Mining Culture-Specific Music Listening Behavior from Social Media Data

Martin Pichl, Eva Zangerle, Günther Specht
Databases and Information Systems
Department of Computer Science, University of Innsbruck, Austria
Email: {martin.pichl, eva.zangerle, guenther.specht}@uibk.ac.at

Markus Schedl
Department of Computational Perception
Johannes Kepler University Linz, Austria
Email: markus.schedl@jku.at

***Abstract*—Incorporating user characteristics and contextual information has shown to be essential when it comes to personalized music retrieval and recommendation. To this end, the current location of a user is often exploited. However, relying solely on GPS coordinates neglects the cultural background of users, which does not necessarily coincide with political borders. In this paper, we analyze culture-specific music listening behavior based on a dataset of 2,724 Spotify users, 62,104 distinct tracks and 104,390 listening events by modeling users jointly via their musical preferences and cultural characteristics. By applying a density-based spatial clustering algorithm, we identify nine clusters which reflect similar users regarding both their musical preference and cultural background. Our findings show that cultural aspects cannot be approximated by GPS coordinates and that incorporating cultural characteristics allows for more precise user characterization. Also, we observe that listening patterns occur on two different levels: we observe country-specific listening patterns as well as cross-country listening patterns that span across several countries.**

I. Introduction

Over the last decade, the way people consume music has changed fundamentally. An ever increasing amount of music is available and being listened to via streaming services and on mobile devices. Given that most of the latter are equipped with localization techniques, such as GPS, geolocalized listening profiles are nowadays available for a substantial amount of social media users who share their listening preferences and habits via social media platforms like Twitter or Facebook and streaming platforms like Spotify or Last.fm. At the same time, the importance of considering user characteristics and context to build personalized music retrieval and recommender systems is widely agreed on (e.g., [1], [2]). The music items retrieved or recommended by such systems are arguably better tailored to the listeners' needs than those uncovered by traditional collaborative filtering or content-based algorithms.

While location plays an important role to describe a listener's context, the use of raw GPS coordinates may be misleading as they do not necessarily reflect differences in culture. Even worse, exploiting GPS coordinates to model *similarity* between listeners, which is key to build recommender systems, leads to systems that are agnostic to cultural characteristics as geographically close users might have a very different cultural background. A common solution to this problem is to map GPS coordinates to countries. However, the underlying assumption that cultural groups coincide with political borders neglects the ethnic fractionalization present within and beyond many countries. Therefore, a measure that integrates musical similarity and cultural similarity beyond countries' geographical borders is called for.

In this paper, we propose a novel approach to model listener similarity by integrating two dimensions: (i) personal listening habits mined from social media and described by acoustic properties and (ii) cultural characteristics based on socio-economic and cultural factors from a publicly available data source. Employing this model, we subsequently address the following research questions:

1) How can we find culture-specific music listening patterns among users?
2) To which extent do a user's musical preferences, a user's cultural embedding and a user's geographical location influence the proposed model?
3) What are the characteristics of the identified cultural groups in terms of music taste?

Please note that while we evaluate the usage of GPS coordinates for locating users precisely in comparison to utilizing the user's country (derived from GPS coordinates), we refrain from going to the sub-country level regarding cultural and socio-economic data at this stage, due to the sparsity of data. We leave this for future work.

To the best of our knowledge, this is the first study modeling user similarity based on two dimensions: (i) personal listening habits and musical preferences and (ii) cultural characteristics extracted from socio-economic data. We believe that our findings in regards to culture-specific listening behavior can advance music information retrieval and music recommender systems as these allow for fine-grained characterization of users and their preferences, therefore contributing to improved personalization capabilities. Our analyses show that a combination of a user's cultural embedding irrespective of country borders (as additional contextual information) and his/her listening preferences contributes to a more precise user characterization, while further complementing the user profiles with GPS coordinates of the user's location does not. This implies that modeling a user's cultural embedding is an important facet of music information retrieval and that culture can be hardly approximated by a user's exact geolocation. As for the resulting clusters, the characteristics of the identified

978-1-5386-2937-6/17 $31.00 © 2017 IEEE
DOI 10.1109/ISM.2017.35

cultural groups show that there are country-specific listening patters as well as listening patterns that span across several countries, each with distinctive characteristics both regarding the cultural and musical dimension. We believe that this study provides new insights for the development of culture-aware retrieval and recommender systems as these allow for complementing broader listening patterns with country-specific listening patterns for countries where users exhibit particular listening behaviors.

The remainder of this paper is structured as follows. Section II introduces related work and Section III presents the data underlying our analyses. Section IV details the methods utilized for the analyses. Section V presents and discusses the results obtained, while Section VI concludes the paper.

