Personal Factors in Music Preference and Similarity: User Study on the Role of Personality Traits

Pei-I Chen, Jen-Yu Liu, and Yi-Hsuan Yang

Academia Sinica
Research Center for Information Technology Innovation
gooa1121@citi.sinica.edu.tw
ciaua@citi.sinica.edu.tw
yang@citi.sinica.edu.tw

Abstract. To discuss how different people measure music similarity in multiple aspects differently in their perception of similarity and preference, we report a user study to explore the effect of personality traits in music preference and the perception of similarity. We build audio-based music auto-taggers to describe music content in the aspects: genre, acoustics, and emotion, and then construct a similarity-based music search system that retrieves clips similar to a query one in one of these three aspects. 21 participants are invited to do a few trials of music retrieval and rate preference and similarity perception. We then study the correlation among the personality traits, the preference and the similarity ratings. The effect of retrieval methods accounting for different music aspects is also discussed. With the subjective nature of similarity perception, the result of this work may shed some light on the creation of a user-centered system for similarity-based music search.

Keywords: Music information retrieval, musical similarity, listening behavior, personality trait

1 Introduction

Music information retrieval (MIR) can be done in multiple ways, which can be broadly categorized into text-based and audio-based ones [7]. In text-based music retrieval systems, the user query can be a few text keywords specifying the song title, artist name, or some key characteristics of the music piece such as genre, instrument, emotion, etc. In audio-based music retrieval systems, the user provides an audio example and requests the system to find similar ones. Such examples include audio fingerprinting, query-by-humming, query-by-tapping, cover song detection, similarity search, etc, differing in the level of specificity, namely the choice of exact or approximate matching [7]. A great amount of research work has been done in the literature to develop advanced text-based or audio-based retrieval systems. Performance bottleneck usually lies in the difficulty of designing audio features that discriminatively and accurately represent the content of the music pieces, and designing unsupervised or supervised algorithms to measure the similarity between music pieces or to map the audio features to high-level semantics such as genre and emotion [25, 42]. Some researchers are more interested in musically meaningful features and the related fundamental music signal processing problems such as multipitch estimation, onset detection, tempo estimation, and chord recognition, whereas some researchers are more interested in developing machine learning algorithms to construct systems that can be useful in various applications. Since its advent in 2005, the annual Music Information Retrieval Evaluation eXchange (MIREX; `http://www.music-ir.org/mirex/wiki/MIREX_HOME`) [9] has provided a community-based formal evaluation framework for all types of music signal processing or information retrieval tasks.

Despite the great research effort that has been made, the majority of research papers published in the annual conference of the International Society for Music Information Retrieval and related venues is concerned with *the music itself*, trying to develop algorithms that for example better *reproduce* the genre labels [34] or similarity ratings [43] provided by users. Although such effort improves the MIR community as a whole toward better understanding music signals and the ways to model them, it provides little knowledge about how users interact with music in daily lives. For example, although a great deal of work has been done to automatically recognize the emotion a piece of music expresses or evokes, relatively little work has been done to investigate how users selects music of different music emotions in different listening moods or activities [44, 45]. As advocated and practiced in a number of existing work [3, 4, 12, 15, 18, 19, 29–32], it is important to understand not only the music but also the *users*, and to involve users in the development and evaluation processes of music information retrieval. The growing interest in the MIR community in understanding user behaviors can be seen from the organization of the MIREX 2014 Grand Challenge on User Experience, which for the first time adopts "holistic, user-centered evaluation of the user experience in interacting with complete, user-facing music information retrieval systems" in MIREX (`http://www.music-ir.org/mirex/wiki/2014:GC14UX`).

Following this research line, we present in this paper a preliminary study that aims at understanding user behaviors for the specific task of *similarity-based music search*, which is an instance of audio-based music retrieval system that responds to a user with music clips similar to the query one. This study is motivated by the observation that music similarity can be measured in various aspects including timbre, rhythm, melody, genre, etc, and that different people may perceive music similarity in fairly different ways. In consequence, it may not be possible for a general similarity-based music search system to work equally well for every user. Intuitively, the performance of music search can be improved by *personalizing* the system for each specific user, but this approach may demand large number of user feedbacks to be effective. We are interested in knowing whether it is possible to find correlations between the *personal traits* of users (e.g. age, gender, personality, music training) and their perception of similarity in music, so that we can devise different retrieval systems for people of different personal traits, before we perform personalization. Moreover, we want to know whether this correlation is effected when we rely on different music aspects to perform similarity search, and how the perception of similarity is related to the preference of music. In other words, the study aims at understanding the correlation among the following four factors: personal traits, music preference, perception of music similarity, and the music aspect considered for performing similarity search.

