Beyond the Big Five personality traits for music recommendation systems

Mariusz Kleć[1*], Alicja Wieczorkowska[1], Krzysztof Szklanny[1] and Włodzimierz Strus[2]

Abstract

The aim of this paper is to investigate the influence of personality traits, characterized by the BFI (Big Five Inventory) and its significant revision called BFI-2, on music recommendation error. The BFI-2 describes the lower-order facets of the Big Five personality traits. We performed experiments with 279 participants, using an application (called Music Master) we developed for music listening and ranking, and for collecting personality profiles of the users. Additionally, 29-dimensional vectors of audio features were extracted to describe the music files. The data obtained from our experiments were used to test several hypotheses about the influence of personality traits and the audio features on music recommendation error. The performed analyses take into account three types of ratings that refer to the cognitive-emotional, motivational, and social components of the attitude towards the song. The experiments showed that every combination of Big Five personality traits produces worse results than using lower-order personality facets. Additionally, we found a small subset of personality facets that yielded the lowest recommendation error. This finding can condense the personality questionnaire to only the most essential questions. The collected data set is publicly available and ready to be used by other researchers.

Keywords Music recommendation systems, Personality traits, The Big Five Inventory-2, Collaborative filtering

1 Introduction

The volume of music data uploaded to the Internet has increased radically. The expanding number of music collections, mobile access to audio files and streaming services pose challenges to finding appropriate songs. Today, thanks to the popularity of streaming services such as Spotify[1], Last.fm[2], Tidal[3], Pandora[4], or Qobuz[5], music discovery and recommendation systems have become much more popular than they were several years ago. Most of these services are hybrid systems (HS) that combine collaborative filtering (CF) and content-based (CB) approaches.

CF analyzes the community's ratings to conclude one's musical preference. The underlying assumption is that if a person *A* highly rates the same music as person *B*, then the system is more likely to recommend to user *A* songs unheard by *A* from the music pool of user *B* than that from any randomly chosen user [1, 2]. Although this approach is widely adopted and computationally fast, it has limitations. First, CF assumes that musical taste is fixed and does not change over time, which is not always true [3]. Another limitation is the tendency to recommend popular music over those pieces that have few ratings. Individual and unique preferences have no chance of being discovered by this algorithm. Therefore, the most critical obstacle is the Cold-Start (CS) problem [4]. It concerns the issue that the system has not yet gathered sufficient information about the user or item to infer precise recommendation. One of the strategies for tackling this problem is to resort to the

---

[1] http://www.spotify.com

[2] https://www.last.fm/

[3] http://www.tidal.com

[4] https://www.pandora.com/

[5] http://www.qobuz.com

*Correspondence:
Mariusz Kleć
mklec@pjwstk.edu.pl
[1] Multimedia Dept., Polish-Japanese Academy of Information Technology, Warsaw, Poland
[2] Institute of Psychology, Cardinal Stefan Wyszyński University, Warsaw, Poland

Kleć *et al. EURASIP Journal on Audio, Speech, and Music Processing* (2023) 2023:4

Page 2 of 17

user's contextual data (e.g., social network) in order to enrich rating profiles. The enhanced information about the user can be further used for clustering "similar" users and personalize the recommendation [5–7]. The user personality is a special case of such contextual data. The assumption is that people with similar personalities have similar interests and behavioral patterns [8], so they also rate the music in a similar way. Personality can be derived implicitly from social networks [9] or explicitly from users [10]. The latter is on asking the user to answer a list of personality questions. However, the personality questionnaires are well established in the psychology field but not for recommendation systems. Additionally, they may be very long (some of them contain even 240 items [11]). Therefore, in this paper we also want to address this problem and select only the most relevant personality traits for doing recommendations. This approach allows to reduce the number of personality questions and presumably increase the satisfaction from using the system.

The CB approach can also alleviate the CS problem. It focuses on the content of items, which can be the metadata or audio features. In this case, a singular song's rating from the user is enough to calculate the similarity of that song's features to the others to make the recommendation. However, it leads to the recommendations that are "too similar", without a chance to surprise the user (low serendipity). Hybridizing these two approaches (i.e., CF and CB) can give satisfactory results. The hybrid approach is used today by large companies like Spotify or Pandora. The significant contribution to this field comes from adopting Deep Learning (DL) [12–14], which allows automatic feature extraction from audio signals [15], or learning latent factors from user-item rating data [16, 17].

However, the factors that influence musical taste vary among individuals. Therefore, music information retrieval systems need to go beyond these approaches to deliver better recommendations. The type of music that one wants to listen to depends not only on listening history but also on one's current disposition, activity, as well as health condition, education, gender, and musical training [18–21].

1.1 Factors underpinning musical preferences

A positive correlation between a specific situation (context) and the preference for the music exists [18, 21]. It is possible to track the listener's context (e.g., time [22], weather [23], location [24]) and derive the musical taste in that context implicitly [25–27]. In the works [3, 28, 29], the authors utilized the surrounding environment (e.g., noise, time, light, and weather) to suggest music.

Other essential factors that influence musical preferences are emotions [20]. While listening to music, people want to relieve stress, change or match their current emotions with those expressed by the music. Descriptions exist on how to communicate emotions via musical structure and how our emotions are influenced by listening to music [30]. Tracking the listener's emotions can help to improve the quality of a recommendation [31]. It is usually achieved implicitly by tracking the context, such as keywords from an extensive collection of documents written by users [32], or extracting the users' texts from social networks [33, 34]. Another approach is to derive emotions from the user's face using the inbuilt camera of a mobile phone [31, 35] or from the signals obtained via wearable physiological sensors [36]. Consequently, research on Context-Aware Music Recommendation Systems (CA-MRS) has gained importance in recent years [37].

However, musical preferences depend not only on the way people regulate their emotions, and current situation, but also on their personality [38]. For example, people who are neurotic (i.e., have low emotional stability) are more likely to use music to foster emotions [20]. Conversely, people who are conscientious and low in creativity (low open-mindedness) are more likely to use music for emotional change and emotional regulation [39].

The systems that incorporate the user's personality into the recommendation process are called Personality-Aware Music Recommendation Systems (PA-MRS) and are a branch of the CA-MRS [40].

