



# Ethnic music exploration guided by personalized recommendations: system design and evaluation



Tiago Fernandes Tavares<sup>1</sup>  · Leandro Collares<sup>2</sup>

Received: 29 July 2019 / Accepted: 22 February 2020 / Published online: 4 March 2020  
© Springer Nature Switzerland AG 2020

## Abstract

Ethnic music collections provide vantage points into the culture of different communities. However, tools that allow the exploration of those music libraries by the general public are lacking. We present and evaluate an ethnic music collection exploration system that comprises a recommendation algorithm and a web-radio interface. The proposed system provides recommendations that adapt to the user's preferences. A blind A/B user study conducted to evaluate the system suggests that personalized recommendations could help individuals from diverse backgrounds explore ethnic music collections. We believe both the system and its evaluation contribute to the growing work in the field of intelligent information systems.

**Keywords** Media data exploration · Ethnic music exploration · Content-based information retrieval · Recommended personalization

## 1 Introduction

Ethnic music embodies language, customs, and family and social interactions of a community. Unfortunately, access to these music libraries by the general public is limited due to the scarcity of appropriate tools [1].

Many systems for accessing, organizing and exploring large collections have been proposed to deal with the increasing number of online media documents. While such systems are available in commercial websites geared toward browsing popular music [2, 3], they are rarely used for ethnic music collection (EMC) exploration.

Exploring EMCs is a far cry from browsing popular music libraries for three main reasons. Firstly, tags related to ethnic music (artist, location, etc) are often unavailable or considered cryptic by most users [4]. Secondly, the number of typical users of EMCs is relatively small, which hinders the use of collaborative filtering techniques [3] or social network behavior [5] in recommendation systems. Lastly,

user intentions while browsing ethnic music are often not clear, as music can be explored according to several personalized criteria [6]. Consequently, a recommendation system for ethnic music exploration has to adapt to user preferences. Building such a system is similar to finding content-based clusters, e.g., aesthetic trends in videos [7] or photographs [8] since it consists in discovering meaningful links between media documents based on personalized criteria.

Several systems have been specially designed for exploring and studying ethnic music. In their interfaces, documents can be visualized according to pitch [9], rhythm [10], repeated melodic phrases [11] and other MIR-related [4] properties. Manual multi-dimensional connections between documents [1], automatic annotation [12] and annotation comparison [13] are features available in some of those interfaces. While the said systems are extremely useful for research purposes [14], they require expert knowledge to be properly used.

✉ Tiago Fernandes Tavares, [tavares@dca.fee.unicamp.br](mailto:tavares@dca.fee.unicamp.br) | <sup>1</sup>School of Electrical and Computer Engineering, University of Campinas, Campinas, SP, Brazil. <sup>2</sup>Department of Psychology, University of Victoria, Victoria, BC, Canada.



Consequently, these systems tend to be unsuitable for the general public.

In this article, we propose a system for EMC exploration that makes personalized recommendations. The system was implemented as a simple web-radio interface that caters to the general public and allows users to navigate through unknown music collections with minimal interaction. Neither collaborative filtering nor curated description vectors were employed in the system.

In the devised system, the media library is mapped to a content-meaningful vector space using an auditory-inspired feature vector [15]. The Euclidean distance is used as an objective measurement of music similarity in the aforementioned space. In this context, searching for music is equivalent to walking a path within the vector space. Users explore different paths in the space by “liking” or “skipping” tracks. As both the mapping and the search processes are amenable to distributed tree [16, 17] and GPU-based [18] parallel searches, the proposed system could be scaled up to accommodate increasing music collections and user bases.

The proposed system is inherently different from those that rely on collaborative filtering or other social network activities since it does not require any prior usage data. It does not use information about the listener’s past behavior and can be employed by listeners that have not been acquainted with the music collection. Therefore, the proposed system implicitly deals with a double cold-start problem in which the system has no prior information about users, and users have no previous information on the contents of the dataset.

Instead of predicting the user’s taste, the music recommendation system fosters a constructive dialogue between the listener and the dataset [19]. In this dialogue, choices made by users help the system learn about their preferences. Consequently, recommendations cannot be simply evaluated as correct or incorrect concerning the ground truth. Rather, they are an integral part of the user experience.

As evaluations conducted with user participation help determine the real-world applicability of Music Information Retrieval (MIR) systems, we assessed the proposed system via an anonymous blind A/B user study in which users interacted freely with the system. Results suggest that the devised system could be used by the general public to explore EMCS.

The main contributions of this work are:

1. An open-source, web-based recommendation system for personalized EMC exploration that does not require any previous knowledge of the collection and does not rely on usage data, and

2. A behavior-based user study that indicates that both the proposed algorithm and the interface are suitable for exploring EMCS listeners are not familiar with.