Specifically, we conduct a user study and invite 21 participants to engage in a *complete, user-facing* similarity-based music search system developed by our lab. The users are asked to perform a few trials of music retrieval by freely selecting query music pieces by their own (instead of a fixed set of query pieces), and then evaluate how they like the retrieved pieces (i.e. music preference) and how they think the retrieved pieces are similar to the query ones (i.e. perception of music similarity). Because we are using a music search system that is developed by our own, we can manipulate the system parameters and use different music aspects when we compare the similarity between songs, making it possible to study the effect of different music aspects. Moreover, the user study situates the participants in a real music searching context, which improves the validity of the user response we can collect. In the end of the user study, we ask the participants to fill a questionnaire [8] to assess their personality in the following dimensions: 'Extraversion,' 'Agreeableness,' 'Conscientiousness,' 'Neuroticism,' and 'Openness.' We consider the above "Big Five" *personality traits* [14] as the main personal factors in this study, for it has been found that these traits are related to music preference [28]. From the 1,008 responses from the participants and the self-reported personality traits, we are able to perform a series of correlation analysis to gain some insights into the behavior of users. To our best knowledge, this study represents the first of its kind to address such research questions using a functional similarity-based music search system.

The paper is organized as follows. Section 2 presents the implemented music retrieval system. Section 3 describes the performed user study. Section 4 reports the correlation analysis and the results. Section 5 discusses the relation between this work and existing ones. Finally, Section 6 concludes the paper.

2 Implemented Retrieval System

Music similarity estimation can be considered as obtaining a suitable distance measurement between music pieces defined on a certain feature space [6]. The feature space can be constructed by either low-level features such as Mel-frequency cepstral coefficients or chroma features [25], or by using high-level semantic descriptors based on the inference of different musical dimensions by discriminative classifiers [41]. These music dimensions, or *aspects*, include genre, culture, emotion, instrument, rhythm, and tempo annotations, amongst others [6]. In our implementation, we develop audio-based music auto-taggers to obtain semantic descriptors of music in the following three aspects: `genre`, `acoustics`, and `emotion`, and then use the semantic descriptors to measure the similarity between music pieces.

The three aspects are selected for the following reasons. First, according to a recent user study [15], genre and emotion (also referred to as mood in the MIR literature) are the two most important factors in playlist creation and management of music collections, besides editorial metadata such as artist names, album names, or song titles. Second, thanks to research on text-based music retrieval, researchers have built datasets that can be used to train auto-taggers for the three aspects. Specifically, the CAL10k dataset constructed by Tingle *et al.* [36] contains genre and acoustics labels, whereas the MER31k dataset compiled by Yang *et al.* [44, 45] provides music emotion labels. Effective ways of training the auto-taggers for these tags have also been proposed (e.g. [39]). Third, the acoustics labels in the CAL10k dataset contain "acoustically objective" tags that describe for example the characteristics of instruments, vocal timbre, tempo, and key in a music piece. Therefore, the three aspects provide a rich yet diverse description of the music content. Although it is possible to further subdivide the acoustic tags into smaller groups, we stay with the three aspects for simplicity.

2.1 Music Auto-tagging

Music auto-tagging, or automatic music annotation, refers to the task of automatically assigning semantic labels (tags) such as genre and emotion to music objects (e.g. artists, tracks, or segments of a track) to facilitate applications such as tag-based music retrieval, similarity search, recommendation, and visualization [20,38]. A number of approaches have been proposed to collect labeled data to train music auto-taggers [37]. The CAL10k dataset contains the annotation of 153 genre tags and 475 acoustic tags for 10,267 Western songs [36]. The tags are labeled by professional music editors employed by the Internet radio company Pandora (`http://www.pandora.com/`). Example genre tags include *rock*, *jazz*, *bebop*, *classical*, and *flamenco*; and acoustic tags include *a ballad tempo*, *a dry recording sound*, *a gravelly male vocalist*, *a lively piano solo*, and *a manic bass line*. The MER31k dataset, on the other hand, consists of the annotation of 190 emotion tags of 31,422 Western songs. The annotations are entered by the crowd of last.fm users (`http://www.last.fm/`) as a result of social tagging [17]. Example emotion tags include *mellow*, *amiable*, *smooth*, *fun* and *ethereal*.