II. Related Work

The work presented in this paper is related to location-specific music information retrieval and music recommender systems. There are several approaches that exploit places of interest, where the idea is to recommend music that suits the environment—in an emotional or cultural sense [3], [4]. Rich sensory devices allow to map a certain location or a trace of locations to a certain activity which can be exploited to provide personalized music recommendations based on the location of a user during the day [5] or even during driving a car [6]. Also, there exist approaches based on user similarity utilizing the geodesic distance between two users [7], where the distance is subsequently incorporated in collaborative filtering recommendation approaches. Furthermore, the cultural distance of users can also be approximated by the country or continent a user is located in [8]. The authors of [9] conclude that if users listen to various different artists, the integration of geospatial information is beneficial. In a later study, they state that countries or continents as geographic entities do not necessarily reflect cultural borders [7]. Moreover, visualizing artist and genre distributions on interactive maps allows researchers explore regional listening patterns [10].

Moore et al. [11] utilized probabilistic embedding methods to capture a joint space of musical taste and geographic information. The authors performed this analysis based on the MMTD (Million Musical Tweets Dataset) by Hauger and Schedl [12]. For the geographic dimension, they relied on the city-level and for the musical (taste) dimension, they relied on the artist level. This allowed them to analyze, for instance, similarities of cities and artists. In the joint space, they find a substantial segmentation of Brazilian, Southeast Asian and American cities. Also, Moore et al. find that a segmentation of genres is clearly visible in the joint space. Furthermore, by applying k-means clustering on a higher-dimensional model they found a very tight French-speaking cluster.

In contrast to the work by Moore et al., the approach presented in this work enriches the analyses with detailed data for both dimensions: we enrich the musical (taste) dimension with content features for each of the tracks and also enrich the geographical model with socio-economic data (on a per-country level). We argue that while an embedding approach allows for a direct comparison of elements (e.g., cities and artists) in the joint space, our proposed approach yields a joint representation of users in an enriched musical and socio-economical space to further investigate the influence of musical and socio-economical features on the similarity of users.

We are not aware of any work aggregating music listening preferences of countries based by their cultural—opposed to geographic—distance and hence, locate a research gap here. We believe that by understanding the cultural embedding of a user, music recommendation and music retrieval approaches can be improved as the gained understanding allows for an improved user model which subsequently enables better personalization.

III. Data

For our analyses, we require information about: (i) music listening behavior of users and their geolocation and (ii) cultural characteristics of countries.

A. Listening Behavior

As the main data source, we use the Spotify playlist dataset we gathered in previous works [13], [14] as one representative of data gathered from streaming platforms. This publicly available dataset contains 143,528 playlists created by 15,345 unique users who listened to 1,878,457 tracks. To detect the location of users, we exploit that Spotify provides the means to share the tracks a user is currently listening to on Twitter (among other social media platforms) and that people often send tweets containing geolocation information via their GPS-enabled mobile devices. Consequently, we search for Spotify user names on Twitter, which allows us to crawl geo-locatable tweets of each user, using only exactly matching user names to reduce the number of false positive matches. A second measure applied to prevent false positives is comparing the last #nowplaying tweets of the user (holding information about the music the user listened to) to the contents of his/her Spotify playlists. If we can find the according tracks in the playlists of the user, we assume that we correctly matched the user's Spotify and Twitter handles. With this approach, we were able to match 22.73% of all user names contained in the original dataset. To map each user to a distinct position, we neglect location shifts beyond the first decimal point of the longitude and latitude values. By analyzing the GPS coordinates of the tweets a user shares, we observe that 80% of the users in the dataset constantly tweet from the same area (i.e., no location shifts beyond the first decimal). To determine the location for the remainder of users, we apply a majority voting approach based on grid rectangles and consider the rectangle in which most tweets were sent as the user's location. We choose the size of the rectangles as 11.1 km x 7 km, as these capture changes of the first decimal of the GPS position. As a result, we can determine a unique country for 2,872 of the 3,335 users and remove the remaining users from the dataset. We could not determine a location for these users as some of the coordinates are located above sea. This could have several

Country	Users	Tracks	TPU	SD
United States	1,131	35,560	31.44	124.51
Spain	417	16,133	38.69	81.12
United Kingdom	279	8,708	31.21	84.62
Mexico	233	15,073	64.69	84.01
Netherlands	91	1,881	20.67	36.10
Sweden	84	2,031	24.18	41.65
France	73	1,533	21.00	27.98
Italy	61	2,963	48.57	102.86
Germany	48	1,441	30.02	47.70
Chile	35	4,406	125.89	174.28

TABLE I: Top-10 countries (TPU=tracks per user, SD=standard deviation).

reasons: malfunctioning devices, devices sending fake GPS coordinates or people tweeting while traveling on ships and on airplanes. Furthermore, we restrict our dataset to countries with listening events of at least 10 distinct users (a total of 25 countries). The resulting dataset contains 104,390 listening events by 2,724 distinct users having listened to 62,104 distinct tracks. The top-10 countries with respect to the number of users are given in Table I.

While the dataset is rather small compared to other available datasets (e.g., [15]), it nevertheless contains the necessary information in sufficient volume. This is also backed by the fact that we found statistically significant differences in features of the different clusters using an analysis of variance (ANOVA), cf. Section V.