The remainder of the paper is organized as follows. Section 2 discusses the background of this work. Section 3 describes the recommendation system. Section 4 details the user study carried out to evaluate the system. Section 5 presents and discusses results yielded by the user study. Finally, Sect. 6 closes the paper with concluding remarks.

## 2 Background

Ethnic music is usually collected for conducting research or preserving traditions. EMCS are built and maintained by institutions from diverse countries, which explains the myriad of technologies employed for storing and providing access to documents. Moreover, ethnic music metadata might be unavailable or comprise terms that are far different from Western standards, e.g., local names of instruments, performers’ background, and the role of the songs in the community [4].

Although researchers, ethnomusicologists and musicologists are typical users of EMCS, these music libraries could be potentially explored by the general public for purposes of education or entertainment. Systems that allow the general public to explore EMCS are, however, few and far between.

Websites that enable individuals to search for ethnic music often rely heavily on text. In those websites, users have to type in labels or keywords that are employed as inputs in document retrieval [20]. Text-based media retrieval results, however, are only meaningful if both the user and the collection curator agree on a vocabulary of descriptive labels regarding genre, artists and track titles, which is usually not the case when EMCS are considered [6, 21]. Consequently, text-based search interfaces may not be suitable for users without previous knowledge of the media collection.

Many music recommendation systems have been proposed in the last two decades to help listeners explore music collections. Several of the solutions hinge on either user- or content-based filtering.

In collaborative filtering, the behavior of a user base is analyzed and recommendations made based on the premise that “users that like  $x$  will probably like  $y$  as well”. This information is represented as elements  $a_{n,j}$  of a matrix  $\mathbf{A}$ , where  $n$  refers to a group of documents and  $j$  refers to the documents in the set that belong to that group. The  $\mathbf{A}$  matrix is multiplied by the  $\mathbf{B}$  matrix, whose elements  $b_{i,n}$  represent how much each user  $i$  likes documents from group  $n$ . The  $\mathbf{A}$  and  $\mathbf{B}$  matrices are obtained by factorizing

the user interaction matrix  $\mathbf{U}$ , whose elements  $u_{ij}$  are 1 if user  $i$  likes document  $j$  and 0 otherwise. Hence, we can use the expression

$$\mathbf{U} \approx \mathbf{BA} \quad (1)$$

and obtain  $\mathbf{B}$  and  $\mathbf{A}$  by minimizing the approximation error regarding  $\mathbf{U}$ .

This technique requires data from a large number of users, making it impractical for niche-specific music collections. Furthermore, it tends to recommend mainstream documents [22, 23] and fails to suggest new documents of the collection, which is undesirable in the context of exploration.

Several variations on collaborative filtering have been proposed to mitigate the said issues [24–28]. These approaches rely on enriching user listening data with additional information, such as acoustic features. However, they still depend on the analysis of large amounts of data related to audio content and user behavior. Consequently, these variations are not suitable for exploring  $\text{EMCs}$ .

Another user-based approach consists in employing social tags to generate recommendations [29–31]. As more popular documents tend to receive more accurate tags, social tagging can be tainted with bias. Therefore, a purely content-based approach is more suitable for  $\text{EMC}$  exploration.

Graphical content-based recommendation systems employ acoustic similarity to organize media files. These systems usually map the musical content of each track to a high-dimensional acoustic feature vector. The music collection, i.e., the vector space can be visualized in a two- or three-dimensional space in which tracks that have similar musical content are placed closed to each other [32–35]. Alternatively, tracks can be displayed in a world map according to their musical origin [6]. While content-based systems allow users to navigate through music collections, they often rely on different types of interactions between the user and the interface [6, 35]. Even though these interactions are relatively simple, user knowledge of audio similarity or music origins is often necessary to ensure a better experience.

Fully automatic content-based recommendation systems have also been devised by Balkema and van der Heijden [36], and Bohra et al. [37]. In these systems, an offline, label-based grouping process precedes the recommendation. As there is often disagreement over vocabulary, labels might not be suitable for  $\text{EMC}$  exploration.

In an interactive content-based approach proposed by Ikeda et al. [38], building a playlist is equivalent to drawing a smooth path within the content vector space. Recommended tracks are chosen among candidates that would keep the continuation of the path between the two last tracks smooth. Consequently, playlists built by the system

are marked by continuous and coherent mood changes. However, the playlist smoothness relies on the diversity of tracks within the content space. This assumption does not hold for smaller collections, which are much sparser in the content space.

Content-based systems have also been used to organize videos [7] and photographs [8] according to their aesthetics.

In the proposed system, the content-based recommendation algorithm works by estimating paths in the content vector space. Unlike the system devised by Ikeda et al. [38], the estimated path does not always result in smooth changes. In our system, the path estimation algorithm makes conservative or extreme recommendations according to user feedback.

