, we consider these items as relevant, all others as irrelevant.<sup>10</sup> For assessing the overall *precision*, *recall*, and  $F_1$ -measure of the evaluated recommender systems, we compute the measures for each individual user and compute the average among all users. For computing the *recall* measure, all relevant items in the test set are considered, independent of the number of recommendations. Thus, there is a natural cap for *recall*, namely the number of recommendations divided by the number of relevant items in the test set.

Regarding the number  $n$  of evaluated recommendations, we argue that exposing a user to more than 10–20 tracks at a time might provoke choice overload and hence, is barely meaningful. The problem of choice overload has been addressed by Bollen et al. (2010) who state that user satisfaction is highest when presenting the user with top-5 to top-20 items—assuming that the recommendation list contains a sufficient number of relevant items for the user.

Hence, we are particularly interested in the performance of the proposed recommendation approaches for lower values of  $n$ . Furthermore, we argue that in the presented scenario, precision is the more important measure to consider from a user perspective as it is able to capture the user's effective utility of the provided recommendations better (Bellogin et al., 2011) and hence, the practical value of the recommender system for the user. Thus, we argue that particularly the precision@10 results are relevant for our evaluation. As for the tuning of XGBoost parameters, we performed a preliminary cross-evaluation aiming to optimize precision values for the proposed models and hence, set the maximum number of trees to learn the models to 1,000. For all other parameters, we rely on the default settings.

## 6 Experimental Results and Discussion

In the following, we first present the findings of the top- $n$  recommendation evaluation task (Section 6.1), before presenting the evaluation of the underlying rating prediction task in Section 6.2. Subsequently, we elaborate on the importance of individual features of the proposed user model (Section 6.3) and discuss the limitations of the approach (Section 6.4).

### 6.1 Top- $n$ Recommendation Evaluation

**Table 4** shows the results obtained by the evaluated user models (cf. **Table 3**), where we consider the top-10 ranked recommended tracks for the evaluation. Regarding the precision of the computed recommendations, we observe that the best results are obtained by the proposed Music + Culture model, which incorporates both the user's general musical preferences and the cultural background of the user. This model reaches a precision@10 of 0.98, whereas the Music model reaches a precision of 0.95 and the Culture model a precision of 0.31, respectively. Compared to the baselines, we observe that using only the country of the user as a proxy for cultural aspects (Country model) achieves a precision value of 0.83, whereas the User + Track model performs worse, reaching a precision value of 0.13.

Regarding the recall values obtained, we observe that again, the Music + Culture model performs best (0.63), followed by the Music (0.59) and Culture (0.52) models. The User + Track baseline again reaches a lower value (0.08), whereas the Country model again performs well (0.52). For the sake of completeness, we also list the  $F_1$  values obtained by the individual models, which are

**Table 4:** Precision, recall, and  $F_1$ -score for all proposed models (sorted by performance; standard deviation in parentheses).

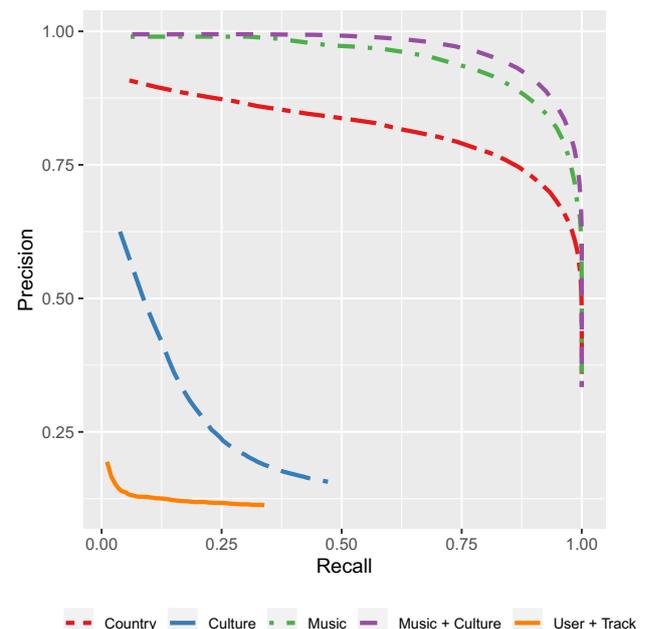
Model	Prec	Rec	$F_1$
Music + Culture	0.98 ( $\pm 0.04$ )	0.63 ( $\pm 0.15$ )	0.75 ( $\pm 0.10$ )
Music	0.95 ( $\pm 0.06$ )	0.59 ( $\pm 0.15$ )	0.72 ( $\pm 0.11$ )
Country	0.83 ( $\pm 0.11$ )	0.52 ( $\pm 0.12$ )	0.63 ( $\pm 0.10$ )
Culture	0.31 ( $\pm 0.15$ )	0.18 ( $\pm 0.08$ )	0.24 ( $\pm 0.09$ )
User + Track	0.13 ( $\pm 0.10$ )	0.08 ( $\pm 0.06$ )	0.13 ( $\pm 0.06$ )

consistent with the individual findings regarding recall and precision. In preliminary baseline experiments, we have also compared our approach with a traditional context-agnostic matrix factorization approach. Singular value decomposition based on implicit feedback achieved a precision of 0.49, a recall of 0.10, and an  $F_1$ -score of 0.17. As already elaborated, we consider the precision metric more relevant in this scenario. Thus, these baseline results show that the proposed models do indeed contribute to recommendation quality.