In 2003, Rentfrow and Gosling [41] empirically found the relationships between personalities and musical preferences. Namely, reflective, complex music (e.g., blues, jazz, or folk) and intense and rebellious music (e.g., rock, alternative or heavy metal) are positively related to Openness to experience. On the other hand, upbeat and conventional music (e.g., country or pop) negatively correlates with Openness, but it positively correlates with Extraversion, Agreeableness and Conscientiousness. Finally, energetic and rhythmic music (e.g., hip-hop, dance or electronic) is positively correlated with Extraversion and Agreeableness. Classical music positively correlates with Neuroticism [42]. In 2011, Rentfrow et al. in [43] provided an improved description of musical preferences. Their findings demonstrate a latent five-factor structure underlying music preferences (further called MUSIC factors): Mellow (comprising smooth and relaxing styles), Urban (defined largely by rhythmic and percussive music), Sophisticated (includes classical, operatic, world music, and jazz), Intense (defined by loud, forceful, and energetic music) and Campestral (comprising a variety of various styles of direct and rootsy music, often found in country and singer-songwriter genres).

In [44] Bansal and co-authors confirmed that the music genre relates to the Big Five personality traits. They analyzed a global music-download database consisting of millions of entries with music metadata describing

people downloading songs onto Nokia mobile phones. They showed that many genres in people's music collections are positively associated with Openness and (unexpectedly) Agreeableness, suggesting that individuals with high Openness and Agreeableness have broader musical tastes than those with high levels of other personality traits. The outcomes also aligned with literature showing that individuals who prefer jazz and folk score highly in Openness [45]. Such persons also tend to avoid genres like pop [46]. Since the level of Openness is related with the level of IQ [47], the findings above also find confirmation in the work of [48]. The authors indicate that people with higher IQ tend to prefer reflective and complex (e.g., jazz, classical, folk, blues) to upbeat and conventional music (e.g., pop). It is because the complex and reflective music is more likely to suit those who seek intellectually stimulating experiences. These people use music in rational or intellectual rather than emotional ways, implying higher levels of cognitive processing.

In [42] the authors indicate strong positive correlations between Neuroticism and classical music preference. Interestingly, they did not find Conscientiousness, Extraversion, or Neuroticism to be predictors of genre exclusivity. However, in [49] they analyzed a large dataset consisting of music listening histories and personality scores of 1415 Last.fm users. Their results show agreements with prior work but also the negative relation between Conscientiousness and folk music. They also report positive relations between Extraversion and such genres as R&B or rap, Agreeableness and country or folk, and also between Neuroticism and alternative music. However, musical genre is a conventional term and often the border between different musical genres is quite blurry. The authors in [50] investigated how different musical taxonomies (e.g., mood, activity, genre) influence the user experience and satisfaction of using music streaming services. Their findings are correlated with the Big Five personality traits. They also describe the link between the musical expertise of the listener and the number of categories within the given taxonomy. Their outcomes show that musically sophisticated users (e.g., experts) enjoy using the system more when exposed to a broader set of categories. This is also confirmed in [51] where experts enjoy the music more when having a more diverse choice of them. Still, there is a need for describing the link between personality and music in a more quantitative way. Such an approach is described in [52]. They correlated such audio features as dynamics, mode, register and tempo with the Big Five. The authors have shown (among others) that slow tempo is rated higher by high in Conscientiousness, major mode is preferred by low in Conscientiousness but high in Extraversion and piano dynamics are rated higher by high in Openness.

In general, audio features are expressed in a quantitative way and can be used together with personality traits in PA-MRS. Interesting approach is described in [53] where authors try to predict the personality trait (Extraversion or Introversion) on the basis of the audio features of the excerpt by employing several classification algorithms.

The authors of [54] showed that the recommendation accuracy could be improved by integrating personality traits. They also demonstrated that the accuracy depends on the recommendation domain: higher accuracy can be achieved in the movie domain than in the music domain.

In another paper [55], the authors analyze the influence of personality traits and emotional states (among others) on ratings. They found that the users with a high degree of Agreeableness rate at least 0.5 stars higher compared to the users with low Agreeableness (on a rating scale from 1 to 5) [56]. In [57] the authors compared the contribution of personality features and physiological signals (recorded by a wearable device) to the accuracy of their recommendation system. They found that the physiological features contributed less than the personality features.

It is also worth mentioning that users with different personalities show different preferences, regarding not only the recommendation accuracy, but also such properties of recommendation as diversity, popularity, and serendipity [58, 59]. The personalization of diversity is described in [60] and used in [61]. The authors demonstrated increased user satisfaction and recommendation diversity when they personalized the system according to the user's personality.

1.2 Personality acquisition

Developing the most efficient acquisition for music recommendation systems (MRS) is a challenge. The review of personality assessment questionnaires can be found in [40]. The most popular one is the Big Five Inventory (BFI) questionnaire, used for Big Five personality acquisition [62, 63]. The Ten Item Personality Inventory (TIPI) is another common option [64]. Generally, the questionnaires vary in the number of questions that the user is to answer. The TIPI is a very short questionnaire containing only 10 items. However, most questionnaires contain more than 50 items (some even 100, 200, and more). Longer questionnaires provide higher reliability, but, at the same time, require more effort from the user. Therefore, researchers try to acquire personality factors implicitly, e.g., using machine learning techniques with features extracted from social media streams [9]. The implicit acquisition does not require any action from the user, but its performance is much worse than explicit methods. For example, in [65], the authors were able to predict personality parameters from Twitter within 11–18% of

their actual value, by looking at the content of the user's tweets. Thus, the obtained results were very low, which was also confirmed in [66].