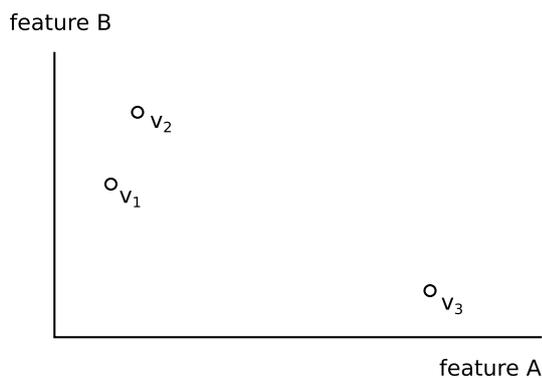
With regard to visual design, users explore  $\text{EMCs}$  via a familiar metaphor—a minimalist web-radio interface—in the proposed system. Moreover, no previous knowledge of the  $\text{EMC}$  is required. The system is thus designed for the general public.

Evaluation methods in music recommendation systems have been analyzed by Bonnin and Jannach [39]. They noted that most music recommendation system evaluations rely exclusively on specialist demonstrations [9, 10] and playlist prediction experiments [24, 26, 40, 41]. These evaluation methods do not consider the impact of the recommendation on the user experience and, consequently, can yield misleading results [42].

Questionnaire-based user surveys have been used to evaluate  $\text{MIR}$  systems [6, 30, 32, 43]. As this method depends on the answers provided by users, it can generate results plagued with bias. Conversely, analyzing the user behavior when interacting with the system can highlight unperceived or undeclared aspects of the user experience [35], provided that testing conditions are similar to real-world situations.

All recommendation systems make assumptions about the music listening process. For example, some assume that users would like to listen to the same tracks as their peers, whereas others presume users favor playlists with coherent tags. When implemented, these assumptions contribute to the construction of user preferences [19]. Such preferences change between users and throughout time. Therefore, it is difficult to state whether a recommendation system is better than the other [44] or to identify reasons underlying the perception of recommendation quality [45]. Evaluations can, however, highlight the feasibility of using a particular system within a specific scenario.

In this work, we conducted a user study using a blind, anonymous  $\text{A/B}$  protocol. While the proposed system made personalized recommendations, the control system recommended random tracks. Results yielded show that the proposed system recommendations outperformed



**Fig. 1** Each track  $k$  in the collection is represented as a point  $v_k$  in the content-inspired feature space. Tracks that are close to each other are similar, whereas tracks that are distant from each other are dissimilar. In the feature space above, the acoustic similarity is higher between  $v_1$  and  $v_2$  than between  $v_1$  and  $v_3$

random recommendations. Longer listening times were also observed when the devised system was used. Therefore, we believe that a web-radio powered by content-based recommendations can be suitable for exploring EMCS.

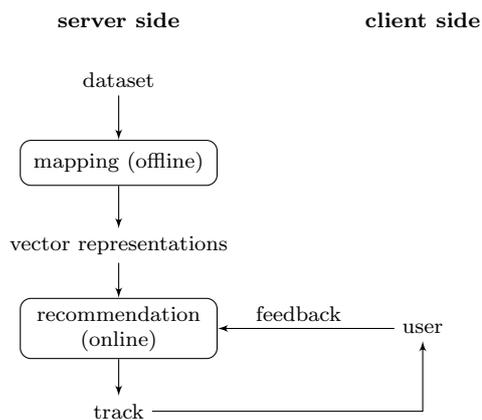
### 3 Recommendation system

The proposed system maps each audio track to a point in a high-dimensional vector space. Within this vector space, Euclidean distances are used to represent perceptual differences, i.e., similar tracks are mapped to points that are close to each other, whereas dissimilar ones correspond to points that are distant from each other. Figure 1 depicts a two-dimensional feature space.

The system hinges on the premise that listeners that like a track would probably enjoy music that is acoustically similar to the said track. If the user gives positive feedback (“like”) on a track, the next recommendation will be acoustically similar to the aforementioned track. Conversely, negative feedback (“skip”) informs the system that the subsequent recommendation has to be acoustically dissimilar to the current one.

The mapping process precedes the system’s operation and is performed offline as shown in Fig. 2. When the system is online, it gathers user feedback to build a path through the vector space. The path tends to cross regions that are close to tracks that received positive feedback and far from tracks that received negative feedback.

Figure 2 also shows that a client–server design guided the system implementation. The client comprises the interface that enables the interaction between users and the music collection. The server makes recommendations and receives user feedback. The interface was developed



**Fig. 2** System outline. A mapping process calculates vector representations for all elements (tracks) in the dataset in an offline stage. Later, an interactive, online recommendation algorithm suggests tracks to the user and collects feedback

in HTML, CSS and JavaScript,<sup>1</sup> whereas the REST API server was implemented in Python (WebPy framework).