As for modeling personal listening taste, we gather content-based audio features for each of the tracks by querying the Spotify API.[1] These acoustic features are extracted from the audio signal of the individual tracks and represent high-level content descriptors for tracks. They include *danceability* (how suitable a track is for dancing), *energy* (perceived intensity and activity), *speechiness* (presence of spoken words in a track), *acousticness* (confidence whether track is acoustic), *instrumentalness* (prediction whether track contains no vocals), *tempo* (in beats per minute) and *valence* (musical positiveness conveyed). Except for tempo, all of these features are given in the range $[0, 1]$. Audio features gathered via the Spotify API have already been exploited by a number of other analyses (e.g., [16]–[18]). Furthermore, in previous work, we have already shown that these audio descriptors can be exploited for clustering tracks based on their audio features and subsequently, can contribute to improved context-aware music recommendations [14], [19].

B. Cultural and Socio-Economic Data

To complement our model with cultural and socio-economic characteristics of countries, we rely on the World Happiness Report (WHR) [20]. We argue that people's cognitive and affective evaluations of their daily life and hence, their subjective well-being [21] provide a good indicator for cultural aspects as these have been shown to be directly influenced by cultural factors [22]. The WHR provides a set of aggregated measures capturing the perceived happiness of 156 countries: *gdp* is the real gross domestic product per capita; *freedom* measures the freedom to make life choices, *healthy life expectancy* states the healthy life expectancy at birth in the country, *generosity* specifies whether people in a country are willing to spend money to a charity; *social support* states if people have people helping them if they need support (i.e., relatives or friends); *corruption* and *happiness* measure the perceived corruption and happiness of citizens.

IV. Methods

For our study, we model a user's personal music listening behavior along with his/her cultural embedding in a single feature vector. First, to model a *user's personal musical preferences* based on his/her listened tracks, we rely on the acoustic features presented in the previous section. For characterizing the musical preferences of each user, we compute the arithmetic mean for each of the acoustic features of the songs contained in the user's playlists.

For the approximation of the *cultural embedding of users*, we rely on the variables of the WHR as described in the previous section. We add these variables to the feature vector as we aim to find cultural listening patterns by computing cultural similarity between users based on these variables. We assume that these variables reflect cultural similarity better than the mere geographic similarity.

Finally, to homogenize values across all variables, we perform centering and scaling such that all elements of the vectors exhibit a mean of 0 and standard deviation of 1 for each of the acoustic and cultural variables. The feature vector representing a user then consists of two parts: a user's individual music preferences captured by acoustic features as well as the user's cultural embedding approximated by socio-economic aspects extracted from the WHR. Employing this user model, we cluster users based on their musical preferences and cultural background and analyze the resulting clusters along their characteristics.

To perform an exploratory data analysis, we obtain a two-dimensional embedding which represents users in the joint 16-dimensional feature vectors by applying t-SNE [23], a state-of-the-art dimension reduction method. Ideally, the latent features jointly represent the personal listening preferences and cultural embedding of the user. To further analyze the influence of the cultural and musical components of the user representation and to select relevant features, we perform a Principal Components Analyses (PCA) [24] complementary to the t-SNE dimension reduction on the set of user representations.

To compute groups of users sharing common listening patterns as well as a common cultural background, we rely on Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [25]. In preliminary experiments we found that this clustering method applied to the t-SNE dimensionality reduced data provides the best results in regards to the variance of the acoustic attributes, even better than applying DBSCAN

[1] A description of the features and the API can be found at https://developer.spotify.com/web-api/get-several-audio-features.

PC/Dimension	Explained Variance	WHR	AF
PC1	14.96%	92.81%	7.19%
PC2	11.36%	10.82%	89.18%
PC3	10.57%	84.84%	15.16%
PC4	9.41%	10.58%	89.42%
PC5	7.84%	18.82%	81.18%
PC6	7.54%	16.63%	83.37%
PC7	7.32%	13.34%	86.66%
PC8	6.27%	76.27%	23.73%
Sum	75.27%	—	—
Mean Impact	—	40.52%	59.49%

TABLE II: PCA: Explained variance for WHR features and acoustic features (AF); WHR and AF columns show loadings of PC with respective dimension.

PC/Dimension	Explained Variance	WHR	AF	GEO
PC1	13.78%	83.72%	5.76%	10.52%
PC2	10.24%	45.72%	38.38%	15.90%
PC3	9.99%	35.04%	55.10%	9.86%
PC4	8.34%	14.98%	84.63%	0.39%
PC5	7.49%	60.41%	12.63%	26.95%
PC6	6.96%	15.55%	80.93%	3.52%
PC7	6.68%	8.95%	87.13%	3.92%
PC8	6.49%	8.06%	87.80%	4.14%
PC9	5.95%	54.43%	13.92%	31.65%
Sum	75.92%	—	—	—
Mean Impact	—	36.21%	51.81%	11.87%

TABLE III: PCA: Explained variance for WHR features, acoustic features (AF) and GPS coordinates of users (GEO); WHR, AF and GEO columns show loadings of PC with respective dimension.