1.3 Contribution

We hypothesize that selecting only the most relevant personality traits for doing recommendations allows for reducing the recommendation error and limiting the number of questions the user needs to answer. To verify this hypothesis, we aimed at selecting the most relevant personality traits. In our study, we decided to use an explicit method for personality acquisition. Since using long questionnaires may be fatiguing for users, we wanted to find a trade-off between the reliability of the user personality representation and the length of the questionnaire. We used the revised version of the BFI (i.e., BFI-2) [67], as it contains 60 items (questions) and allows to go beyond the Big Five personality traits, by also measuring the lower-order level (i.e., facets) of the Big Five. We developed an application (called Music Master) for gathering users' personality information, listening to music, and rating it. Based on the data collected from the listening sessions, a memory-based hybrid music recommendation system has been developed and evaluated in an offline manner. The memory-based approach allows us to measure (among others) the similarities between users, in terms of various subsets of their personality traits, and clearly interpret the recommendation process. Based on the results, we selected only those traits (and their corresponding questions from the BFI-2 questionnaire) that contributed most to the system's performance. To the best of our knowledge, the BFI-2 has not been used before in any recommendation system. Additionally, we have published the collected data with ratings, features, and personality traits, to make them available for further investigations by other researchers.

2 Personality

Personality describes how individuals differ in their permanent emotional, interpersonal, experiential, attitudinal and motivational styles [68]. Over the past quarter-century, personality psychology has been dominated by theories of traits. There are several established and at the same time competing models of personality trait structure, such as the so-called Giant Three model by Eysenck [69], six-factor HEXACO model [70], or Two-Factor Model of higher-order personality factors [71, 72]. However, the Five-Factor Model, which is also known as the Big Five [11, 62, 73], is the prevailing conceptualization of personality structure and its basic dimensions. According to the Big Five model, most of the significant individual differences in people's patterns of thinking, feeling,

and behaving are embraced by five personality domains: Extraversion, Agreeableness, Conscientiousness, Neuroticism (or Negative Emotionality) and Openness to experience (alternatively labeled Intellect or Open-Mindedness) [11, 62, 67]. These domains are basic personality dimensions, and each of them is a quantitative variable with a positive and negative pole (e.g., the negative pole of Extraversion is introversion, and the negative pole of Neuroticism is emotional stability).

Most papers focus on the Big Five model [40], possibly because of the ease of its interpretation and because the results can be expressed quantitatively [74]. A discussion on the usability of this model in recommendation systems can be found in [75]. However, in our study, the revised version of the BFI (i.e., BFI-2) [67] was used. This psychometric model contains scales for the 5 primary domains and 15 subscales, nested within the primary ones (in total 20 personality dimensions, further referred to as traits). Brief characteristics of primary personality domains, and a list of their lower-order subscales (further referred to as facets) are given below:

- *Extraversion*: characterizes the activity (energy) level, the number of social interactions and social self-confidence, as well as positive emotionality.
  - Sociability, Assertiveness, Energy Level;
- *Agreeableness*: general disposition toward other people: positive, trustful, polite, empathic and altruistic vs. negative, antagonistic, and egocentric.
  - Compassion, Respectfulness, Trust;
- *Conscientiousness*: revealed in relation to work, rules and obligations and characterizes the level of orderliness, dutifulness, as well as perseverance and diligence.
  - Organization, Productiveness, Responsibility;
- *Neuroticism*: contains negative emotionality, over-sensitivity, volatility and irritability, as well as vulnerability, lack of resistance to stress, and low self-esteem.
  - Anxiety, Depression, Emotional Volatility;
- *Openness*: positive (cognitive) attitude towards novelty and both intellectual stimuli (abstract ideas), as well as aesthetic (or artistic) experiences; vivid imagination and complex thinking.
  - Aesthetic Sensitivity, Intellectual Curiosity, Creative. Imagination

3 Music Master application

We developed an application to gather the listener's personality profiles and musical ratings. The application communicates with the server using the TCP/IP protocol. The client part is called Music Master (MM). Its User Interface (UI) is divided into three main views: personality registering, personality visualization, and music player (see Fig. 1).
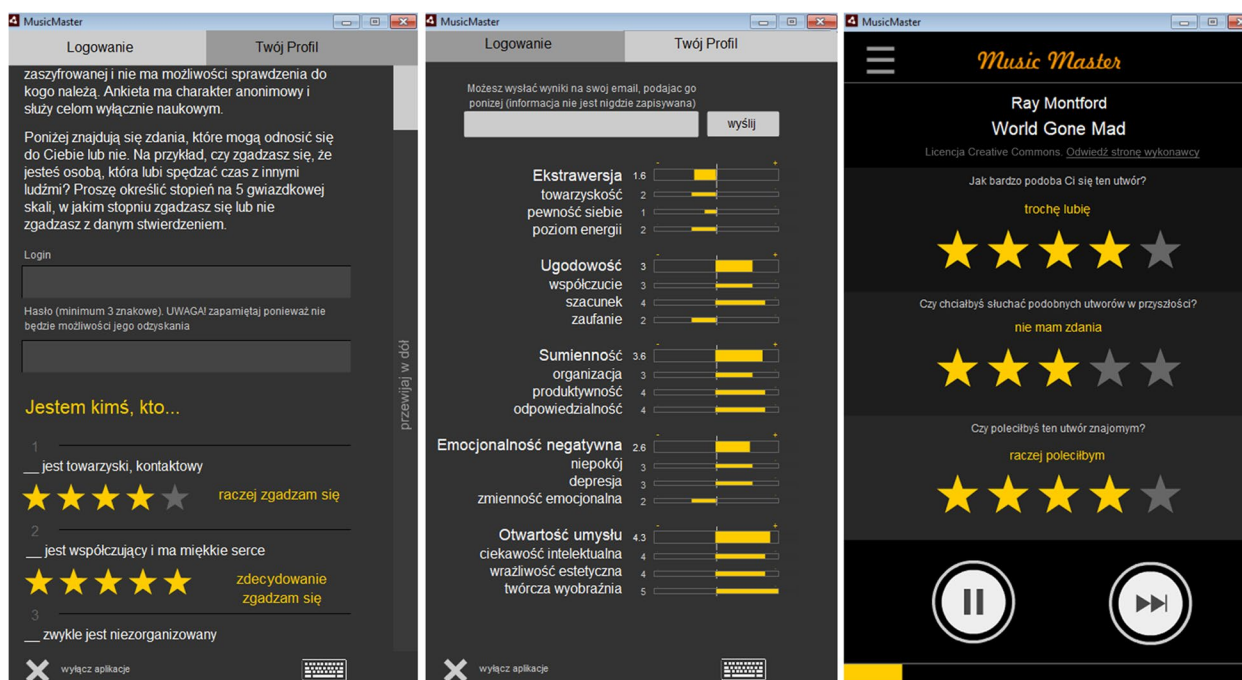
Fig. 1 The main screens of the Music Master application used in the experiments. The left screen shows the registration process, with 60 questions from the BFI-2 questionnaire. The middle screen displays the intensity of negative and positive poles of the personality dimensions. The right screen presents a music player where the participants can rate the song in terms of three different aspects (cognitive, motivational and social), using a 5-point Likert scale. The application is for Polish users, so the user interface is in Polish