Section 3.1 describes the process of mapping the music collection tracks onto the vector space, and Sect. 3.2 presents the recommendation algorithm.

#### 3.1 Mapping

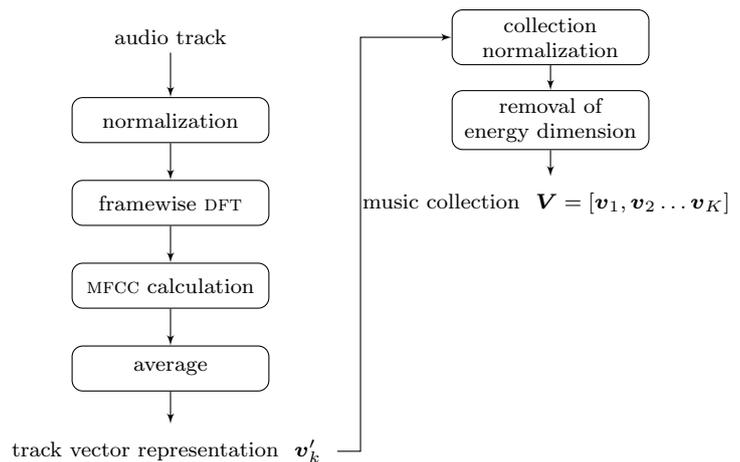
The mapping process produces a content-based organization of the music collection. Each track is represented by a high-dimensional feature vector comprising audio features. In the vector space, the Euclidean distance is employed as a proxy for auditory similarity.

The vector space has to address the trade-off between dimensionality and timbre description capability. A high-dimensional vector space is likely to yield richer timbre descriptions, whereas a low-dimensional one allows for quicker interactions as it relies on fewer data. The proposed system is based on the work by Tzanetakis and Cook [15], who observed that the average Mel-Frequency Cepstral Coefficients ( $M_{FCCS}$ ) calculated over a sliding window have great discriminatory power concerning timbre.

The audio features were computed in several steps. First, each of the  $K$  audio tracks in the collection were individually normalized so that its samples present zero mean and unit variance. Then, each track was split into 46-ms frames with a step size of 12.5 ms and multiplied by a Hanning window. After that, each frame had its Discrete Fourier Transform (DFT) calculated, and 20  $M_{FCCS}$  were computed from the DFT. Finally, the framewise average for

<sup>1</sup> Based on the Skeuomorphic Audio Player, by Josh Bader—<https://codepen.io/joshbader/pen/GqXbVvk>.

**Fig. 3** Digital signal processing steps for mapping musical collections onto a vector space



each MFCC was calculated, yielding a 20-dimensional vector  $\mathbf{v}'_k$  to represent each track  $k, k \in [1 \dots K]$ .

Track vectors were normalized across the music collection so that each dimension was restricted to the 0–1 range. The first dimension of each vector was removed as it represents energy, which does not play an important role in recommendation due to the normalization in the mapping process. The remaining 19 dimensions were used to represent the track in the vector space. This process outputs the collection of  $K$  track representation vectors  $\mathbf{v}_k$  and is depicted in Fig. 3.

The recommendation algorithm suggests a path across the content-based vector space and is detailed in the next subsection.

### 3.2 Recommendation algorithm

The recommendation algorithm hinges on the user’s interactions with the interface shown in Fig. 4. Each time the interface is loaded; a new session is started on the server. The interface streams a randomly chosen track from the server. Listeners can pause or play the track using the corresponding button and navigate through the track with the slider. When listeners click/tap on “like” or “skip” buttons, interactions are labeled by the server as positive or negative, respectively.

An interaction is labeled as positive if a track is listened to until the end, or the user clicks/taps on the “like” button while the track is playing. An interaction is labeled as negative if the user clicks/taps on the “skip” button before a track is finished. Labels of the most recent interactions are used in track recommendation. A new track is suggested after the track is played till the end or when the user clicks/taps on the “skip” or “like” buttons.

The recommendation algorithm records the set  $p$  of tracks that have been already played in the current session, and the user interaction labels  $l_1$  and  $l_2$ , which correspond

to the last and second last tracks played, respectively. Vectors to which these tracks have been mapped to are named  $\mathbf{t}_1$  and  $\mathbf{t}_2$ .