**Figure 1** shows a precision/recall plot of the evaluated approaches for  $n = 1 \dots 50$  track recommendations. From this plot, we again observe the superior performance of the music-cultural user model across all evaluated lengths of recommendation lists  $n$ . The plot also highlights the difference between the two models that incorporate acoustic features for describing musical preferences (Music + Culture and Music) and the remaining user models that do not exploit this information, where precision and recall are both substantially lower. These findings underline that the musical preference of users is paramount for recommendation scenarios. We can also observe that using the user's country as a proxy for their cultural background does indeed contribute. Naturally, including a set of cultural features to describe the user's cultural background also allows to exploit a more comprehensive, multi-dimensional notion of similarity between users (Schedl and Schnitzer, 2013), which can be exploited by the recommender system. We also have experimented with combining musical features and country code, however, this did not increase performance compared to using only musical features.

### 6.2 Rating Prediction Evaluation

Besides the decision-based evaluation regarding recall and precision, we are also interested in the prediction accuracy of the individual user models. **Table 5** presents the RMSE



**Figure 1:** Precision-recall-curves for top- $n = 1 \dots 50$  recommendations for all models.

and MAE per user across all tracks within the user’s test set. These findings are in line with the decision-based findings as the lowest RMSE is again achieved by the Music + Culture model (RMSE of 0.15). In comparison, relying solely on acoustic features to describe users and tracks (Music model) achieves a RMSE of 0.17, whereas relying on cultural aspects only results in a RMSE of 0.88. The baseline approaches reach RMSE values of 0.36 (Country model) and 0.93 (User + Track model), respectively. The evaluation of mean absolute errors of the individual models is consistent with the findings for RMSE.

### 6.3 Influence of Features

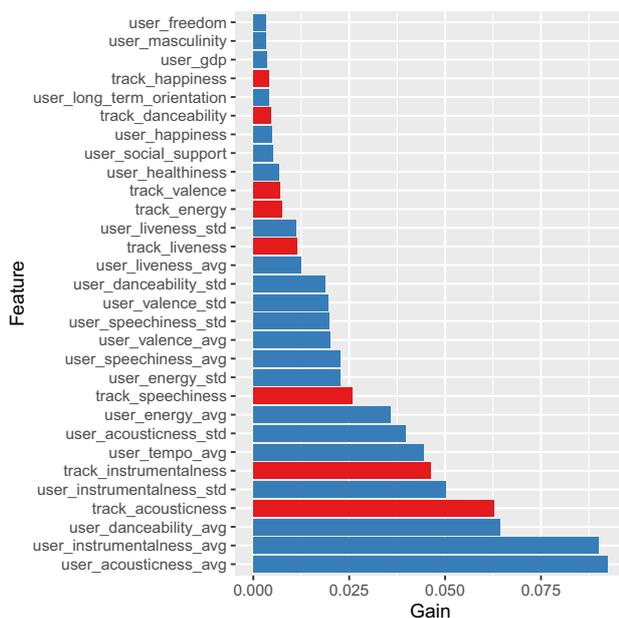
Apart from the performance of the proposed music-cultural user model in regard to recommendation quality, we are also interested in the contribution of the individual features of the user model to the trained XGBoost classification model. Therefore, we utilize the gain of each feature in the XGBoost model (Chen and Guestrin, 2016), which is a measure for the improvement in accuracy when adding a split on the given feature to the tree. This gain is computed for each feature in every tree of the trained model and is then averaged to a final gain value for each feature. **Figure 2** shows the contribution of the top-30 individual features to classification performance of the proposed music-cultural user model. Please recall that in

the proposed model, both users and tracks are described by musical and cultural features (cf. **Table 3**). Hence, we color the bars of user features in blue and track features in red. In total, acoustic features account for 93% of the gain (76% user features, 17% track features), WHR features account for 4% and Hofstede’s dimensions for 3% of the gains.

The results show that the major contributing features are related to the acoustic features that describe the user’s musical preference and the tracks. This high importance of acoustic features when it comes to describing users is congruent with the analyses of Pichl et al. (2017) and in line with the findings of the top- $n$  recommendation evaluation, where the Music model was the second best performing model. The features that contribute most to the classification accuracy (and hence, recommendation performance) are the average acousticness (`user_acousticness_avg`), instrumentality (`user_instrumentality_avg`) and danceability (`user_danceability_avg`) of tracks the user has listened to. As for the track features, acousticness and instrumentality are also the main contributing features. This high contribution of instrumentality and acousticness is in line with previous findings (Pichl et al., 2016), where these two features have been shown to discriminate tracks well in a principal component analysis. These findings are also congruent with the results of the evaluation conducted, where the user model that solely relies on the user’s preferences achieved the second best recall and precision values (performing substantially better than the Culture, Country, and User + Track models). However, while socio-economic factors are not among the top contributing features, socio-economic features nevertheless contribute to the recommendation quality and make a decisive difference regarding recommendation performance. The user features contributing most are healthiness, social support, happiness, GDP and masculinity and for tracks, the happiness and social support features provide the highest gain. While WHR features contribute more in our scenario, features stemming from both sources (WHR and Hofstede’s cultural dimensions) are among the top-contributing features; this also supports our choice to include both social and economic features in the user model as both contribute to higher recommendation performance.

**Table 5:** RMSE and MAE of all models.

Model	RMSE	MAE
Music + Culture	0.15	0.02
Music	0.17	0.03
Country	0.36	0.13
Culture	0.88	0.77
User + Track	0.93	0.85



**Figure 2:** Information gain of the top 30 individual user and track features of the Music + Culture model.

### 6.4 Discussion and Limitations

We believe that the proposed music-cultural user model and the conducted evaluation are an important first step towards culture-aware music recommender systems. The obtained results show that the proposed music-cultural user model outperforms all other evaluated models. However, we still see a few limitations of our approach, which we will elaborate on in the following. First, we currently represent the musical preferences of a user by utilizing the average of the acoustic features of the tracks the user has listened to and the standard deviation thereof. While we believe that this method is sufficiently elaborate for the experiments conducted, this is a rather naive approach towards representation and does not reflect the diverse and often context-related musical preferences of

users. Similarly, we currently use a rather simple majority voting approach for assigning cultural features to tracks. However, in the paper at hand, we are particularly interested in the influence of individual features and characteristics of users, their cultural background, and tracks on the recommendation performance and, hence, deliberately refrain from utilizing a more comprehensive user model. Nevertheless, looking into creating more comprehensive and complex user models based on the cultural background of users is part of our future research agenda. For instance, Zangerle and Pichl (2018) employed Gaussian Mixture Models (GMM) for modeling a user's diverse tastes of music and showed that utilizing such a GMM approach in combination with the acoustic features of the tracks the user listened to is able to capture a user's musical preferences well.