First, the user needs to create an account by assigning a username and password and then rates the phrases about oneself. The phrases came from BFI-2, e.g., "I am someone who is outgoing" or "I am someone who is compassionate" [67]. Next, the personality profile is calculated and presented visually. When saving, the data is encrypted to ensure anonymity. Setting up a new account allows the user to start listening to music. The application is prepared to propose one song at a time or to generate a set of songs as a playlist. However, only the first option was used for gathering the data described in this paper. The client part has been written in Action-Script 3.0 in the Adobe Animate CC software. It allowed easy deployment on various platforms, such as for PC or mobile applications with Android or iOS operating systems. The server part has been written in JAVA. Its role is to communicate with the client and stream audio files. It saves the music meta-data, audio features, the user's profiles, ratings, and user actions. The recommendation engine has been written in Matlab. The 29-dimensional feature vector represents each song. The description of the features is presented below.

3.1 Audio features

There were 29 features calculated from each of the songs: 11 amplitude-based features, 6 spectrum-based features, 4 high-level features, and 8 emotion-based features. They were calculated using 50 ms frame length with Hamming windowing and half-frame overlapping by means of the MIRtoolbox in Matlab [76, 77]. The values of the audio features were averaged across all the frames within the length of the audio file. Some features are based on the statistics of occurring sudden bursts of signal energy that usually corresponds to such events as notes, chords and rhythm beats. Additional information about each feature can be found in [76–78].

3.1.1 Amplitude-based features

- Attack time: the mean, standard deviation, slope, and entropy of the duration of events' attack phase, detected in the amplitude of the signal (*AttackTimeMean*, *AttackTimeStd*, *AttackTimeSlope*, *AttackTimeEntropy*).
- Attack slope: the mean, standard deviation, slope, and entropy of the average slope of events' attack phase, detected in the amplitude of the signal (*AttackSlopeMean*, *AttackSlopeStd*, *AttackSlope-Slope*, *AttackSlopeEntropy*).
- Zero crossing rate (*Zerocross*) is a simple indicator of the noisiness of the signal. It counts the average

number of times that the signal changes sign in the frame.

- *RMS* measures the global energy of the signal. It is defined as the root mean square of the energy of the amplitude.
- *Lowenergy* is the percentage of frames that show less than average energy [79].

3.1.2 Spectrum-based features

- Centroid, spread, skewness, kurtosis, flatness, entropy are statistical descriptions of spectral distribution and are described by statistical moments.

  - *Centroid* indicates the center of mass of the spectrum. It has a connection with the impression of the brightness of a sound. A higher value of centroid corresponds to a brighter sound (i.e., with more energy of the signal being concentrated within higher frequencies).
  - *Spread* is the indicator of how a spectrum is spread in the frequency domain. Noises have a high spectral spread, whereas sounds with isolated peaks in the spectrum have a low spectral spread. Noisy signals are more challenging to interpret. Spectral spread is used as an indication of the dominance of a tone because the spread is low in this case; pitched sounds have low spectral spread. For complex sounds, the spread increases as the tones diverge and decreases as the tones converge.
  - *Skewness* measures the symmetry of the distribution. A distribution can be positively skewed in the case when it has a long tail to the right, while a negatively skewed distribution has a longer tail to the left. Symmetrical distribution has a skewness of zero. For harmonic signals, the spectral skewness indicates the relative strength of higher and lower harmonics.
  - *Kurtosis* measures the flatness or non-Gaussianity of the spectrum around its centroid. It is used to indicate the "peakiness" of a spectrum. For example, if the white noise is occurring within the signal, then the kurtosis decreases.
  - *Flatness* can be used to distinguish between a harmonic (flatness close to zero) and a noisy signal (flatness close to one for white noise).
  - *Entropy* is low for a spectrum with many distinct spectral peaks and high for a flat spectrum. Spectral entropy is a measure of signal irregularity.

3.1.3 Higher-level features

- *EventDensity* estimates the average frequency of events per second.
- *PulseClarity* estimates the rhythmic clarity, indicating the strength of the beats [80].
- *Inharmonicity* estimates the number of partials that are not multiples of the fundamental frequency. It takes into account the amount of energy outside the ideal harmonic series.
- *Brightness*. Although spectral centroid can be used as brightness predictor, we decided to use an improvement to it, namely to calculate centroid only for signal energy above a particular frequency; we chose 1500 Hz [81, 82]. This feature might be used to quantify the sensation of sharpness, related to the high frequency content of a sound.

3.1.4 Emotion-based features

- *Activity*, *Valence*, *Tension*, *Happy*, *Sad*, *Tender*, *Anger*, *Fear*: emotions evoked in music can be described using two paradigms: in terms of five basic emotions (i.e., happy, sad, tender, anger, and fear) and in terms of three dimensions: activity (or energetic arousal), valence (a pleasure-displeasure continuum) and tension (or tense arousal). The output of the predictive model of emotions, found on the basis of parameters from musical signal [77, 83] gives the localization of emotional content within the five basic classes and within the three dimensions.

4 The experiment setup

In the presented work, 279 participants were invited to take part in the experiment. They were mainly students from the Faculty of Information Technology and the Faculty of New Media Arts of the Polish-Japanese Academy of Information Technology. The listening sessions were organized only for volunteers in classrooms with few students. Each participant was asked to set up an account with their personality profile in the Music Master (MM) application. It was preceded by a short presentation about the data encryption in the code because it was necessary to convince the participants that the research was entirely anonymous. Over-ear semi-open headphones AKG K-240 were used in the experiments. The participants were informed that they can listen to as many songs as they want for at least 10 min and they should not perform any other tasks on the computer.

The songs were on Creative Commons license, randomly chosen from the pool of 745 songs downloaded from the magnatune.com website. The details about the pool of songs used in our experiments are given in Table 1.