We obtain the 30- or 60-seconds MP3 audio previews of these songs via the 7digital API (`http://www.7digital.com`). For MER31k, we have gathered all the 31,422 songs. For CAL10k, we are able to collect 7,799 previews. We only keep the tags with more than 50 previews to supply sufficient data to train the auto-taggers. 140 tags remain in `Genre` and 435 tags remain in `Acoustics`.

We employ the AWtoolbox to compute the audio words [13, 35, 46] as the audio feature in this paper. First the log magnitude spectrogram is extracted. By using a basis of vectors, called a *dictionary*, the log magnitude spectrogram is decomposed with respect to the basis with 1-norm sparse constraint, using sparse coding techniques [23]. We used a dictionary with 1,024 vectors learned from USPOP [10]. Audio words capture local and time-varying characteristics

Aspect	Training data	Number of tags	Average AUC
`Genre`	CAL10k [36]	140	0.811
`Acoustics`	CAL10k [36]	435	0.786
`Emotion`	MER31k [44,45]	190	0.725

Table 1: The three music aspects, number of tags, and the accuracy of music auto-tagging in cross validation evaluations

in the audio signal and its effectiveness has been demonstrated in several tasks, including the classification of genre, acoustics, emotion, and instruments [13,35, 46]. The representation for a track is a vector formed by mean-pooling the audio words across frames.

We build the auto-tagging models with Fixed-Point Continuation Approximation (FPCA) [22]. FPCA is a state-of-the-art efficient optimization method. It aims at building a model that maps the input vector to the target vector while minimizing the rank of the mapping matrix. We apply FPCA on the auto-tagging problem by taking the audio words as input and taking the label vector as the target. The n-th term in the label vector is 1 only if the song is labeled with the n-th tag, and other terms are 0s. The FPCA model is suitable for large datasets as the prediction is produced by simply multiplying the input vector by the mapping matrix.

Table 1 lists the three aspects and the area under the receiver operating characteristic curve (AUC) of the implemented auto-taggers in five-fold cross validation evaluations. AUC is a performance measure for tag-based music retrieval using the predicted tags. Its value falls in [0, 1] and a larger value indicates better performance. Although there is still room for improvement, empirically we have observed that the tags predicted by our auto-taggers are fairly reasonable.

2.2 Similarity-based Music Retrieval

We collect another collection of 84,164 audio previews from 7digital to be used as the *retrieval database* in the music retrieval system. As described in [21], the database is supposed to cover a wide range of Western music. We employ the auto-taggers trained from CAL10k and MER31k to predict the genre, emotion and acoustics tags for the songs in the retrieval database. Then, the similarity between two songs is estimated by comparing the predicted tags of the two songs. For example, if the `acoustics` aspect is considered, we will represent the content of a song by a 435-dimensional feature vector indicating the likelihood (a.k.a. *label confidence*) of the song being associated with each of the tag, and then use the l_1 distance (i.e. the sum of element-by-element absolute difference) between the feature vectors of two songs to measure dissimilarity. Given a query song, the songs with the minimal dissimilarity with it are retrieved. Although it is possible to measure similarity/dissimilarity by mixing the three aspects or by using other advanced distance metrics (e.g. [24,33]), we opt for using the aspects independently and the simple l_1 distance in this study for simplicity.

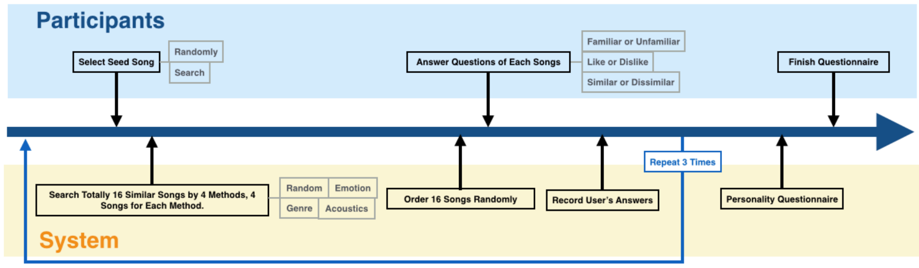

Fig. 1: Flowchart of the user study.