The test set creation procedure applied (random 50 positive and 500 negative samples per user) allows for evaluating the ability to distinguish positive and negative samples. We have also experimented sampling 10 relevant and 100 irrelevant tracks for each user, however, we argue that given the high number of listening events per user in the dataset, sampling 50 positive and 500 negative tracks reflects a more suitable scenario. The results achieved were high in precision and low on the prediction error metrics, showing that the proposed models were able to detect the 50 positive samples and rank these on top.

As already stated in Section 4.4, we consider the classification-based approach for the computation of recommendations as a baseline regarding the actual recommender system. However, we believe that even though the method is rather simple, it provides us with conclusive results regarding the user models evaluated, which was our focus.

## 7 Interplay Between Country Characteristics and Music Preferences

In the following, we analyze the cultural/socio-economic and acoustic features on a country level more thoroughly, aiming to uncover country-specific patterns of their inhabitants' music preferences in terms of acoustic features and to identify similarities and differences between countries (Section 7.1). We further investigate to which extent cultural/socio-economic and acoustic features correlate with each other, on a per-feature-basis (Section 7.2).

### 7.1 Country-specific Differences of Acoustic Feature Preferences

To obtain insights into country-specific particularities of the acoustic properties of music consumption, we provide an overview of the investigated acoustic features (and their standard deviations) per country, computed over all users in each country in **Table 6**. Overall, we observe pronounced differences between countries for most of the properties, but also non-negligible standard deviations within countries, indicating partly substantial variances in music preferences among citizens. Highest danceability in music preferences can be found in France (0.533), Colombia (0.532), and Mexico (0.529); the lowest in Iran (0.455). Notably, Iran is also the country with the lowest

music energy (0.599) in its population's preferences. In contrast, the populations of Finland (0.806), Bulgaria (0.801), and Hungary (0.800) like highly energetic music. This is further evidenced when investigating their preferred music styles, which include several variants of the genre metal. As for speechiness, the lowest figures are found in Indonesia and Argentina (both 0.048), whereas music listeners in Poland (0.065) tend to listen more commonly to music featuring spoken words such as hip-hop or rap. Acousticness is lowest for Finland (0.062) and Bulgaria (0.063); by far highest for Iran (0.278), China (0.232), and Turkey (0.199). As for instrumentality, by far the lowest-scoring countries are Brazil (0.029), Indonesia (0.040), and Argentina (0.059). At the other end, users in Romania (0.224) and Greece (0.198) particularly like non-vocal instrumental music. Regarding liveness, Iran (0.133) and Turkey (0.137) show the lowest values, whereas Finland (0.166) has the highest figures for this attribute. This may be explained by Finns having a particular preference for live music and by Finland having a very vivid music performing culture and therefore a large number of hobby musicians as well as (semi-)professional bands. Music listened to by Iranian users scores by far the lowest on the dimension of valence, on average (0.298). In stark contrast, music consumed in South and Middle America scores highest on this dimension; in particular, users in Colombia (0.486), Mexico (0.485), Argentina (0.482), and Brazil (0.478) tend to listen to a substantial amount of music that is suited to evoke positive emotions. Finally, when it comes to tempo, users in Iran and Turkey tend to prefer slower music, around 120 BPM on average. On the other hand, Venezuela, New Zealand, Hungary, and Germany prefer faster music, on average around 125 BPM.

### 7.2 Correlations Between Cultural Background and Music Preferences

To uncover possible relationships between acoustic properties of a country's inhabitants' music preferences and the cultural or socio-economic characteristics, we investigate the correlation between each of the acoustic features and the cultural/socio-economic dimensions. **Tables 7** and **8** depict Spearman's rank-order correlation coefficients for Hofstede's cultural features and WHR socio-economic characteristics, respectively. We use rank-order correlation to cope with the different value ranges of the various dimensions investigated and compute these correlations considering all users in our dataset as observations. To describe each user's aggregated musical feature vector, we follow the same approach as detailed in Section 4.3. Correlations larger than 0.1 (or less than  $-0.1$ ) are highlighted in bold. Statistically significant correlations are marked with an asterisk.

As a general observation, while almost all correlations are significant (even at  $p < 0.001$ ), most are only weak, which hints at the different nature of aspects to compare. Nevertheless, some interesting observations can be made. Focusing on **Table 7**, we observe notable correlations for the cultural trait of indulgence (IND). More precisely, a positive correlation between IND and acousticness (0.125) as well as valence (0.114) is identified. This means that societies that like to engage in joyful activities tend

**Table 6:** Means and standard deviations (in parentheses) of acoustic preferences of each country's users. The highest value of each acoustic property is printed in bold; the lowest in italic. Countries are sorted alphabetically according to their country code.