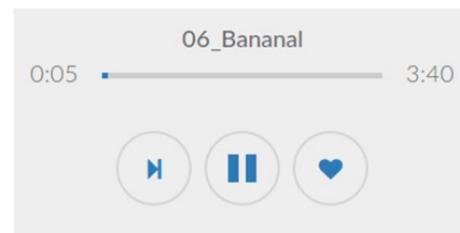
For each recommendation, the system calculates a reference point  $\mathbf{r}$  that is used to search for a new recommendation. The recommendation consists of the  $\hat{k}$ -th track in the collection, which is the closest one to  $\mathbf{r}$  that has not yet been played in the current session, that is:

$$\hat{k} = \arg \min_k \|\mathbf{r} - \mathbf{v}_k\|^2, \quad \mathbf{v}_k \notin p. \tag{2}$$

If both  $l_1$  and  $l_2$  are labeled as positive, then the system assumes that the current recommendations are suitable and yields a conservative estimation. In this estimation,  $\mathbf{r}$  is the mean between the previous track vectors, that is:

$$\mathbf{r} = \frac{\mathbf{t}_1 + \mathbf{t}_2}{2}. \tag{3}$$

Conversely, if both  $l_1$  and  $l_2$  are labeled as negative, recommendations were deemed unsuitable by the user. Thus the algorithm performs a random search in the content vector space and suggests a random vector  $\mathbf{r}$ . A random vector is also suggested during the cold start as there is no information about the user preferences, and when only



**Fig. 4** Web-radio interface elements. Track title, elapsed time and track duration are displayed. Users can use the slider, which also works as a progress bar, to navigate through the track. From left to right: “skip”, play/pause and “like” buttons

**Table 1** Recommendation algorithm use cases

Tracks		
Last	Second last	Next
$I_1$	$I_2$	$r$
N	N	Random
P	N	$t_1$
N	P	$t_2$
P	P	$\frac{t_1+t_2}{2}$

$n$  refers to negative or non-existent interactions.  $p$  is a positive interaction

one track has been suggested and considered inadequate by the user.

If  $I_1$  is labeled as positive and  $I_2$  is labeled as negative, then the only available reference for a search region is the last interaction, thus  $r = t_1$ . This behavior also addresses the situation in which only one track has been suggested and received positive feedback. Finally, if  $I_1$  and  $I_2$  are labeled as negative and positive, respectively, then the system backtracks to the second last recommendation by using  $r = t_2$ .

These four cases are summarized in Table 1.

The algorithm allows users to explore different regions of the vector space, depending on the way they interact with the system. Tracks that receive positive feedback causes the algorithm to suggest acoustically similar tracks, whereas negative feedback makes the system recommend tracks located in other areas of the music space. The source code for the recommendation system is available on GitHub.<sup>2</sup>

The proposed algorithm relies only on minimum Euclidean distances, which can be computed efficiently using distributed tree structures [16, 17] or GPU-based implementations [18]. The proposed system is thus horizontally scalable and suitable for distributed applications.

The system was evaluated using a blind A/B user study, as described in the next section.

## 4 Evaluation

The evaluation consisted of an anonymous A/B user study. It was conducted to assess whether the proposed recommendation algorithm could be effectively employed by the general public to explore ethnic music datasets.

**Table 2** Number of tracks, number of positive interactions and interaction time for each recommendation algorithm

Algorithm	Metric	Mean/SD	Median
Proposed	# Tracks	19.1 ± 16	12
	# Positive interactions	5 ± 4.9	4
	Time (s)	8992 ± 23039.3	555.7
Random	# Tracks	9.1 ± 7	7
	# Positives interactions	3.5 ± 3.3	2
	Time (s)	752.3 ± 1185.3	325.4

The ethnic music dataset used in the evaluation was the *Música das Cachoeiras* collection.<sup>3</sup> It was recorded with a mobile studio during a Rio Negro expedition in the Amazon forest that took place in 2013. The dataset comprises 85 documents among aboriginal and ritual music, and regional, Western-inspired songs. As the same studio was used during recording and audio mastering, there is no “producer effect” within the collection.

The study was advertised online using social media to encourage the participation of individuals interested in music. The system was kept online for 30 days.

During the experiment, individuals interacted with the system freely. Each access to the system was labeled as a different session. Within each session the subject was randomly assigned to either the proposed recommendation algorithm described in Sect. 3 or a baseline algorithm that simply makes random recommendations. All interactions were recorded for further analysis, as shown in the next section.

## 5 Results and discussion

In this section, user study results are presented and analyzed. Some interactions between user study participants and the ethnic music dataset are highlighted as well.

Sessions in which only the initial recommendation was made were not included in the result analysis as there was no interaction with the system in this scenario. Sessions in which no positive feedback was given were also excluded from the analysis since the algorithm only recommended random tracks. As a result, only 20 of 60 sessions (11 using the proposed algorithm and 9 using the random algorithm) were further analyzed. The sample size is compatible with previous work in the field [39, 43].