3 User Study

We recruit 21 undergraduate or graduate students (12 females, 9 males, age range: 21–30) from universities to participate in the user study. All the participants are born and raised in Taiwan.

As shown in Figure 1, the user study consists of the following steps.

- From the retrieval database, a participant selects a favored seed song by searching artist name, song title, or using a random song picked by the system. The participant listens to the songs before deciding the one to be used as the seed song.
- The system would then choose 4 songs by each of the following 4 methods from the retrieval database: similar in `genre`, similar in `acoustics`, similar in `emotion`, or random selection. The 16 retrieved songs are shuffled and presented to the participant with song titles and artist names displayed.
- Without knowing how the song list is generated, the participant evaluates whether she or he likes the retrieved song or not, whether the retrieved song is similar to the seed song (whatever the participant defines similarity), and whether she or he is familiar with the retrieved song, all in binary forced-choice fashion.
- These above steps are repeated three times. That is, the participant needs to choose three seed songs and evaluate 48 retrieved songs in total. The participants are asked to choose seed songs with diverse styles.
- In the end, the participant is asked to fill the Chinese version of the 60-item NEO-Five Factor Inventory [8] to assess their personality in terms of the Big Five traits [14]. The values for each personality trait range from 1–5.

The experiment interface and music retrieval system are implemented as a website using PHP and Python codes. The participants are asked to wear a headset and stay calm in a silent room while performing the study over the Internet. It takes each of them 30–50 minutes to complete the experiment. The participants are monetary rewarded and are assured that the collected data would be kept in private. In total, $21 \times 48 = 1,008$ responses are collected.

Figure 2 shows the histograms of the participants' personality traits, with five histogram bins. We see that the participants are biased toward high Agree-

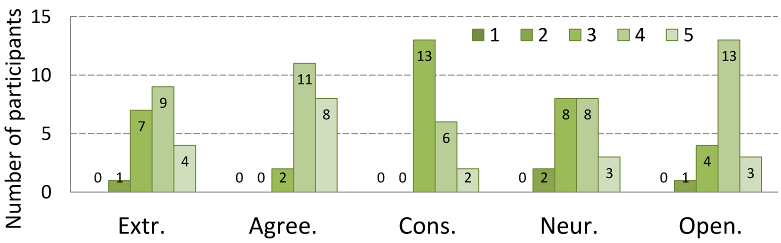

Fig. 2: Distribution of participants' personality traits values in the Big Five dimensions.

ableness and high Conscientiousness, and that there are better bell shapes for Extraversion and Neuroticism. Admittedly, the size and the diversity of our participants are fairly limited, so the findings of this study may not generalize well. However, hopefully the methodology presented in this paper can still be helpful in analyzing user behavior in future larger-scale user studies.

4 Analysis and Result

Our analysis is divided into three parts. First, we analyze the relationship between music aspects, music preference and similarity perception. Second, we investigate the role of personality in the personal perception of similarity. Finally, we combine the factors of personality and similarity in different music aspects and discuss the implications of the result on similarity search. We do not make use of the familiarity data in this analysis.

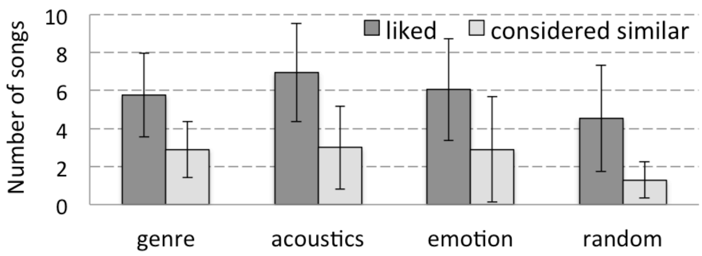

Fig. 3: Average number of liked songs (ranging from 0 to 12) and similar songs (ranging from 0 to 8) for retrieval methods using different music aspects.