PC/Dimension	Explained Variance	AF	GEO
PC1	16.53%	95.63%	4.37%
PC2	13.78%	93.25%	6.75%
PC3	12.25%	18.93%	81.07%
PC4	11.41%	86.67%	13.33%
PC5	11.20%	73.36%	26.64%
PC6	10.69%	84.34%	15.66%
Sum	75.86%	—	—
Mean Impact	—	77.94%	22.06%

TABLE IV: PCA: Explained variance for acoustic features (AF) and GPS coordinates of users (GEO); AF and GEO columns show loadings of PC with respective dimension.

to the original feature matrix. As we aim to find cultural listening patterns across both underlying feature dimensions, we naturally strive to maximize the variance of the acoustic features between the clusters. The sum of all standard deviations (SD) of all acoustical attributes for DBSCAN is $SD = 4.25$ compared to k-means ($SD = 2.33$) and spectral clustering ($SD = 3.62$) and hence, we chose to utilize DBSCAN for the clustering of users. We moreover use the maximization of cluster variance for determining the DBSCAN parameters $minPts$ and ϵ. The first parameter, $minPts$, defines how many points have to be within the range ϵ, such that those points are considered as core points and form a cluster. Hence, in our setting, $minPts$ defines how many users have to be grouped inside the range ϵ by DBSCAN to actually form a cultural cluster. Accordingly, we tune the parameters of DBSCAN by maximizing cluster variance and as a result, set these parameters to $minPts = 20$ and $\epsilon = 2$ for the given dataset (cf. Section III).

V. Results and Discussion

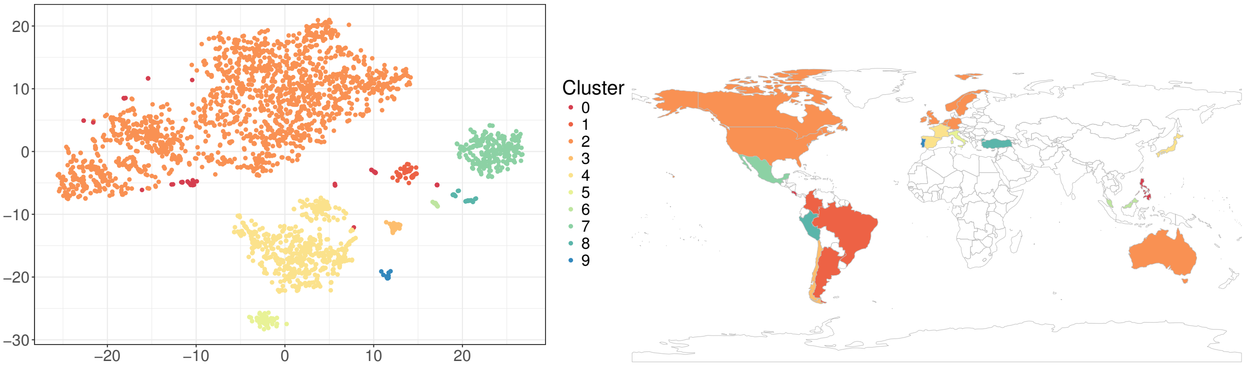
In the following section, we present and discuss the results of our analyses. Firstly, we examine the impact of individual features on the clustering. Subsequently, we present the results of the conducted user clustering. Next, we elaborate on the results of a correlation analysis of the features of both dimensions underlying our approach and discuss the patterns that are characteristic for certain countries.

A. User Models and Impact of Components

Figure 1 shows a two-dimensional representation of all users and their cluster assignments obtained by t-SNE with a perplexity parameter of 60. This perplexity parameter can be considered as the average number of neighbors and has to be determined empirically. Preliminary experiments showed that setting this parameter to 60 works best for reducing the dimension of the dataset from 14 dimensions based on seven acoustic and seven cultural features to two dimensions. From the t-SNE projection depicted in Figure 1, we can observe spherical groups of users (which is in line with our choice of DBSCAN as a clustering method). Applying DBSCAN provides us with a set of nine clusters and one additional group of noise points, which we refer to as cluster 0 in the subsequent plots.

To get a deeper understanding of the impact of the individual cultural and acoustic features, we perform a PCA. As we are interested in the influence of the cultural and acoustic features and also aim to evaluate the suitability of GPS coordinates for the localization of users (in contrast to mapping a user's GPS location to the country level), we perform PCA on a set of different user models: (i) user feature vectors as described in Section IV (holding cultural and musical features); (ii) user feature vectors holding both cultural and musical features complemented with longitude and latitude information of the location of each user; (iii) user feature vectors solely containing musical features and the longitude and latitude information and hence, neglecting any cultural features in this model.

For the conducted PCA, we set a minimum threshold of 75% explained variance, which is reached between the sixth and ninth PC depending on the model. In Table II we present the PCs for the user model containing cultural and acoustic data including the explained variance of each PC and the relative loadings of the cultural and acoustic features reflecting the feature's impact. We observe that the mean impact across all eight PCs of the WHR data is 41% and the mean impact of the acoustic features is 59%. Table III presents the PCA analyses for the user model complemented with GPS coordinates of



(a) t-SNE jointly applied to cultural and acoustic features.