For each session, we calculated the number of tracks listened to, the number of positive interactions, i.e., the

<sup>2</sup> <https://github.com/tiagoft/macunaima>.

<sup>3</sup> <http://www.musicadascachoeiras.com.br/> and <https://soundcloud.com/musicadascachoeiras>.

**Table 3** Pearson and Spearman correlations for number of tracks, number of positive interactions and interaction time

Algorithm	Metrics	Pearson	Spearman
Proposed	# Tracks · # positives	<b>0.76</b>	<b>0.76</b>
	# Tracks · time	<b>0.86</b>	<b>0.88</b>
	# Positives · time	<b>0.64</b>	<b>0.93</b>
Random	# tracks · # positives	<b>0.87</b>	<b>0.67</b>
	# Tracks · time	-0.07	<b>0.56</b>
	# Positives · time	0.02	<b>0.80</b>

Numbers in bold are significant to a 5% level ( $p < 0.05$ )

number of tracks that received a “like” or were listened till the end, and the total interaction time (from the start of the session until the last recommendation). Mean, standard deviation and median values for these metrics are shown in Table 2.

Table 2 shows that user behavior is marked by great variance. Nevertheless, both mean and median consistently suggest a better user experience when individuals were assigned to the proposed algorithm. Moreover, values show that users interacted with the interface longer when recommendations were made by the proposed algorithm, which means the proposed algorithm recommendations outperformed random ones.

To assess the impact of the proposed algorithm on individual sessions, we calculated session-wise Pearson and Spearman correlations between the three metrics—number of tracks listened to, number of positive interactions and total interaction time. Results are shown in Table 3.

Table 3 results highlight that Spearman correlation values are consistently higher for the proposed algorithm metrics. Individuals that clicked/tapped on the “like” button more often tended to listen to more tracks for a longer time when the proposed algorithm made recommendations. Pearson correlation values regarding usage time were also higher when the proposed algorithm was employed.

As mentioned before, user study sessions in which only one recommendation was made or no positive feedback was given were excluded from the analysis. Consequently, only  $\frac{1}{3}$  of sessions were analyzed. This fact suggests that most participants entirely rejected the music collection employed in the user study. The said behavior was observed in previous research as well [6, 32, 43]. We speculate that rejection can be due to the incompatibility between the dataset and participants’ musical preferences.

The high variance in results presented in Table 2 shows that users behaved in different ways. The median, however, was always lower than the mean, that is, the distribution is skewed to the left. Therefore, most subjects were *light* users, whereas some subjects were *heavy* users. This

finding corroborates the common 80–20 hypothesis, i.e., few users have most of the impact on system usage.

All user interactions are the result of a dialogue between users and the system. The user’s perspective on the collection has a considerable bearing on those interactions. It is possible that some users are more prone to listening to different tracks while others are not. Some users might have interacted with the collection briefly due to lack of time or dislike of the music styles. Nevertheless, results suggest that content-based recommendations can provide better user experiences than random ones.

The user study indicated that acoustic similarity can help listeners during the exploration process and, as a consequence, increase the interaction time with the system. Different search criteria in the feature space can evoke diverse user experiences. A story told by a curator or tracks listened by one’s social network could guide or influence music collection exploration as well. All these vantage points to music space exploration are useful depending on the context.

The next section presents conclusive remarks.

## 6 Conclusion

We presented a music recommendation system for exploring EMCS with a simple web-radio interface and conducted a user study to evaluate it. The analysis of usage time and number of tracks listened to indicates that the proposed system provides a better user experience than a system that merely suggests random tracks.

The proposed system comprises a simple interface with “skip” and “like” buttons, and a MFCC-based mapping of audio tracks to a vector space. User interactions and the vector space are employed by the recommendation algorithm to suggest tracks similar to those that received positive feedback by the listener and different from those that received negative feedback.

As ethnic music user bases are often small and EMCS may be entirely unknown to most people, music exploration using collaborative filtering and labels commonly associated with ethnic music can be challenging. We believe our system, which quickly builds a user-specific model for music clustering, addresses these issues by offering individuals the possibility of navigating unknown music collections with minimal interaction.

Evaluating system performance when the music space contains thousands of tracks and is accessed by more users is a possible research avenue. Conducting user studies in such conditions could assess the impact of computing Euclidean distances using distributed tree structures or GPU-based implementations on the user experience.

The proposed system draws information solely from acoustic features. It is useful in some situations, and suitable for a restricted number of users. Curator- and user-based recommendation systems lend themselves to specific scenarios. Consequently, it is difficult to rank recommendation systems using objective quality measures. Investigating scenarios in which each recommendation paradigm could be employed constitutes an important research theme for future work.