4.1 Music Preference and Similarity Perception in Different Music Aspects

Figure 3 shows the average number of songs liked (heavy gray) and considered similar to the query (light gray) across all the participants, for each of the four retrieval methods. As each retrieval method selects four songs for each of the three query songs, the values here should range from 0 to 12. However, due to an unfortunate bug, we do not record the similarity ratings for the last query song, so the number of similar songs ranges from 0 to only 8.

From Figure 3 we see that `acoustics` has the highest number of average liked songs (mean=6.95, SD=2.67), followed by `emotion` (mean=6.05) and then `genre` (mean=5.76). Unsurprisingly, the participants less favor the songs chosen randomly (mean=4.52). In one-way ANOVA analysis, the retrieval method has significant effect on the number of liked songs (F=3.4, p=0.0283).

From Figure 3 we also see that on average 3 out of the 8 retrieved songs by each of the three content-based methods are considered similar by the participants. The result of `genre` appears to have smaller standard deviation (SD=1.48).

In view of the poor result of the random method (mean=1.29), we see that finding similar songs is a fairly challenging task. The three music aspects do not differ much in the average number of retrieved similar songs. Although it is possible to improve the performance of similarity search by fusing the three aspects, this is not the focus of this study.

Liked	`Genre`	`Acoustics`	`Emotion`	Random
Extraversion	0.365	0.009	0.267	0.335
Agreeableness	**0.487**	0.275	**0.603**	0.287
Conscientiousness	0.253	0.240	**0.476**	0.069
Neuroticism	–0.317	0.082	–0.372	–0.112
Openness	**0.561**	0.238	**0.466**	0.315
Similar	`Genre`	`Acoustics`	`Emotion`	Random
Extraversion	0.022	0.085	0.033	0.300
Agreeableness	–0.044	0.117	0.243	0.084
Conscientiousness	0.159	0.175	**0.513**	0.133
Neuroticism	0.180	0.197	0.136	0.020
Openness	–0.042	–0.195	–0.008	0.126

Table 2: The correlation between personality traits and the number of songs retrieved by different methods that are liked (upper) or deemed similar (lower). Bold font indicates significant correlation (p <0.05).

4.2 The Relationship between Personality and the Number of Liked or Similar Songs

We then investigate the effect of personality by correlating each of the personality trait with the number of liked or similar songs selected by different retrieval methods. Specifically, we compute the Pearson correlation coefficient (ranging from −1 to 1) between for example the Openness values of the 21 participants and the number of songs each of the 21 participants likes among the 12 songs retrieved by using the `Genre` aspect. The result is shown in Table 2.

Table 2 shows that songs retrieved by `Genre` and `Emotion` can be easily preferred by people of certain personality traits, yet such songs are not favored by people who rate high in Neuroticism. The correlation between the number of liked songs and Neuroticism is negative, except for the `Acoustics` method. On the other hand, it seems that the performance of different music aspects in retrieving similar songs has little to do with personality traits. Comparing to the number of songs liked, the correlation coefficients between personality traits and the number of songs considered similar appear to be smaller (especially for Extraversion and Openness) for all the music aspects, and we observe significant correlation between only `Emotion` and Conscientiousness.

To gain more insights, in Table 3 we further show the correlation coefficients between the ratings in the five personality traits and the following four quantities:

- Total number of liked songs (from 0 to 48) and similar songs (from 0 to 36) per participant.
- For each participant, we divide the number of liked/similar songs retrieved by each retrieval method by the total number of liked/similar songs, and then take the standard deviation of the liked/similar rate across the four retrieval methods. This indicates whether the participant has strong preference for a specific music retrieval method in finding favorite or similar songs.

The following observations can be made from Table 3. First, people who rate high in Agreeableness or Openness tend to like more songs, while people who rate high in Neuroticism like less. Second, there is no significant correlation between any personality trait and the total number of songs considered similar. Third, people who rate high in Extraversion or Openness do not have specific preferred music aspect in liking songs, but again no significant correlation is found for the perception of similarity.

Remarks — Our study so far shows that the three music aspects retrieve similar number of songs that are considered to be similar to the query ones by the participants, regardless of their personalities. People of certain personalities would be more selective in the songs they would like, but it is not clear whether these liked songs are those considered similar by the users. This leads to last part of our analysis.