(b) Countries and cluster assignments.

Fig. 1: Clustering results (same colors indicate same cluster assignments across all subfigures).

users. The average impact of cultural data (WHR) drops from 41% to 36% as the users in the same country feature the same WHR values and hence, a geographic similarity is implicitly covered while it is now explicitly covered by the mean impact of the GPS coordinates. I.e., the cultural components act as a proxy in this scenario. The geographic distance based on the GPS coordinates has an average impact of 12%. This is, as part of the information is already implicitly covered by the variables of the WHR. Particularly for small countries it holds that users originating from the same country have similar GPS coordinates and naturally, identical WHR variables. To complement this analysis, we also apply a PCA to a restricted dataset solely containing musical features and GPS coordinates. This analysis aims at getting a deeper understanding of the extent to which GPS coordinates may act as a proxy and allow for approximating cultural information in terms of user modeling. Therefore, in this analysis, we neglect any cultural features. The results of this PCA are shown in Table IV. We observe that in such a setting, the GPS coordinates explain 22% of the variance. Therefore, we argue that the geographic distance explains a substantially smaller fraction of the explained variance in comparison to the WHR data representing cultural aspects (41%). This is also reflected in the relative loadings of the PCs shown in Tables III and IV: In Table III, there is no loading in the geographic dimension that substantially influences any PC. Although 27% in PC5 is not small, it is less than half of the variation explained by the WHR (60%). In Table IV, solely PC3 is substantially influenced by the geographic dimension.

To conclude, we argue that in the current form, where cultural aspects are modeled on a country-level, adding a geographic user similarity in terms of GPS coordinates does not improve the result substantially. This is why we conduct the main analysis on the dataset comprising WHR data and acoustic features.

For examining the characteristics of the obtained clusters in terms of musical taste and cultural characteristics, we provide an interactive web interface which allows to compute clusters based on various clustering algorithms and settings and to interactively explore and visualize the obtained clusters and their characteristics.[2]

In the following, we first discuss the results of the clustering and subsequently focus on the individual characteristics of culture-specific listening patterns across and within individual countries.

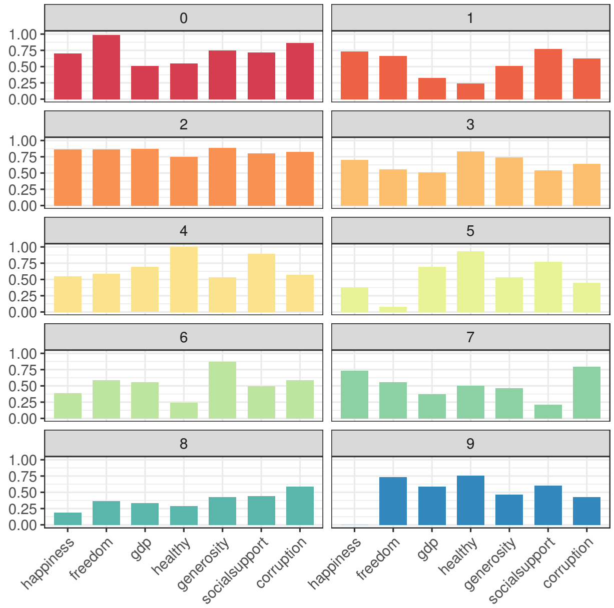
B. Country Analysis

In the following sections, we present the results obtained by applying the clustering method as described in Section IV. We firstly focus on the assignments of users to clusters, then look at the distribution of features among the clusters and subsequently, elaborate on the correlation of cultural as well as acoustic features and perform a deeper analysis of the characteristics of the obtained clusters.

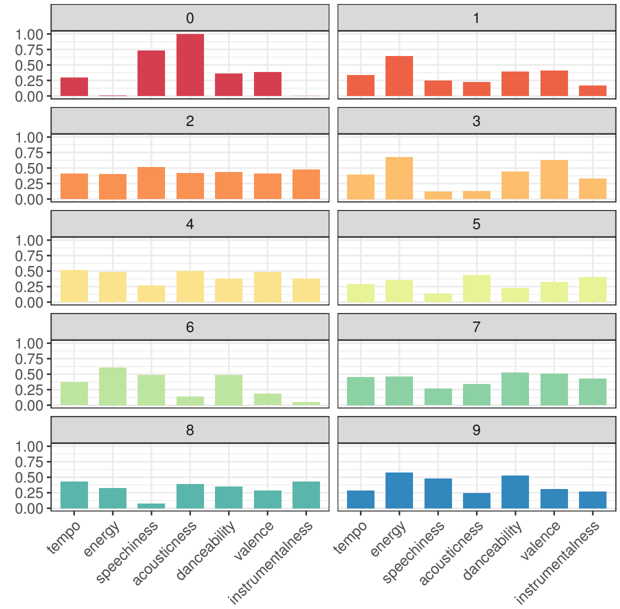
1) Country-Cluster Assignments: In Figure 1a, we depict the clusters resulting from applying DBSCAN on the user model containing both acoustic and cultural features. Moreover, we provide a world map depicting which countries belong to which cluster in Figure 1b. We indicate user-cluster assignments by individual colors in both subfigures. For countries where users belong to several clusters, we apply a majority vote, where the country is assigned to the cluster to which most of its users belong. Here we can observe that, e.g., countries in Northern Europe mostly share similar listening characteristics with Canada, the United States, and Australia. Furthermore, we find a South American cluster, but Chile and Peru forming own clusters.