Country	Danceability	Energy	Speechiness	Acousticness	Instrumentalness	Liveness	Valence	Tempo
AR	0.512 (0.091)	0.739 (0.140)	<i>0.048</i> (0.017)	0.113 (0.163)	0.059 (0.166)	0.145 (0.034)	0.482 (0.122)	123.113 (7.756)
AT	<i>0.476</i> (0.102)	0.766 (0.172)	0.059 (0.025)	0.106 (0.182)	0.127 (0.227)	0.154 (0.042)	0.405 (0.133)	124.400 (8.483)
AU	0.491 (0.100)	0.746 (0.157)	0.057 (0.028)	0.112 (0.172)	0.119 (0.228)	0.153 (0.043)	0.435 (0.129)	123.562 (9.116)
BE	0.507 (0.106)	0.718 (0.170)	0.056 (0.029)	0.143 (0.198)	0.165 (0.260)	0.148 (0.045)	0.428 (0.129)	122.783 (8.825)
BG	0.491 (0.101)	0.801 (0.135)	0.062 (0.029)	0.063 (0.123)	0.117 (0.215)	0.159 (0.044)	0.418 (0.131)	124.052 (10.034)
BR	0.509 (0.089)	0.758 (0.148)	0.053 (0.024)	0.114 (0.173)	<i>0.029</i> (0.112)	0.154 (0.054)	0.478 (0.121)	124.566 (10.589)
CA	0.495 (0.098)	0.736 (0.159)	0.056 (0.028)	0.126 (0.180)	0.117 (0.222)	0.153 (0.048)	0.441 (0.128)	123.161 (8.588)
CH	0.518 (0.106)	0.706 (0.169)	0.053 (0.025)	0.161 (0.197)	0.134 (0.251)	0.142 (0.037)	0.442 (0.140)	122.438 (8.510)
CL	0.495 (0.099)	0.769 (0.136)	0.054 (0.022)	0.091 (0.155)	0.072 (0.170)	0.151 (0.041)	0.455 (0.131)	124.367 (7.929)
CN	0.502 (0.118)	0.643 (0.197)	0.051 (0.041)	0.232 (0.249)	0.153 (0.279)	0.145 (0.074)	0.393 (0.153)	121.190 (13.016)
CO	0.532 (0.097)	0.755 (0.129)	0.050 (0.017)	0.099 (0.154)	0.073 (0.169)	0.142 (0.036)	<b>0.486</b> (0.141)	123.085 (7.644)
CZ	0.487 (0.097)	0.769 (0.154)	0.057 (0.024)	0.094 (0.166)	0.139 (0.235)	0.157 (0.051)	0.418 (0.137)	123.901 (8.317)
DE	0.502 (0.110)	0.776 (0.154)	0.063 (0.039)	0.094 (0.166)	0.114 (0.227)	0.158 (0.048)	0.445 (0.138)	124.570 (9.937)
DK	0.524 (0.099)	0.701 (0.172)	0.052 (0.026)	0.161 (0.203)	0.107 (0.220)	0.147 (0.059)	0.445 (0.125)	121.128 (8.498)
EE	0.504 (0.095)	0.755 (0.144)	0.056 (0.028)	0.091 (0.151)	0.147 (0.246)	0.147 (0.037)	0.428 (0.124)	124.531 (10.383)
ES	0.514 (0.101)	0.733 (0.163)	0.052 (0.023)	0.141 (0.196)	0.085 (0.194)	0.148 (0.038)	0.474 (0.136)	123.432 (8.257)
FI	0.487 (0.103)	<b>0.806</b> (0.132)	0.062 (0.032)	<i>0.062</i> (0.131)	0.122 (0.219)	<b>0.166</b> (0.042)	0.428 (0.136)	123.707 (8.277)
FR	<b>0.533</b> (0.113)	0.704 (0.159)	0.057 (0.035)	0.152 (0.193)	0.152 (0.249)	0.144 (0.046)	0.452 (0.145)	120.900 (9.452)
GR	0.473 (0.091)	0.709 (0.161)	0.049 (0.020)	0.124 (0.193)	0.198 (0.267)	0.144 (0.033)	0.397 (0.127)	121.519 (8.147)
HR	0.473 (0.101)	0.752 (0.157)	0.056 (0.026)	0.110 (0.165)	0.158 (0.245)	0.151 (0.038)	0.418 (0.132)	122.991 (8.289)
HU	0.494 (0.116)	0.800 (0.144)	0.064 (0.033)	0.066 (0.140)	0.189 (0.283)	0.162 (0.045)	0.408 (0.146)	124.793 (10.081)
ID	0.510 (0.089)	0.716 (0.165)	<i>0.048</i> (0.023)	0.150 (0.195)	0.040 (0.144)	0.147 (0.048)	0.448 (0.126)	123.762 (12.311)
IE	0.503 (0.092)	0.696 (0.174)	0.051 (0.024)	0.164 (0.211)	0.120 (0.222)	0.146 (0.040)	0.445 (0.125)	122.503 (8.780)
IN	0.487 (0.104)	0.704 (0.186)	0.053 (0.037)	0.158 (0.234)	0.143 (0.266)	0.145 (0.058)	0.398 (0.134)	121.598 (11.939)
IR	<i>0.455</i> (0.101)	<i>0.599</i> (0.215)	0.049 (0.031)	<b>0.278</b> (0.265)	0.181 (0.281)	<i>0.133</i> (0.038)	<i>0.298</i> (0.137)	<i>119.224</i> (12.176)
IT	0.501 (0.090)	0.705 (0.166)	0.051 (0.023)	0.158 (0.199)	0.085 (0.186)	0.144 (0.036)	0.444 (0.130)	122.752 (8.591)
JP	0.512 (0.102)	0.729 (0.189)	0.056 (0.032)	0.153 (0.220)	0.156 (0.268)	0.153 (0.060)	0.474 (0.159)	123.181 (13.594)
LT	0.477 (0.105)	0.750 (0.154)	0.054 (0.020)	0.097 (0.165)	0.182 (0.264)	0.146 (0.037)	0.393 (0.124)	122.687 (8.250)
LV	0.494 (0.099)	0.730 (0.172)	0.056 (0.033)	0.122 (0.192)	0.158 (0.263)	0.149 (0.046)	0.399 (0.125)	121.961 (12.291)
MX	0.529 (0.091)	0.757 (0.124)	0.051 (0.023)	0.091 (0.145)	0.079 (0.191)	0.146 (0.040)	0.485 (0.130)	124.044 (8.197)
NL	0.518 (0.100)	0.705 (0.171)	0.053 (0.029)	0.154 (0.202)	0.115 (0.235)	0.144 (0.040)	0.446 (0.130)	122.553 (9.230)
NO	0.507 (0.101)	0.710 (0.162)	0.052 (0.024)	0.147 (0.193)	0.117 (0.225)	0.145 (0.037)	0.435 (0.130)	122.500 (8.098)
NZ	0.486 (0.100)	0.771 (0.144)	0.059 (0.026)	0.085 (0.154)	0.136 (0.252)	0.158 (0.044)	0.432 (0.134)	124.857 (9.177)
PL	0.504 (0.102)	0.766 (0.145)	<b>0.065</b> (0.046)	0.093 (0.155)	0.099 (0.208)	0.154 (0.048)	0.436 (0.137)	122.569 (10.738)
PT	0.478 (0.107)	0.736 (0.178)	0.056 (0.028)	0.129 (0.203)	0.145 (0.241)	0.150 (0.041)	0.407 (0.132)	122.887 (9.709)
RO	0.476 (0.113)	0.720 (0.166)	0.053 (0.023)	0.121 (0.184)	<b>0.224</b> (0.285)	0.142 (0.034)	0.373 (0.139)	121.389 (7.864)
RS	0.499 (0.119)	0.745 (0.154)	0.059 (0.034)	0.102 (0.167)	0.139 (0.240)	0.151 (0.041)	0.424 (0.143)	121.517 (8.257)
RU	0.485 (0.099)	0.790 (0.146)	0.061 (0.032)	0.071 (0.149)	0.141 (0.247)	0.161 (0.049)	0.415 (0.136)	124.464 (10.373)
SE	0.512 (0.096)	0.725 (0.159)	0.053 (0.028)	0.138 (0.185)	0.115 (0.227)	0.147 (0.036)	0.454 (0.123)	123.027 (7.834)
SK	0.479 (0.103)	0.755 (0.172)	0.064 (0.040)	0.109 (0.178)	0.184 (0.263)	0.156 (0.040)	0.381 (0.136)	122.172 (9.100)
TR	0.498 (0.095)	0.669 (0.184)	0.049 (0.023)	0.199 (0.228)	0.128 (0.238)	0.137 (0.040)	0.398 (0.125)	119.935 (9.252)
UK	0.512 (0.096)	0.723 (0.163)	0.054 (0.027)	0.134 (0.192)	0.110 (0.227)	0.148 (0.041)	0.465 (0.128)	123.424 (9.642)
US	0.507 (0.100)	0.721 (0.163)	0.057 (0.044)	0.140 (0.194)	0.108 (0.221)	0.150 (0.049)	0.461 (0.130)	122.624 (9.813)
VE	0.515 (0.101)	0.777 (0.113)	0.054 (0.022)	0.070 (0.120)	0.082 (0.198)	0.151 (0.042)	0.476 (0.152)	<b>124.961</b> (10.287)