	Personal total liked songs	Liked rate SD	Personal total similar songs	Similar rate SD
Extraversion	0.299	**–0.551**	0.103	–0.021
Agreeableness	**0.506**	–0.336	0.142	0.273
Conscientiousness	0.319	–0.219	0.322	–0.039
Neuroticism	–0.214	0.315	0.185	–0.260
Openness	**0.481**	**–0.446**	–0.082	0.324

Table 3: Correlation between personality and statistics of personal music preference and similarity perception. Bold font indicates significant correlation (p <0.05).

4.3 Personality and Influence of Perceived Similarity on Music Preference

In this part, we investigate whether people would like the songs they consider similar to the query songs. Specifically, Table 4 shows the correlation coefficients between the ratings in the five personality traits and the following five quantities:

- Correlation coefficient between the preference ratings and the similarity ratings (both binary) of 36 retrieved songs per participant; a scalar for each participant.
- Total number of songs considered 'similar & liked' simultaneously (from 0 to 36) for each participant.
- Similarly, the number of songs for 'similar & disliked,' 'dissimilar & liked,' and 'dissimilar & disliked.'

While there is no significant correlation between any personality trait and the correlation between preference ratings and the similarity ratings, we see significant correlations in the last four columns of Table 4. In particular, we note the following interesting findings:

- People who rate high in Conscientiousness tend to like the songs they find similar.
- Neuroticism is the only personality trait with moderately positive correlation (though not significant) for the case 'similar & disliked,' implying that for these people similarity does not imply preference.
- People who rate high in Agreeableness can more easily like the songs they find dissimilar. Actually, Agreeableness is the only personality trait with positive correlation (though not significant) for this case.
- Similarly, people who rate high in Agreeableness tend NOT to dislike the songs they find dissimilar.

The above result may imply that, for example, it is fine to diversify the retrieval result a little bit for people who rate high in Neuroticism or Agreeableness. In contrast, for people who rate high in Conscientiousness, it is a good idea to suggest songs that are indeed similar to the query one.

	Overall correlation	Similar & liked	Similar & disliked	Dissimilar & liked	Dissimilar & disliked
Extraversion	–0.047	0.148	–0.113	–0.121	–0.279
Agreeableness	0.155	0.240	–0.258	0.309	**–0.440**
Conscientiousness	0.199	**0.402**	–0.180	–0.085	–0.273
Neuroticism	–0.186	0.068	0.365	–0.390	0.042
Openness	0.114	0.028	–0.328	**–0.520**	–0.350

Table 4: Correlation between personality and the relation between all liked and similar songs, the number of songs classified to 4 dimensions of similarity and preference: 'similar & like,' 'similar & dislike,' 'dissimilar & like,' and 'dissimilar & dislike.' Bold font indicates significant correlation (p <0.05).

5 Related Work

The relationship between personality and preferred music genres or styles has been studied before [27,28]. However, the methodology adopted in existing work is mostly based on self-reported questionnaires or field studies, rather than user responses while interacting with a functional music information retrieval computer system.

A great deal of research work has been devoted to audio-based and text-based music similarity estimation [1, 2, 5, 6, 11, 16]. Recent years have also witnessed a growing interest in improving the performance of music retrieval by either giving users more control in customizing the system parameters, or by personalizing the system using user feedbacks [26, 40, 47]. Instead of exploring novel similarity measures, this work presents the methodology and a corresponding correlation analysis to investigate the role of personality in music similarity search, an issue that has not been well studied in the literature. Moreover, we investigate whether the retrieved similar songs would be in favor by real users in the context of similarity-based music retrieval.

6 Conclusions

In this paper, we have presented an attempt to evaluate the performance of an audio-based music similarity search system by real users. The goal of the study is to understand the role of personality traits in such a music retrieval problem. We have analyzed the correlation between the self-reported personality traits of 21 participants and the number of songs considered similar by the participants among the 48 retrieved songs in response to three query songs. Result shows that there is no obvious correlation between personality traits and the preferred music aspects in similarity search, but people with different personalities do behave fairly differently when it comes to whether they are in favor of the retrieved songs that they consider similar to the query one. This finding may be used in setting up the default setting for a personalized similarity-based music retrieval system, and related tasks such as content-based music recommendation and playlist generation.