2) Feature Distribution: To also get a deeper understanding of the individual acoustic and cultural characteristics of all clusters, we provide bar plots for all clusters in Figure 2. Particularly, Figure 2a depicts the average cultural aspects of the nine clusters detected and Figure 2b presents the average acoustic features and their distribution across all clusters. By

[2]http://dbis-mcc.uibk.ac.at



(a) Cultural characteristics of clusters.



(b) Musical characteristics of clusters.

Fig. 2: Detailed characteristics of clusters.

applying an ANOVA analysis using the individual feature vectors of the users in each cluster, we found that the differences between cluster means (compute among the contained users) are significant across all features (p-value < 0.01).

3) Feature Correlation and Cluster Analyses: To analyze listening patterns based on the proposed user model, we perform a correlation analysis of acoustic and cultural features. Using Pearson's correlation coefficient, we compare acoustic and cultural features and depict the obtained results in Figure 3. Besides several low positive and negative correlations between the acoustic and cultural features, we observe that happiness correlates well with valence (correlation coefficient $\rho = 0.61$). When inspecting clusters with respect to their valence and happiness values using the segment charts in Figure 2, clusters 1, 2, and 7 feature the highest values for both of these features. Cluster 1 groups users from Argentina, Brazil, and Columbia; cluster 2 groups users from Northern Europe, the United States, and Canada, whereas cluster 7 solely contains Mexican users. Hence, these countries (stemming from three different clusters) feature high happiness values and tend to listen to high-valence music.

Besides valence and happiness, in Figure 3, we additionally observe moderate correlation between healthiness and valence with $\rho = 0.46$. This pattern is found mainly among user in cluster 2 and hence found among users from Nothern Europe. In other parts of the world we detect particularly high danceability values, which surprisingly do not correlate with happiness and only slightly correlate with valence ($\rho = 0.11$). Danceability is particular high clusters 7 and 9. Cluster 7 contains users solely from Mexico, whereas cluster 9 solely contains Portuguese users. Portuguese users have similar listening patterns as other western users, however additionally consume music characterized by high danceability.

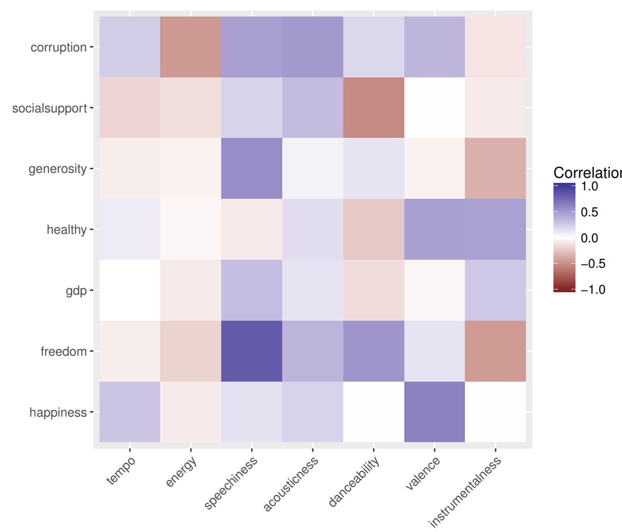


Fig. 3: Correlations between cultural (y-axis) and acoustic (x-axis) features across all clusters.

Focusing on the cultural dimension, we plot the cultural characteristics of the individual clusters in Figure 2a. In this analysis, we detect that cluster 2 (Northern European countries, U.S. and Canada) features the highest values for the gdp-feature along with high values for generosity, health, and freedom. Generally, we observe these characteristics for most western countries. In contrast, cluster 7 (Mexico) features especially low values for social support and high corruption compared to both cluster 1 (Argentina, Brazil, and Colombia) and cluster 2 (Northern European countries, U.S., and Canada). Besides these cultural differences, we moreover observe differences in the acoustic features plotted in Figure 2b: compared to cluster 2, cluster 1 features lower instrumentalness and speechiness values. Cluster 7 is relatively similar to cluster 2 with respect to the acoustic characteristics, however, has lower speechiness values. Along with that we find that cluster 2 features the highest speechiness and instrumental values among all clusters. Solely cluster 4 features similarly high values for speechiness and instrumentalness. This cluster contains user from France, Italy, Japan, and Spain. The slightly lower speechiness and instrumentalness values are also accompanied by lower generosity values. Furthermore, we see a strong correlation between French and Japanese users based on the cultural features ($\rho = 0.91$). While most of the feature values differ (but correlate), we found similar values for freedom and corruption. We argue, that in this special case, the variables of the WHR represent cultural aspects where these two countries appear similar, however, we assume that for Japan there exist other cultural aspects where those countries differ, but are not captured by the WHR.