**Table 7:** Spearman rank-order correlations between users' acoustic properties of listening behavior and cultural features (Hofstede). Correlations  $>0.1$  are highlighted in **bold** face. Statistically significant correlations at  $p < 0.001$  are marked with an asterisk (\*).

	PD	IDV	MAS	UA	LTO	IND
Danceability	-0.035*	0.044*	0.023*	-0.052*	-0.024*	0.072*
Energy	0.056*	-0.102*	-0.014	<b>0.116*</b>	0.076*	<b>-0.115*</b>
Speechiness	0.022*	-0.034*	0.016*	0.085*	0.065*	-0.096*
Acousticness	-0.056*	<b>0.105*</b>	0.026*	<b>-0.122*</b>	-0.086*	<b>0.125*</b>
Instrumentalness	-0.012	0.011	-0.029*	0.038*	0.055*	-0.055*
Liveness	0.021*	-0.042*	-0.014	0.059*	0.035*	-0.065*
Valence	-0.042*	0.059*	0.047*	-0.076*	-0.063*	<b>0.114*</b>
Tempo	0.009	-0.041*	0.008	0.031*	0.043*	-0.025*

**Table 8:** Spearman rank-order correlations between users' acoustic properties of listening behavior and socio-economic features (WHR). Correlations  $>0.1$  are highlighted in **bold** face. Statistically significant correlations at  $p < 0.001$  are marked with an asterisk (\*).

	Happiness	GDP	Social Sup.	Life Exp.	Freedom	Trust	Generosity
Danceability	0.035*	0.036*	-0.010	0.049*	0.037*	0.051*	0.052*
Energy	-0.036*	-0.067*	0.056*	-0.056*	-0.026*	-0.033*	<b>-0.101*</b>
Speechiness	-0.018*	-0.007	0.059*	-0.017*	0.011	-0.004	-0.067*
Acousticness	0.055*	0.079*	-0.046*	0.070*	0.039*	0.048*	<b>0.118*</b>
Instrumentalness	-0.031*	0.030*	0.042*	0.040*	0.006	0.001	-0.044*
Liveness	0.005	-0.019*	0.056*	-0.030*	0.001	-0.008	-0.048*
Valence	0.071*	0.047*	0.008	0.051*	0.044*	0.064*	0.084*
Tempo	0.004	-0.025*	0.046*	-0.015*	0.001	0.003	-0.016*

to listen to music that has a higher probability of being acoustic and that evokes positive emotions, which makes sense. At the same time, indulging populations tend to prefer lower energy levels in music (correlation of  $-0.115$ ), which hints at a preference for more relaxing music. Furthermore, uncertainty avoidance (UA) is positively correlated with music energy level (0.116), but negatively with acousticness ( $-0.122$ ). Societies characterized by stiff codes and laws therefore tend to prefer more energetic music, but lower amounts of acoustic tracks. Also, there is a positive correlation between individualism (IDV) and acousticness (0.105).