The participants were informed that they could skip the song after the minimum 20 seconds of continued listening (with the option to pause or skip to any desired point) and when the song received ratings. The three types of ratings we gathered denote answers, using a five-point Likert scale, to the following three questions:

- Q1: How much do you like this song?
  - (1) "I definitely don't like it," (2) "I rather do not like it," (3) "I have no opinion," (4) "I rather like it," and (5) "I definitely like it."
- Q2: Would you like to listen to similar songs in the future?
  - (1) "I definitely would not want to," (2) "I rather would not want to," (3) "I have no opinion," (4) "I rather would want to," and (5) "I definitely would want to."
- Q3: Would you recommend this song to your friend?
  - (1) "I definitely would not recommend," (2) "I rather would not recommend," (3) "I have no opinion," (4) "I would rather recommend," and (5) "I would definitely recommend."

The majority of music recommendation systems ask users about "how much do you like this song?" (Q1 rating type) and try to predict the same for unknown songs. This question refers to the cognitive-emotional component of the attitude towards a particular song (i.e., simply to the actual opinion and belief concerning the reaction to music). The question Q2, "would you like to listen to similar songs in the future?", refers to the motivational component of the attitude, reflecting possible engagement in future contacts with the song. It is worth noting that the prediction of future engagement with similar songs is something the recommendation systems try to do. Finally, the question Q3 "would you recommend this song to your friend?" refers to the social component of the attitude, reflecting a willingness to share a given song with the user's friends. To summarize, we can say that while Q1 refers to just intrapsychic elements of the song preference, Q2 and Q3 are markers of its more extrinsic and behavioral aspects.

5 Collected data

In total, 5278 data items have been recorded. Each item represents the ratings of a particular song by one user, according to the three questions (Q1, Q2, and Q3). The answers to these three questions are further referred to as three rating types. The collected data set contains the values of 20 personality traits, the ratings for Q1, Q2, and Q3, and audio features extracted from musical files. The data set is publicly available.

Afterwards, three user-item matrices (each containing a different rating type) with 279 rows (users) and 745 columns (songs) were created. The sparsity of the matrices is equal to 0.9764. The global averages of the ratings are 2.85 for Q1, 2.58 for Q2, and 2.28 for Q3. We also studied the relationships between personality traits and ratings. We used Pearson's correlation coefficient to measure the strength and direction of each relationship (see Fig. 2). The correlation between Q1 and Q2 equals 0.871, between Q1 and Q3 0.764 and between Q2 and Q3 0.806. Figure 3 presents the distribution of each rating type across the Likert scale.

6 Proposed methodology

The role of MRS is to predict the user's rating value for an unknown song. The prediction is perfect when it is equal to the rating value that the user would give. More formally, having the group of users U and the set of songs S, the system's task is to learn a function f, which predicts the recommendation value $r \in R$ for a song s to user u: $f(u,s) : U \times S \rightarrow R$.

Model based approaches, especially those incorporating DL techniques, can learn the recommendation function f to predict ratings with high accuracy [84]. It requires the amount of data that prevents the models from being over-fitted during the training. However, the size of our data-set is not sufficient for DL models. Moreover, we wanted to obtain high interpretability of the learning process and to analyze the results and interactions between variables in the prediction phase from a psychological point of view. This would be cumbersome or impossible in the case of DL. Therefore, we decided to implement an easy to interpret memory-based Collaborative Filtering (CF) algorithm. It utilizes the k most similar users (user-based) or similar items (item-based) for predicting rating for a given item, i.e., song [1, 2]. Cosine similarity is one of the most common measures used to calculate the similarity of two vectors of ratings [2], and we decided to use this measure.

Table 1 The number of songs per genre used for the experiment

Genre	Number of songs	Genre	Number of songs
Classical	123	World	142
Jazz	63	Hard rock	113
Alternative rock	145	Electronic rock	42
Electronica	117		
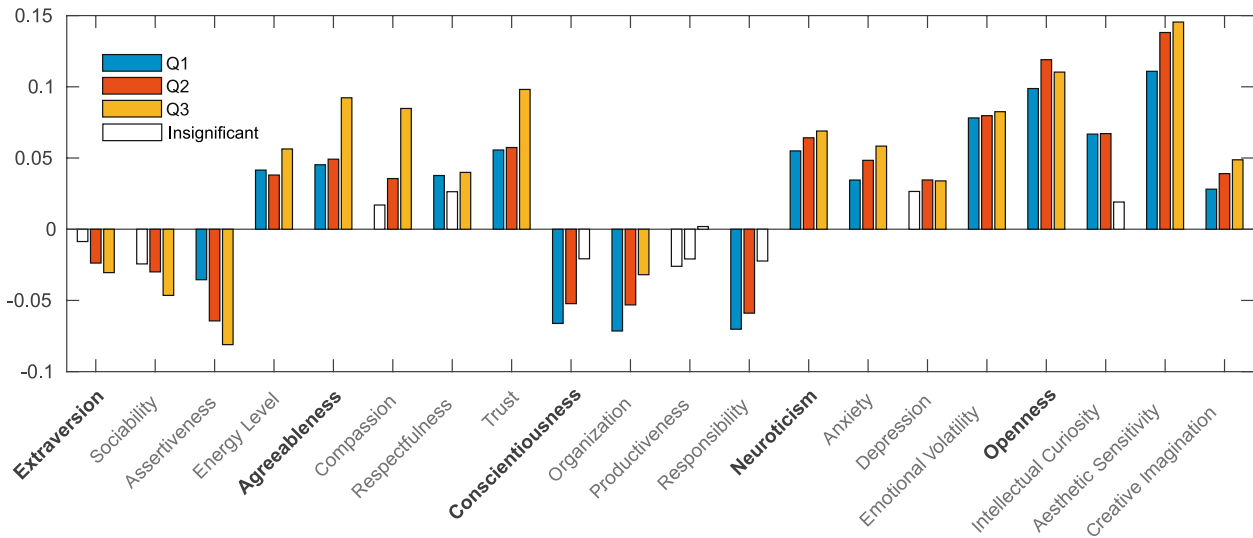
Fig. 2 Pearson's correlation coefficients, describing correlations between three rating types and personality domains. The Big Five main traits are marked in bold. Statistically insignificant correlations ($p > 0.05$) are marked in white
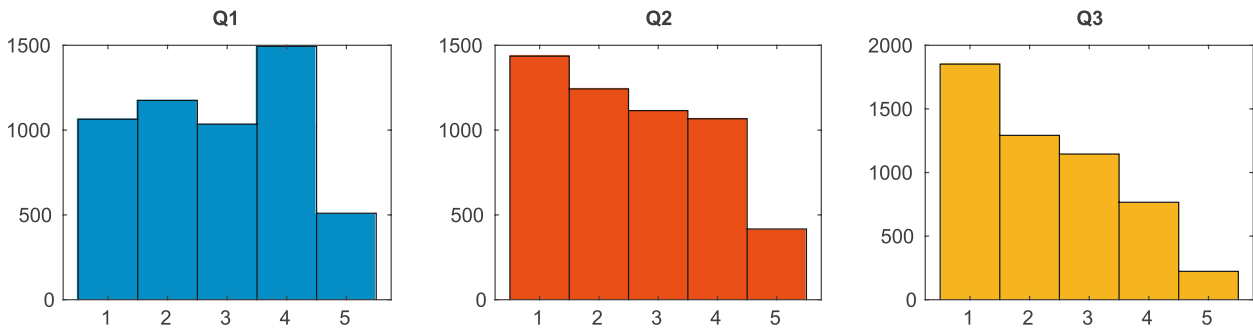
Fig. 3 The histograms of each rating type in the collected data set