Another interesting finding is the observation that high speechiness values correlate strongly ($\rho = 0.81$) with high freedom values in the dataset. We presume that clusters 2 and 4 containing western countries are characterized by a culture where freedom is important and along with that, music with high speechiness values and hence, mostly rap music is popular and important. We detect a similar pattern for the medium correlation between healthiness and instrumentalness ($\rho = 0.45$): cluster 2 and 4 group countries, where instrumental music is more popular than in other countries (and clusters). At the same time, cluster 4 (grouping European and Japanese users) as well as cluster 2 (grouping Northern Europe users, U.S., Canada, and Australia) are characterized by the highest health values. In fact, cluster 2 is characterized by the highest health values among all clusters.

4) Country-Specific Patterns: Besides looking at patterns that span across countries as presented in the previous section, we also aim to detect country-specific listening patterns. Therefore, we have a deeper look at the clusters solely grouping users of a single country, which can be considered outliers, in the remainder of this section. Particularly, we are interested in the features that make these countries stand out and hence, form individual clusters.

Cluster 3, a cluster solely grouping user from Chile, exhibits similar patterns to cluster 1 (grouping user from Argentina, Brazil, and Colombia), besides 28% higher instrumentalness values. Italian users are located in clusters 4 (4.92% of all Italian users) and 5 (95.08%, respectively). In contrast to cluster 4, where other European user are located, Italian users in cluster 5 are characterized by lower tempo (-2.23%), valence (-4.35%) and danceability (-2.41%) values. Cluster 6, containing users from Malaysia, is characterized by the lowest acousticness values among all clusters, which are 13.17% lower than the mean computed. This is accompanied by 10.42% lower instrumentalness values. Finally, Mexican users form their own cluster and hence have their own music listening pattern. The music listening behavior of Mexican users is similar to the western cluster 2, but we observe 13.05% lower speechiness values along with 5.32% lower acousticness values.

C. Discussion

We consider our analysis still as early, as we are aware of the fact that our findings based on Spotify users do not necessarily represent the world's music taste aside of music streaming. Further, we note that the number of users analyzed in this study is still limited and hence, naturally affects the generalizability of our study. Nevertheless, the results of our clustering approach, bringing together the cultural embedding of a user with his or her musical preferences, suggests that there exist several culture-specific music listening patterns. We categorize those listening patterns into two groups: we observe country-specific listening patters as well as cross-country listening patterns that span across several countries. The latter are not restricted to neighboring countries or continents, as we see in the bias towards instrumental and rap music for western countries reaching from Australia over Europe to the United States and Canada or commonalities in the music listening between Europe and Japan. Besides those cross-country listening patterns, we find country-specific patterns, for instance, the bias of Italian and Portuguese users towards music with high danceability values.

Based on these findings, we are particularly interested in analyzing cultural music listening patterns beneath the country level. An analysis beneath the country-level would mitigate the weakness of performing a majority vote to assign a country to a cluster as well as allow a more fine-grained analysis of cultural listening patterns, probably revealing regional listening patterns. Whereas analyzing regional listening patterns is possible due to precise GPS coordinates, the cultural analysis in contrast is challenging as we are not aware of a consistent data source observing cultural aspects beneath the country-level. Our analyses show, in contrast, that if cultural features are leveraged on the country-level, the incorporation of precise GPS coordinates is not beneficial. Besides a sub-country analysis, it would be also interesting to look into optimizing the user model: in the current approach, we use the arithmetic mean of each individual acoustic feature for aggregating values of each track, which implies the simplification that each user follows a single, homogeneous listening pattern. Hence, our analysis may benefit from a more comprehensive user model revealing multiple listening patterns per user.

Our findings show that Japan and France feature a high correlation based on cultural features. While might not seem obvious, according to the socio-economic features contained in the WHR data, the correlation holds. However, this also signals that characterizing a user's cultural embedding by WHR data only does not fully capture the cultural background of users. Therefore, we also aim to extend the description of cultural embeddings with further characteristics to improve precision in this regards in future work.

Nevertheless, based on our findings, we argue that a music recommender or retrieval system incorporating a user's cultural background allows for providing more fine-grained and personalized results. We particularly consider the finding that there are on the one hand clusters that span across multiple countries, and on the other hand, clusters that only comprise users of a single country as highly relevant. Furthermore, given that cultural information explains 41% of the variance in our dataset, therefore nearly as much as music content information (59%), we argue that cultural information is an important contextual variable that allows for better characterizing users. We believe that these findings can contribute to providing more personalized and culture-aware music recommendations by integrating country-specific listening patterns and cultural information.