Comparing the acoustic features with the WHR dimensions, cf. **Table 8**, we can only observe two correlations exceeding the threshold. Both relate to the aspect of generosity. More precisely, we see a positive correlation between generosity and acousticness (0.118), whereas a negative one with energy ( $-0.101$ ). More generous populations therefore tend to prefer less energetic music, with a more acoustic sound.

## 8 Conclusion and Future Work

The contributions of this work are two-fold: (i) we introduced a novel music-cultural user model that jointly relies on acoustic song features and culture-related features

to describe the user's musical preferences and cultural background and (ii) we proposed a recommender system that leverages these features as contextual information. Our evaluations based on a dataset comprising more than 55,000 users showed that the proposed user model is able to outperform models that incorporate either solely musical aspects or cultural aspects and the evaluated baseline methods (relying on user's country as a proxy for culture, utilizing solely the user's and track's identifiers). In regard to both recall and precision, we show that adding contextual information obtained via incorporating audio features of tracks, data extracted from the World Happiness Report and Hofstede's cultural dimensions, contributes to improved recommendations when compared to the baseline approaches. Particularly, we find that a combination of acoustic features of the songs a user listened to (describing the individual music preferences of a user) and the World Happiness Report as a description of the cultural/socio-economic background of the user performs best.

Future work includes extending the user models with further data utilized for capturing cultural aspects of users (e.g., the Quality of Government dataset (Dahlberg et al., 2016)). Moreover, we are particularly interested in analyzing the country-specific influence of each of the

individual features of the proposed user models on the overall recommendation performance to get a deeper understanding for cohesive features that constitute listening patterns. Regarding the representation of both the musical preferences and cultural aspects, we plan to investigate more sophisticated modeling approaches. Particularly regarding the representation of musical preferences of users, we believe that, e.g., using Gaussian mixture models will allow for a more differentiated representation of users and their (possibly diverse and broad) preferences. Finally, we aim to transcend the country level for our culture-based analyses, e.g., focusing on culturally similar users that live in the same cultural region (but not necessarily in the same country).

## Notes

- <sup>1</sup> A listening event is defined as a quintuple <user, artist, album, track, timestamp>.
- <sup>2</sup> <https://www.last.fm>.
- <sup>3</sup> <https://developer.spotify.com/web-api/search-item/>.
- <sup>4</sup> <https://developer.spotify.com/web-api/get-several-audio-features/>.
- <sup>5</sup> To foster further research, we provide the dataset at <https://doi.org/10.5281/zenodo.3477842>.
- <sup>6</sup> <https://www.iso.org/iso-3166-country-codes.html>.
- <sup>7</sup> A description of these features and the API can be found at <https://developer.spotify.com/web-api/get-several-audio-features/>.
- <sup>8</sup> <https://www.hofstede-insights.com/models/national-culture/>.
- <sup>9</sup> <http://worldhappiness.report/>.
- <sup>10</sup> Please note that this distinction between the positive and negative class is also utilized by XGBoost for binary classification tasks based on logistic regression.

## Competing Interests

The authors have no competing interests to declare.

## References

- Adomavicius, G., & Tuzhilin, A.** (2011). Context-aware recommender systems. In *Recommender Systems Handbook*, pages 217–253. Springer, New York, NY, USA. DOI: [https://doi.org/10.1007/978-0-387-85820-3\\_7](https://doi.org/10.1007/978-0-387-85820-3_7)
- Andersen, J. S.** (2014). Using the Echo Nest's automatically extracted music features for a musicological purpose. In *4th International Workshop on Cognitive Information Processing (CIP)*, pages 1–6.
- Ankolekar, A., & Sandholm, T.** (2011). Foxtrot: A soundtrack for where you are. In *IWS '11 Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*, pages 26–31. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2019335.2019341>
- Ayaki, T., Yanagimoto, H., & Yoshioka, M.** (2017). Recommendation from access logs with ensemble learning. *Artificial Life and Robotics*, 22(2), 163–167. DOI: <https://doi.org/10.1007/s10015-016-0346-x>
- Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Lüke, K.-H., & Schwaiger, R.** (2011a). InCarMusic: Context-aware music recommendations in a car. In *International Conference on Electronic Commerce and Web Technologies*. DOI: [https://doi.org/10.1007/978-3-642-23014-1\\_8](https://doi.org/10.1007/978-3-642-23014-1_8)
- Baltrunas, L., Ludwig, B., & Ricci, F.** (2011b). Matrix factorization techniques for context-aware recommendation. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, pages 301–304. ACM. DOI: <https://doi.org/10.1145/2043932.2043988>
- Bellogin, A., Castells, P., & Cantador, I.** (2011). Precision-oriented evaluation of recommender systems: An algorithmic comparison. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, pages 333–336. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2043932.2043996>
- Bollen, D., Knijnenburg, B. P., Willemsen, M. C., & Graus, M.** (2010). Understanding choice overload in recommender systems. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, pages 63–70. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/1864708.1864724>
- Braunhofer, M., Kaminskas, M., & Ricci, F.** (2011). Recommending music for places of interest in a mobile travel guide. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, pages 253–256. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2043932.2043977>
- Braunhofer, M., Kaminskas, M., & Ricci, F.** (2013). Location-aware music recommendation. *International Journal of Multimedia Information Retrieval*, 2(1), 31–44. DOI: <https://doi.org/10.1007/s13735-012-0032-2>
- Breese, J. S., Heckerman, D., & Kadie, C.** (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 43–52. San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M.** (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96, 668–696. DOI: <https://doi.org/10.1109/JPROC.2008.916370>
- Celma, O.** (2010). *Music Recommendation and Discovery: The Long Tail, Long Tail, and Long Play in the Digital Music Space*. Springer Publishing, 1st edition.
- Chen, T., & Guestrin, C.** (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22<sup>nd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2939672.2939785>
- Cheng, Z., & Shen, J.** (2014). Just-for-me: An adaptive personalization system for location-aware social music recommendation. In *Proceedings of the 2014 ACM International Conference on Multimedia Retrieval*, pages 1267–1268. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2600428.2611187>
- Cremonesi, P., Koren, Y., & Turrin, R.** (2010). Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the Fourth*