For rating similarity calculations, the item and user ratings were first normalized using $rnorm_{u,i} = \mu + b_i + b_u$, to remove user and item bias. The $rnorm_{u,i}$ represents the normalized rating for user u and item i, μ denotes the global rating average, b_i and b_u are item and user bias, respectively. The biases are calculated as the difference between the global average and the average item or user ratings.

In order to determine the set of k most similar users (user-based) or items (item-based), we first calculate a similarity matrix for each approach, using cosine similarity, and k most similar users/items are found in the corresponding similarity matrix. In an item-based approach, the rating prediction for a song s and a user u is determined according to the following formula:

$$predItemBased(u,s) = \frac{\sum_{n \in K} sim(n,s) * (r_{u,n})}{\sum_{n \in K} sim(n,s)} \quad (1)$$

where $sim(n, s)$ denotes the similarity between the song s and its n'th most similar neighbor. The $r_{u,n}$ is the rating for the n'th item given by the user u.

In a user-based approach, the rating prediction can be defined according to the following formula:

$$predUserBased(u,s) = \frac{\sum_{n \in K} sim(n,u) * (r_{s,n})}{\sum_{n \in K} sim(n,u)} \quad (2)$$

where $sim(n, u)$ denotes the similarity between the user u and its n'th most similar neighbor. The $r_{s,n}$ is the rating for the item s given by the n'th user.

Next, the user and item based approach were combined in the following formula of the hybrid rating prediction:

$$predHybrid(u,s) = \frac{\sum_{n \in K} sim(n,s) * (r_{u,n}) + \sum_{n \in K} sim(n,u) * (r_{s,n})}{\sum_{n \in K} sim(n,s) + \sum_{n \in K} sim(n,u)} \quad (3)$$

Besides the similarity of ratings, we also used the similarity of audio features (instead of $sim(n, s)$) and personality

domains (instead of $sim(n, u)$) in our experiments. These data were normalized to have zero mean and standard deviation equal to one (z-score normalization), and cosine similarity was applied.

We evaluated the experiments by calculating Root Means Square Error (RMSE), using the 10-fold cross validation approach (10-CV) to evaluate the predictions of ratings. Therefore, the "recommendation quality" in the further text refers to the quality measured by RMSE obtained via the 10-CV procedure, and the lower the RMSE, the higher the recommendation quality. We will report the results for all three rating types: Q1, Q2, and Q3.

7 Experiments

In our experiments we studied two main hypotheses:

1. The recommendation quality differs when employing various personality domains (user-based approach) or audio features (item-based approach).
2. There is a difference in recommendation quality when using solely Big Five personality traits, or their low-level facets (using a hybrid approach).

In order to examine these hypotheses, baseline recommendation quality values (in terms of RMSE) were calculated first for various settings. First of all, a global average value of ratings was calculated as a baseline prediction. Next, we calculated baseline RMSE values for simple user-based and item-based CF. Subsequently, the similarity of ratings ($sim(n, u)$ and $sim(n, s)$) were replaced with the similarity of all personality traits and the similarity of all the audio features. Finally, we calculated baselines for the hybridized approaches. The results of the baseline RMSE values are presented in Table 2.

In order to study the influence of individual personality traits on the quality of music recommendations, we used a user-based CF. The influence of individual audio features was examined using an item-based approach. In order to measure the similarity between individual personality domains and individual audio features,

which are 1-dimensional vectors (scalars), $1 - d$ was applied as a similarity measure (instead of cosine similarity), where d denotes Euclidean distance, The results are presented in Figs. 4 and 5.

For studying the differences in recommendation quality between Big Five and their low-level personality facets, we used the hybrid model for rating prediction (see Eq. 3). First, we calculated the similarity values for simplified models, namely for each pair consisting of one personality trait and one audio feature, and these values were applied to calculate predictions, for each rating type (see Fig. 6). Next, from all performed experiments, two minimum RMSE values were chosen for each rating type: (1) belonging to one of the Big Five traits and (2) belonging to one of the personality facets. Together with their corresponding audio feature, these results were saved for further experiments. The pair (personality dimension and audio feature) that gave the lowest RMSE for each rating type, will be further called the "best pair". Therefore, we obtained 6 best pairs, i.e., two pairs for each of the three ratings Q1, Q2, and Q3.

In the next steps, we gradually improved the results. We started with the two pairs, for which minimal values of Q1 are obtained (see Fig. 6), i.e., curiosity, tender, and Openness, tender. Next, for each of the two previously selected best pairs, the next best pair was added and selected in the same manner as the first one. Namely, we added one personality trait (domain or facet) and one audio feature, together with the previous pair yields minimal RMSE for Q1. The difference was that the first selection used Euclidean distances (as we had one-dimensional vectors, for which cosine distance would not work), and in the next steps, cosine similarities were applied (as in this case we had multi-dimensional vectors). Every selection was performed in two ways: selecting only among Big Five domains and only low level facets. This process was repeated step by step until the RMSE error started to grow. In each step, we reported the minimum RMSE results. The same procedure was also performed for Q2 and Q3. The results are presented in Fig. 7.