VI. Conclusion

In this paper, we presented an analysis of culture-specific music listening behavior of Spotify users. We model a user's listening behavior by aggregating the acoustic features of the tracks the user has listened to and complement this information with socio-economic and cultural information gathered from the World Happiness Report. We find that the variance in our dataset can be explained to 41% by cultural and to 59% by musical features, which we believe is particularly interesting due to the two levels on which a user may be characterized: on a cultural level and on a musical level. These findings stress the importance of incorporating cultural information as a contextual variable into a music recommender system as it allows for describing users and their characteristics more accurately. Related to this, we could also show that the cultural distance (or similarity) of users cannot be replaced or even substantially improved by adding their exact GPS location.

To find cultural music listening patterns, we rely on the DBSCAN clustering algorithm. We find a set of nine clusters where each cluster is a group of countries which share common music listening patterns *and* common culture-specific characteristics. Also, we observe that some of these cluster span across several countries and hence, group users located in several different countries, who share common listening behavior. On the other hand, we also find country-specific clusters that contain users of a single country only.

Future work will include incorporating further data sources as well as extending the set of features, particularly regarding the cultural dimension. Furthermore, we aim to perform a more fine-grained analysis on subcultures within countries.

References

[1] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin, "Context-aware recommender systems," *AI magazine*, vol. 32, no. 3, pp. 67–80, 2011.

[2] M. Schedl, A. Flexer, and J. Urbano, "The neglected user in music information retrieval research," *Journal of Intelligent Information Systems*, vol. 41, no. 3, p. 523, 2013.

[3] M. Kaminskas, F. Ricci, and M. Schedl, "Location-aware music recommendation using auto-tagging and hybrid matching," in *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 2013, pp. 17–24.

[4] M. Braunhofer, M. Kaminskas, and F. Ricci, "Recommending music for places of interest in a mobile travel guide," in *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 2011, pp. 253–256.

[5] X. Wang, D. Rosenblum, and Y. Wang, "Context-aware mobile music recommendation for daily activities," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 99–108.

[6] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K.-H. Lüke, and R. Schwaiger, "Incarmusic: Context-aware music recommendations in a car," *E-Commerce and web technologies*, pp. 89–100, 2011.

[7] M. Schedl and D. Schnitzer, "Location-aware music artist recommendation." in *MMM (2)*, 2014, pp. 205–213.

[8] M. Schedl, A. Vall, and K. Farrahi, "User geospatial context for music recommendation in microblogs," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 2014, pp. 987–990.

[9] M. Schedl and D. Schnitzer, "Hybrid retrieval approaches to geospatial music recommendation," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013, pp. 793–796.

[10] D. Hauger and M. Schedl, "Exploring geospatial music listening patterns in microblog data." in *Adaptive Multimedia Retrieval*, 2012, pp. 133–146.

[11] J. L. Moore, T. Joachims, and D. Turnbull, "Taste space versus the world: an embedding analysis of listening habits and geography." in *ISMIR*, 2014, pp. 439–444.

[12] D. Hauger, M. Schedl, A. Košir, and M. Tkalcic, "The million musical tweets dataset: What can we learn from microblogs," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 2013.

[13] M. Pichl, E. Zangerle, and G. Specht, "Towards a context-aware music recommendation approach: What is hidden in the playlist name?" in *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1360–1365.

[14] ——, "Improving context-aware music recommender systems: Beyond the pre-filtering approach," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '17. New York, NY, USA: ACM, 2017, pp. 201–208. [Online]. Available: http://doi.acm.org/10.1145/3078971.3078980

[15] M. Schedl, "The lfm-1b dataset for music retrieval and recommendation," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*. ACM, 2016, pp. 103–110.

[16] C.-M. Chen, M.-F. Tsai, J.-Y. Liu, and Y.-H. Yang, "Using emotional context from article for contextual music recommendation," in *Proceedings of the 21st ACM International Conference on Multimedia*, ser. MM '13. New York, NY, USA: ACM, 2013, pp. 649–652. [Online]. Available: http://doi.acm.org/10.1145/2502081.2502170

[17] A. Schindler and A. Rauber, *Capturing the Temporal Domain in Echonest Features for Improved Classification Effectiveness*. Springer International Publishing, 2014.

[18] A. Goel, M. Sheezan, S. Masood, and A. Saleem, "Genre classification of songs using neural network," in *International Conference on Computer and Communication Technology*. IEEE, 2014, pp. 285–289.

[19] M. Pichl, E. Zangerle, and G. Specht, "Understanding Playlist Creation on Music Streaming Platforms," in *Proceedings of the IEEE Symposium on Multimedia (ISM)*. IEEE, 2016.

[20] J. F. Helliwell, H. Huang, and S. Wang, "The distribution of world happiness," *WORLD HAPPINESS*, p. 8, 2016.

[21] E. Diener, "Subjective well-being: The science of happiness and a proposal for a national index." *American psychologist*, vol. 55, no. 1, p. 34, 2000.

[22] U. Schimmack, P. Radhakrishnan, S. Oishi, V. Dzokoto, and S. Ahadi, "Culture, personality, and subjective well-being: integrating process models of life satisfaction." *Journal of personality and social psychology*, vol. 82, no. 4, p. 582, 2002.

[23] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[24] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[25] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.