- ACM Conference on Recommender Systems, pages 39–46. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/1864708.1864721>
- Cremonesi, P., Turrin, R., Lentini, E., & Matteucci, M.** (2008). An evaluation methodology for collaborative recommender systems. In *2008 International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution*, pages 224–231. DOI: <https://doi.org/10.1109/AXMEDIS.2008.13>
- Dahlberg, S., Holmberg, S., Rothstein, B., Khomenko, A., & Svensson, R.** (2016). *Quality of Government (QoG) Basic Dataset 2016*. The Quality of Government Institute, University of Gothenburg.
- Diener, E.** (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1), 34. DOI: <https://doi.org/10.1037/0003-066X.55.1.34>
- Dror, G., Koenigstein, N., Koren, Y., & Weimer, M.** (2012). The yahoo! music dataset and KDDcup' 11. In *Proceedings of KDD Cup 2011 Competition, JMLR Proceedings*, volume 18, pages 3–18. JMLR.org.
- Elkahky, A. M., Song, Y., & He, X.** (2015). A multiview deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*, pages 278–288. International World Wide Web Conferences Steering Committee. DOI: <https://doi.org/10.1145/2736277.2741667>
- Ferwerda, B., & Schedl, M.** (2016). Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A crosscountry analysis. In *Proceedings of the 24th International Conference on User Modeling, Adaptation and Personalization: Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems*.
- Hauger, D., Schedl, M., Kosir, A., & Tkalcic, M.** (2013). The Million Musical Tweet Dataset: What we can learn from microblogs. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*.
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S.** (2017). Neural collaborative filtering. In *Proceedings of the 26th International Conference on WorldWideWeb*, pages 173–182. Geneva, Switzerland. International World Wide Web Conferences Steering Committee. DOI: <https://doi.org/10.1145/3038912.3052569>
- Helliwell, J. F., Layard, R., & Sachs, J.** (2016). *World Happiness Report*. Sustainable Development Solutions Network.
- Hofstede, G. H.** (1980). *Culture's Consequences: International Differences in Work-Related Values*. Sage Publications, Beverly Hills, CA.
- Hofstede, G., Hofstede, G. J., & Minkov, M.** (1991). *Cultures and Organizations: Software of the Mind*, volume 2. McGraw-Hill.
- Hu, Y., Koren, Y., & Volinsky, C.** (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 263–272. Washington, DC, USA. IEEE Computer Society. DOI: <https://doi.org/10.1109/ICDM.2008.22>
- Hu, Y., & Ogihara, M.** (2011). Nexttone player: A music recommendation system based on user behavior. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*. Miami, FL, USA.
- Kaminskas, M., Fernández-Tobías, I., Ricci, F., & Cantador, I.** (2012). Knowledge-based Music Retrieval for Places of Interest. In *Proceedings of the Second International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, pages 19–24. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2390848.2390854>
- Kaminskas, M., & Ricci, F.** (2012). Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 6(2), 89–119. DOI: <https://doi.org/10.1016/j.cosrev.2012.04.002>
- Kaminskas, M., Ricci, F., & Schedl, M.** (2013). Location-aware music recommendation using auto-tagging and hybrid matching. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 17–24. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2507157.2507180>
- Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N.** (2010). Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, pages 79–86. ACM. DOI: <https://doi.org/10.1145/1864708.1864727>
- Kim, J.-Y., & Belkin, N. J.** (2002). Categories of music description and search terms and phrases used by non-music experts. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, volume 2, pages 209–214.
- Knees, P., & Schedl, M.** (2013). A survey of music similarity and recommendation from music context data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 10(1). DOI: <https://doi.org/10.1145/2542205.2542206>
- Knees, P., & Schedl, M.** (2016). *Music Similarity and Retrieval – An Introduction to Audio- and Webbased Strategies*. Springer, Berlin and Heidelberg, Germany. DOI: [https://doi.org/10.1007/978-3-662-49722-7\\_1](https://doi.org/10.1007/978-3-662-49722-7_1)
- Lee, J. H., & Downie, J. S.** (2004). Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In *Proceedings of the 5th International Conference on Music Information Retrieval*, volume 2004.
- Lays, C., Ley, C., Klein, O., Bernard, P., & Licata, L.** (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. DOI: <https://doi.org/10.1016/j.jesp.2013.03.013>
- Liu, M., Hu, X., & Schedl, M.** (2017). Artist preferences and cultural, socio-economic distances across countries: A big data perspective. In *Proceedings of*