Other research routes include the assessment of the system in museums, expositions, music schools, and multimedia installations. Even though the system was designed for ethnic music exploration, evaluating its use when interacting with experimental, indie, regional or personal music collections could be relevant to both MIR and intelligent information systems fields.

**Acknowledgements** We thank Agenor Vasconcelos for kindly authorizing the use of the music collection in this research.

**Funding** This work was funded by FAPESP, São Paulo Research Foundation, Grant 2015/18671-4.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

## References

- Dattolo A (2010) Structuring and navigating ethnic music archives. *Sig Process* 90(4):1077–1082
- Glaser WT, Westergren TB, Stearns JP, Kraft JM (2006) Consumer item matching method and system. US Patent US7003515B1
- Davidson J, Liebold B, Liu J, Nandy P, Van Vleet T, Gargi U, Gupta S, He Y, Lambert M, Livingston B, Sampath D (2010) The YouTube video recommendation system. In: Proceedings of the fourth ACM conference on recommender systems, RecSys '10, pp 293–296, New York, NY, USA, ACM
- Cornelis O, Lesaffre M, Moelants D, Leman M (2010) Access to ethnic music: advances and perspectives in content-based music information retrieval. *Sig Process* 90(4):1008–1031 (special section: ethnic music audio documents: from the preservation to the fruition)
- Fields B, Jacobson K, Rhodes C, d'Inverno M, Sandler M, Casey M (2011) Analysis and exploitation of musician social networks for recommendation and discovery. *IEEE Trans Multimed* 13(4):674–686
- Magas M, Proutskova P (2013) A location-tracking interface for ethnomusical collections. *J New Music Res* 42(2):151–160
- Aran O, Biel JI, Gatica-Perez D (2014) Broadcasting oneself: visual discovery of vlogging styles. *IEEE Trans Multimed* 16(1):201–215
- Hong R, Zhang L, Zhang C, Zimmermann R (2016) Flickr circles: aesthetic tendency discovery by multi-view regularized topic modeling. *IEEE Trans Multimed* 18(8):1555–1567
- Six J, Cornelis O (2011) Tarsos: a platform to explore pitch scales in non-western and western music. In: 12th International society for music information retrieval conference (ISMIR-2011), pp 169–174. University of Miami, Frost School of Music
- Antonopoulos I, Cornelis O, Pikrakis A, Leman M, Moelants D (2007) Music retrieval by rhythmic similarity applied on Greek and African traditional music. In: Proceedings of ISMIR 2007
- Kroher N, Pikrakis A, Diaz-Banez JM (2017) Discovery of repeated melodic phrases in folk singing recordings. *IEEE Trans Multimed* 20:1291–1304
- Cornelis O, Six J, Holzapfel A, Leman M (2013) Evaluation and recommendation of pulse and tempo annotation in ethnic music. *J New Music Res* 42(2):131–149
- Ness SR, Biró DP, Tzanetakis G (2010) Computer-assisted cantillation and chant research using content-aware web visualization tools. *Multimed Tools Appl* 48(1):207–224
- Serra X (2012) Opportunities for a cultural specific approach in the computational description of music. In: 2nd CompMusic Workshop, pp 1–4, Istanbul, Turkey, 12/07/2012
- Tzanetakis G, Cook P (2002) Musical genre classification of audio signals. *IEEE Trans Speech Audio Process* 10(5):293–302
- Patwary MMA, Satish NR, Sundaram N, Liu J, Sadowski P, Racah E, Byna S, Tull C, Bhimji W, Prabhat, Dubey P (2016) Panda: extreme scale parallel k-nearest neighbor on distributed architectures. In: 2016 IEEE international parallel and distributed processing symposium (IPDPS), pp 494–503, May
- Cahsai A, Ntarmos N, Anagnostopoulos C, Triantafyllou P (2017) Scaling k-nearest neighbours queries (the right way). In: 2017 IEEE 37th international conference on distributed computing systems (ICDCS), pp 1419–1430, June
- Douze M, Johnson J, Jegou H (2017) Faiss: a library for efficient similarity search. <https://code.facebook.com/posts/1373769912645926/faiss-a-library-for-efficient-similarity-search/>
- Hail S (1980) Encoding/decoding. In: Culture, media, language. Working Papers in Cultural Studies.
- Lidy T, Silla CN, Cornelis O, Gouyon F, Rauber A, Kaestner CAA, Koerich AL (2010) On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections. *Sig Process* 90(4):1032–1048 (special section: ethnic music audio documents: from the preservation to the fruition)
- Gómez E, Herrera P, Gómez-Martin F (2013) Computational ethnomusicology: perspectives and challenges. *J New Music Res* 42(2):111–112
- Chou S-Y, Yang YH, Lin Y-C (2015) Evaluating music recommendation in a real-world setting: on data splitting and evaluation metrics. In: 2015 IEEE international conference on multimedia and expo (ICME), pp 1–6, June
- Terveen L, Hill W (2001) Beyond recommender systems: helping people help each other. In: Carroll J (ed) HCI in the new millennium. Addison-Wesley, Boston, pp 1–21
- McFee B, Barrington L, Lanckriet G (2012) Learning content similarity for music recommendation. *IEEE Trans Audio Speech Lang Process* 20(8):2207–2218
- Lee K, Lee K (2014) Using dynamically promoted experts for music recommendation. *IEEE Trans Multimed* 16(5):1201–1210
- Lu Y, Qiao Z, Zhang P, Guo L (2016) Ranking-based music recommendation in online music radios. In: 2016 IEEE first international conference on data science in cyberspace (DSC), pp 614–619, June
- Vassiliou C, Stamoulis D, Martakos D (2006) A recommender system framework combining neural networks. In: Collaborative filtering, WSEAS transactions on information science and applications, vol 3(7), pp 1202–1207
- Borges R, Queiroz M (2018) Automatic music recommendation based on acoustic content and implicit listening feedback. *Rev Música Hodie* 18(1):31–43
- Nanopoulos A, Rafailidis D, Symeonidis P, Manolopoulos Y (2010) Musicbox: personalized music recommendation based