7 Acknowledgement

This work was supported by a grant from the Ministry of Science and Technology under the contract MOST 102-2221-E-001-004-MY3 and the Academia Sinica Career Development Program.

References

1. Allan, H., Müllensiefen, D., Wiggins, G.A.: Methodological considerations in studies of musical similarity. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 473–478 (2007)

2. Aucouturier, J.J., F., P.: Improving timbre similarity: How high's the sky? J. Negative Results in Speech and Audio Sciences 1(1) (2004)
3. Barrington, L., Oda, R., Lanckriet, G.R.G.: Smarter than Genius? Human evaluation of music recommender systems. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 357–362 (2009)
4. Baur, D., Büttgen, J., Butz, A.: Listening factors: A large-scale principal components analysis of long-term music listening histories. In: Proc. ACM Int. Conf. Human Factors in Computing Systems. pp. 1273–1276 (2012)
5. Berenzweig, A., Logan, B., Ellis, D.P.W., Whitman, B.: A large-scale evaluation of acoustic and subjective music-similarity measures. Computer Music Journal 28(2), 63–76 (2004)
6. Bogdanov, D., Serra, J., Wack, N., Herrera, P., Serra, X.: Unifying low-level and high-level music similarity measures. IEEE Trans. Multimedia 13(4), 687–701 (2011)
7. Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M.: Content-based music information retrieval: Current directions and future challenges. Proceedings of the IEEE 96(4), 668–696 (2008)
8. Costa, P.: Revised Neo Personality Inventory (Neo Pi-R) and Neo Five-factor Inventory (Neo-Ffi): Professional Manual. Psychological Assessment Resources (1994)
9. Downie, J.S., Hu, X., Lee, J.H., Choi, K., Cunningham, S.J., Hao, Y.: Ten years of MIREX (music information retrieval evaluation exchange): Reflections, challenges and opportunities. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 657–662 (2014)
10. Ellis, D., Berenzweig, A., Whitman, B.: The "uspop2002" Pop Music data set (2003), [Online] `http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html`
11. Flexer, A.: On inter-rater agreement in audio music similarity. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 245–250 (2014)
12. Hu, X., Kando, N.: User-centered measures vs. system effectiveness in finding similar songs. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 331–336 (2012)
13. Jao, P.K., Yeh, C.C.M., Yang, Y.H.: Modified lasso screening for audio word-based music classification using large-scale dictionary. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. pp. 5207–5211. IEEE (2014)
14. John, O.P., Srivastava, S.: The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research 2(1999), 102–138 (1999)
15. Kamalzadeh, M., Baur, D., Möller, T.: A survey on music listening and management behaviours. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 373–378 (2012)
16. Knees, P., Schedl, M.: A survey of music similarity and recommendation from music context data. ACM Trans. Multimedia Computing, Communications, and Applications 10(1), 2:1–2:21 (2013)
17. Lamere, P.: Social tagging and music information retrieval. J. New Music Research 37(2), 101–114 (2008)
18. Law, E.: The problem of accuracy as an evaluation criterion. In: Proc. Int. Conf. Machine Learning (2008)
19. Lee, J.H., Waterman, N.M.: Understanding user requirements for music information services. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 253–258 (2012)