- the 18th International Society for Music Information Retrieval Conference*, pages 103–111.
- Liu, M., Hu, X., & Schedl, M.** (2018). The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study. *PLOS ONE*, 13(12), 1–29. DOI: <https://doi.org/10.1371/journal.pone.0208186>
- Logan, B.** (2002). Content-based playlist generation: Exploratory experiments. In *Proceedings of the 3<sup>rd</sup> International Conference on Music Information Retrieval*, pages 295–296.
- McVicar, M., Freeman, T., & De Bie, T.** (2011). Mining the correlation between lyrical and audio features and the emergence of mood. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 783–788.
- Miotto, R., Barrington, L., & Lanckriet, G.** (2010). Improving Auto-tagging by Modeling Semantic Co-occurrences. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*.
- Pacuk, A., Sankowski, P., Wegrzycki, K., Witkowski, A., & Wygocki, P.** (2016). RecSys Challenge 2016: Job recommendations based on preselection of offers and gradient boosting. In *Proceedings of the Recommender Systems Challenge*, pages 10:1–10:4, New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2987538.2987544>
- Pan, R., Zhou, Y., Cao, B., Liu, N. N., Lukose, R., Scholz, M., & Yang, Q.** (2008). One-class collaborative filtering. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 502–511. Piscataway, NJ, USA. IEEE. DOI: <https://doi.org/10.1109/ICDM.2008.16>
- Pichl, M., & Zangerle, E.** (2018). Latent feature combination for multi-context music recommendation. In *Proceedings of the Conference on Content-Based Multimedia Indexing*. IEEE. DOI: <https://doi.org/10.1109/CBMI.2018.8516495>
- Pichl, M., Zangerle, E., & Specht, G.** (2016). Understanding playlist creation on music streaming platforms. In *IEEE International Symposium on Multimedia*, pages 475–480. IEEE Computer Society. DOI: <https://doi.org/10.1109/ISM.2016.0107>
- Pichl, M., Zangerle, E., Specht, G., & Schedl, M.** (2017). Mining culture-specific music listening behavior from social media data. In *IEEE International Symposium on Multimedia*, pages 208–215. IEEE Computer Society. DOI: <https://doi.org/10.1109/ISM.2017.35>
- Schedl, M.** (2013). Leveraging microblogs for spatiotemporal music information retrieval. In *European Conference on Information Retrieval*, pages 796–799. Springer. DOI: [https://doi.org/10.1007/978-3-642-36973-5\\_87](https://doi.org/10.1007/978-3-642-36973-5_87)
- Schedl, M.** (2016). The LFM-1b dataset for music retrieval and recommendation. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, pages 103–110. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2911996.2912004>
- Schedl, M.** (2017). Investigating country-specific music preferences and music recommendation algorithms with the LFM-1b dataset. *International Journal on Multimedia Information Retrieval*, 6(1), 71–84. DOI: <https://doi.org/10.1007/s13735-017-0118-y>
- Schedl, M.** (2019). Deep Learning in Music Recommendation Systems. *Frontiers in Applied Mathematics and Statistics*, 5, 44. DOI: <https://doi.org/10.3389/fams.2019.00044>
- Schedl, M., Lemmerich, F., Ferwerda, B., Skowron, M., & Knees, P.** (2017). Indicators of country similarity in terms of music taste, cultural, and socio economic factors. In *Proceedings of the 19th IEEE International Symposium on Multimedia*. DOI: <https://doi.org/10.1109/ISM.2017.55>
- Schedl, M., & Schnitzer, D.** (2013). Hybrid retrieval approaches to geospatial music recommendation. In *Proceedings of the 35th Annual International Conference on Research and Development in Information Retrieval*, pages 793–796. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2484028.2484146>
- Schedl, M., & Schnitzer, D.** (2014). Location-aware music artist recommendation. In *Proceedings of the 20th International Conference on MultiMedia Modeling*, pages 205–213. Springer. DOI: [https://doi.org/10.1007/978-3-319-04117-9\\_19](https://doi.org/10.1007/978-3-319-04117-9_19)
- Schedl, M., Vall, A., & Farrahi, K.** (2014). User geospatial context for music recommendation in microblogs. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 987–990. New York, NY, USA. ACM. DOI: <https://doi.org/10.1145/2600428.2609491>
- Schimmack, U., Radhakrishnan, P., Oishi, S., Dzokoto, V., & Ahadi, S.** (2002). Culture, personality, and subjective well-being: Integrating process models of life satisfaction. *Journal of Personality and Social Psychology*, 82(4), 582. DOI: <https://doi.org/10.1037/0022-3514.82.4.582>
- Tran, N. K.** (2016). Classification and learning-to-rank approaches for cross-device matching at CIKM Cup 2016. *arXiv preprint arXiv:1612.07117*.
- Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G.** (2008). Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2), 467–476. DOI: <https://doi.org/10.1109/TASL.2007.913750>
- Viglienconi, G., & Fujinaga, I.** (2017). The music listening histories dataset. In *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pages 96–102.
- Wang, X., Rosenblum, D., & Wang, Y.** (2012a). Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM International Conference on Multimedia*, pages 99–108. ACM. DOI: <https://doi.org/10.1145/2393347.2393368>
- Wang, X., Rosenblum, D., & Wang, Y.** (2012b). Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM International Conference on Multimedia*, pages 99–108. New York,

NY, USA. ACM. DOI: <https://doi.org/10.1145/2393347.2393368>

**Zangerle, E., & Pichl, M.** (2018). The many faces of users: Modeling musical preference. In *Proceedings of the 19th International Society for Music Information Retrieval Conference*, pages 709–716.

**Zangerle, E., Pichl, M., Gassler, W., & Specht, G.** (2014). #nowplaying music dataset: Extracting listening behavior from twitter. In *Proceedings of the 1st International Workshop on Internet-Scale Multimedia Management*, pages 21–26. DOI: <https://doi.org/10.1145/2661714.2661719>

**How to cite this article:** Zangerle, E., Pichl, M., & Schedl, M. (2020). User Models for Culture-Aware Music Recommendation: Fusing Acoustic and Cultural Cues. *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 1–16. DOI: <https://doi.org/10.5334/tismir.37>

**Submitted:** 04 June 2019

**Accepted:** 15 November 2019

**Published:** 05 March 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

**]u[** *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 