- on cubic analysis of social tags. *IEEE Trans Audio Speech Lang Process* 18(2):407–412
30. Shao B, Wang D, Li T, Ogihara M (2009) Music recommendation based on acoustic features and user access patterns. *IEEE Trans Audio Speech Lang Process* 17(8):1602–1611
  31. Eck D, Lamere P, Bertin-mahieux T, Green S (2008) Automatic generation of social tags for music recommendation. In: Platt JC, Koller D, Singer Y, Roweis ST (eds) *Advances in neural information processing systems*, vol 20. Curran Associates Inc, Red Hook, pp 385–392
  32. Knees P, Schedl M, Pohle T, Widmer G (2007) Exploring music collections in virtual landscapes. *IEEE Multimed* 14(3):46–54
  33. Lillie AS (2008) *MusicBox: navigating the space of your music*. Master's thesis, Massachusetts Institute of Technology
  34. Stober S, Nürnberger A (2010) *MusicGalaxy—an adaptive user-interface for exploratory music retrieval*. In: *Proceedings of 7th sound and music computing conference*
  35. Collares L, Tavares TF, Gooch A, Tzanetakis G (2017) Personalizing self-organizing music spaces with anchors: design and evaluation. *Multimed Tools Appl* 77:5525–5545
  36. Balkema W, van der Heijden F (2010) Music playlist generation by assimilating GMMs into SOMs. *Pattern Recogn Lett* 31(11):1396–1402
  37. Bohra TS, Kumar V, Ganesan S (2015) Segmenting music library for generation of playlist using machine learning. In: *2015 IEEE international conference on electro/information technology (EIT)*, pp 421–425
  38. Ikeda S, Oku K, Kawagoe K (2017) Analysis of music transition in acoustic feature space for music recommendation. In: *Proceedings of the 9th international conference on machine learning and computing, ICMLC 2017*, pp 77–80, New York, NY, USA, ACM
  39. Bonnin G, Jannach D (2014) Automated generation of music playlists: survey and experiments. *ACM Comput Surv* 47(2):26:1–26:35
  40. Kamehkhosh I, Jannach D (2017) User perception of next-track music recommendations. In: *Proceedings of the 25th conference on user modeling, adaptation and personalization, UMAP '17*, pp 113–121, New York, NY, USA, ACM
  41. Chen C-W, Lamere P, Schedl M, Zamani H (2018) Recsys challenge 2018: automatic music playlist continuation. In: *Proceedings of the 12th ACM conference on recommender systems, RecSys '18*, pp 527–528, New York, NY, USA, ACM
  42. Jannach D, Kamehkhosh I, Bonnin G (2016) Biases in automated music playlist generation: a comparison of next-track recommending techniques. In: *Proceedings of the 2016 conference on user modeling adaptation and personalization, UMAP '16*, pp 281–285, New York, NY, USA, ACM
  43. Wang X, Wang Y, Hsu D, Wang Y (2014) Exploration in interactive personalized music recommendation: a reinforcement learning approach. *ACM Trans Multimed Comput Commun Appl* 11(1):7:1–7:22
  44. Kovačević A, Milosavljević B, Konjović Z, Vidaković M (2010) Adaptive content-based music retrieval system. *Multimed Tools Appl* 47(3):525–544
  45. López-Sintas J, Cebollada À, Filimon N, Gharhaman A (2014) Music access patterns: a social interpretation. *Poetics* 46:56–74

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