Table 2 RMSE (10-CV) calculated in baseline experiments, for each rating type. The k denotes the number of neighbors used for prediction, chosen experimentally

Experiment name	RMSE for:	Q1	Q2	Q3
Item-based with ratings similarity ($k = 50$)		1.192	1.173	1.052
User-based with ratings similarity ($k = 10$)		1.326	1.313	1.199
Item-based with all features similarity ($k = 10$)		1.163	1.144	1.029
User-based with all personalities similarity ($k = 10$)		1.365	1.361	1.262
Hybrid with ratings similarity ($k_s = 50, k_u = 10$)		1.180	1.163	1.043
Hybrid with all personalities and features similarity ($k_s = 10, k_u = 10$)		1.150	1.139	1.028

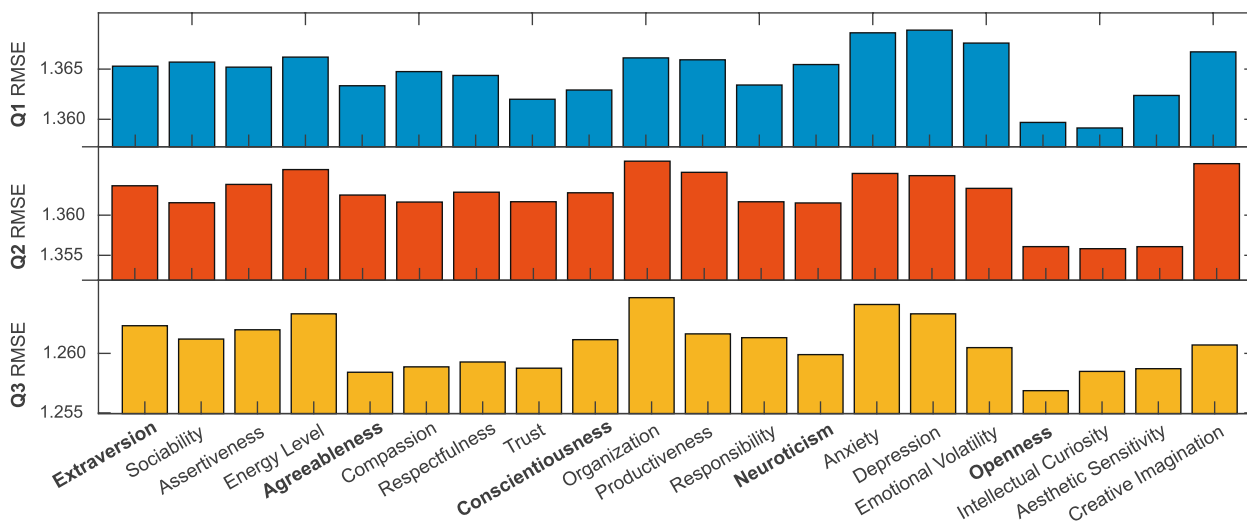**Fig. 4** The comparative analysis of recommendation quality in a user-based approach, with singular personality domains used in similarity calculations
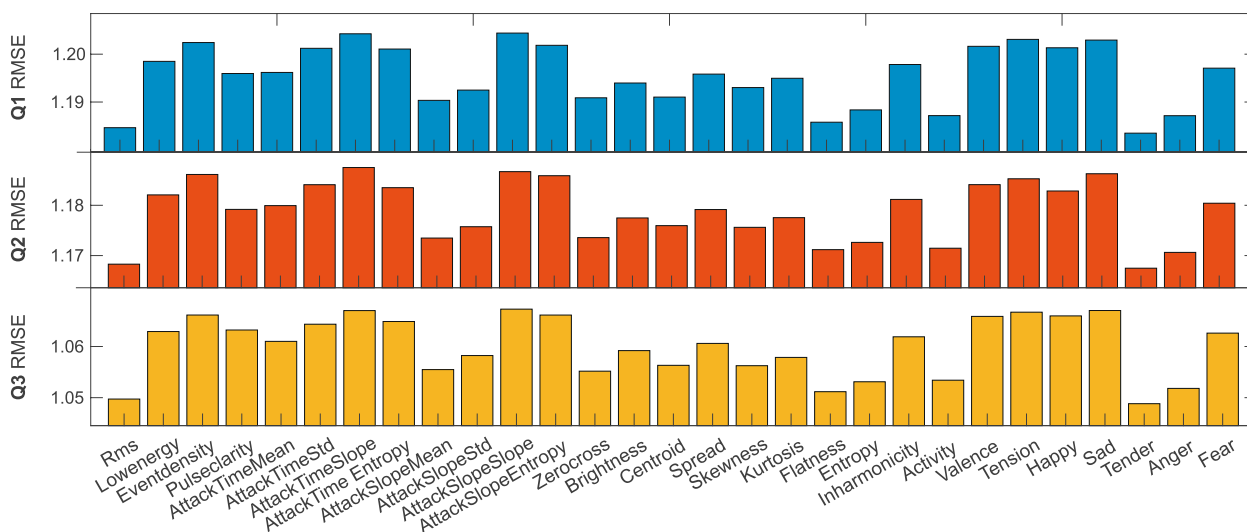

Fig. 5 The comparative analysis of recommendation quality in an item-based approach, with singular audio features used in similarity calculations

8 Results and discussion

Every comparative analysis of results in the description below will concern the values of Q1, unless indicated otherwise.

Aesthetic Sensitivity (Openness's facet) has the highest and positive correlation with all rating types (see Fig. 2). Interestingly, Assertiveness (Extraverion's facet) negatively correlates with all ratings. We believe that this can be explained by the genres used in our experiment (classical, world, jazz, hard rock, alternative rock, electronic rock, and electronica). Rentfrow et al. [85] show that people with high Openness usually prefer more complex music, like blues, jazz, folk, and rock, than Extravert people, who usually appreciate upbeat music like hiphop, funk and electronic [53]. Our experiments corroborate these findings.