20. Levy, M., Sandler, M.: Music information retrieval using social tags and audio. IEEE Trans. Multimedia 11, 383–395 (2009)
21. Liu, J.Y., Liu, S.Y., Yang, Y.H.: Lj2m dataset: Toward better understanding of music listening behavior and user mood. In: Proc. IEEE Int. Multimedia Expo. pp. 1–6 (2014)
22. Ma, S., Goldfarb, D., Chen, L.: Fixed point and bregman iterative methods for matrix rank minimization. Mathematical Programming 128(1-2), 321–353 (2011)
23. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proc. Int. Conf. Machine Learning. pp. 689–696 (2009)
24. McFee, B., Barrington, L., Lanckriet, G.R.G.: Learning content similarity for music recommendation. IEEE Trans. Audio, Speech & Lang. Processing 20(8), 2207–2218 (2012)
25. Müller, M., Ellis, D.P.W., Klapuri, A., Richard, G.: Signal processing for music analysis. IEEE J. Sel. Topics Signal Processing 5(6), 1088–1110 (2011)
26. Nanopoulos, A., Rafailidis, D., Symeonidis, P., Manolopoulos, Y.: MusicBox: Personalized music recommendation based on cubic analysis of social tags. IEEE Trans. Audio, Speech & Lang. Processing 18(2), 407–412 (2010)
27. Pearson, J.L., Dollinger, S.J.: Music preference correlates of jungian types. Personality and Individual Differences 36(5), 1005–1008 (2004)
28. Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. J. personality & Social Psychology 84(6), 1236 (2003)
29. Schedl, M., Flexer, A.: Putting the user in the center of music information retrieval. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 385–390 (2012)
30. Schedl, M., Flexer, A., Urbano, J.: The neglected user in music information retrieval research. J. Intelligent Information Systems 41(3), 523–539 (2013)
31. Schedl, M., Hauger, D., Farrahi, K., Tkalčič, M.: On the influence of user characteristics on music recommendation. In: Proc. European Conf. Information Retrieval (2015)
32. Schedl, M., Stober, S., Gómez, E., Orio, N., Liem, C.C.: User-aware music retrieval. In: Müller, M., Goto, M., Schedl, M. (eds.) Multimodal Music Processing, Dagstuhl Follow-Ups, vol. 3, pp. 135–156. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2012)
33. Schlüter, J.: Learning binary codes for efficient large-scale music similarity search. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 581–586 (2013)
34. Sturm, B.L.: Classification accuracy is not enough : On the evaluation of music genre recognition systems. J. Intelligent Information Systems 41(3), 371–406 (2013)
35. Su, L., Yeh, C., Liu, J., Wang, J., Yang, Y.: A systematic evaluation of the bag-of-frames representation for music information retrieval (2014)
36. Tingle, D., Kim, Y.E., Turnbull, D.: Exploring automatic music annotation with acoustically-objective tags. In: Proc. ACM Int. Conf. Multimedia Information Retrieval. pp. 55–62 (2010)
37. Turnbull, D., Barrington, L., Lanckriet, G.R.G.: Five approaches to collecting tags for music. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 15–20 (2008)
38. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.R.G.: Semantic annotation and retrieval of music and sound effects. IEEE Trans. Audio, Speech & Lang. Processing 16(2), 467–476 (2008)
39. Vaizman, Y., McFee, B., Lanckriet, G.: Codebook based audio feature representation for music information retrieval. IEEE/ACM Trans. Audio, Speech & Lang. Processing 22(10), 1483–1493 (2014)

40. Vignoli, F., Pauws, S.: A music retrieval system based on user driven similarity and its evaluation. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 272–279 (2005)
41. Wang, J.C., Lee, H.S., Wang, H.M., Jeng, S.K.: Learning the similarity of audio music in bag-of-frames representation from tagged music data. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 85–90 (2011)
42. Wang, J.C., Yang, Y.H., Wang, H.M., Jeng, S.K.: Modeling the affective content of music with a gaussian mixture model. Affective Computing, IEEE Transactions on 6(1), 56–68 (2015)
43. Wolff, D., Stober, S., Nurnberger, A., Weyde, T.: A systematic comparison of music similarity adaptation approaches. In: Proc. Int. Soc. Music Information Retrieval Conf. pp. 103–108 (2012)
44. Yang, Y.H., Liu, J.Y.: Quantitative study of music listening behavior in a social and affective context. IEEE Trans. Multimedia 15(6), 1304–1315 (2013)
45. Yang, Y.H., Teng, Y.C.: Quantitative study of music listening behavior in a smartphone context. ACM Transactions on Interactive Intelligent Systems (TiiS)
46. Yeh, C.C.M., Jao, P.K., Yang, Y.H.: AWtoolbox: Characterizing audio information using audio words. In: Proc. ACM Int. Conf. Multimedia. pp. 809–812 (2014), [Online] `http://mac.citi.sinica.edu.tw/awtoolbox/`
47. Zhang, B., Xiang, Q., Wang, Y., Shen, J.: Compositemap: a novel music similarity measure for personalized multimodal music search. In: Proc. ACM Int. Conf. Multimedia. pp. 973–974 (2009)