As shown in Fig. 2, persons of high Openness usually give higher ratings, in contrast to the persons of high Extraversion, who usually give lower ratings. Additionally, the genres used in the experiments seem to be preferred by persons high in the trait of Openness. However, as described in [44], people of high Openness have broader musical tastes (and enjoy more genres) than Extraverted people. Therefore, they may rate the music higher because they generally like to listen to it, not only because of the preferred genres. To summarize, even
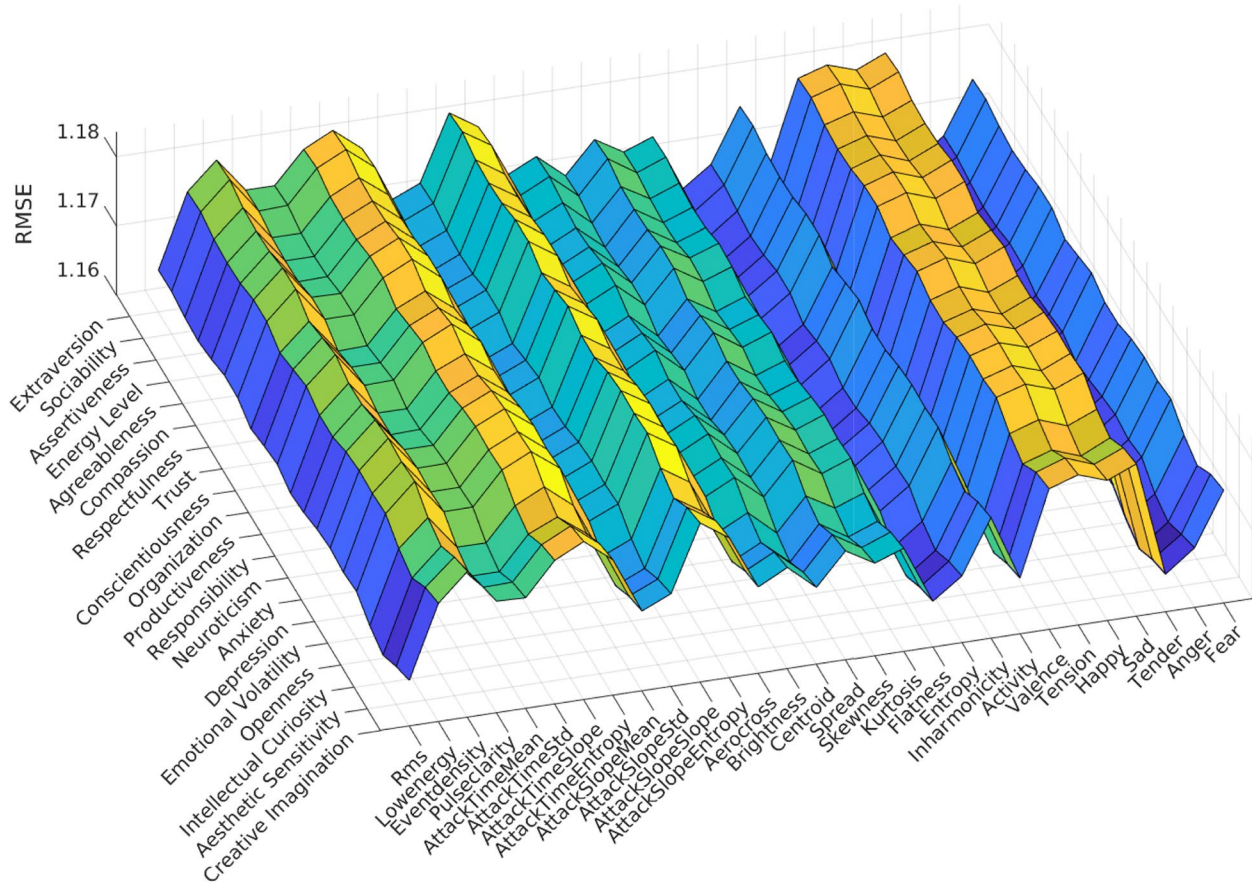
Fig. 6 The prediction error for Q1 ratings using a hybrid model based on the similarity of a single personality trait and a single audio feature. The minimum point (best pair) with respect to the 15 personality facets is obtained for curiosity and tender (RMSE = 1.1628). The best pair with respect to the Big Five personality domains is obtained for Openness and tender (RMSE = 1.1639)

though we found statistically significant correlations between ratings and personality domains, these correlations are relatively weak. The performed meta-analysis described in [86] confirms weak connections between personality and five-dimensional MUSIC factors for music preferences. The authors in [52] also confirm that associations between personality and acoustic features exist, though this association is relatively weak. Nevertheless, it is worth noting that some lower-order facets show higher correlation with ratings than their main personality domains (see Fig. 2).

Looking at the baseline results presented in Table 2, we can conclude that predicting Q2 and Q3 gives lower RMSE than predicting Q1 in all performed experiments. This means that our models make more accurate predictions for Q2 (*how much the user wants to listen to similar songs in the future*) than for Q1 (*how much the user likes the song*). Furthermore, the models perform even better when predicting Q3 (*how much the user would like to share the song with friends*). We believe these differences can be explained by the different distribution of

these rating types (see Fig. 3). In the case of Q2 and Q3, we can see that participants tended to give lower ratings more often than for Q1. Therefore, a system that predicts lower ratings for Q2 and Q3 will achieve lower RMSE. Q1 refers to the opinion or belief concerning a particular song that the listener has just heard, and therefore it could be treated as a somewhat superficial aspect of the attitude. In contrast, Q2 and Q3 reflect socio-motivational and therefore more behavioral aspects of the attitude towards the music, requiring more engagement. Therefore, Q2 and Q3 can be seen as concerning more profound psychological characteristics, more strongly related to stable personality dispositions (traits).

When comparing all the rating-based CF results, we can see that the user-based approach performs much worse than the item-based one (1.326 vs 1.192). It is not surprising as the user-item matrix had only 275 users but 745 songs. Thus, the model had a more limited number of user neighbors for making the prediction, compared to songs. Regarding the item-based approach, replacing rating with audio feature-based similarities improves the
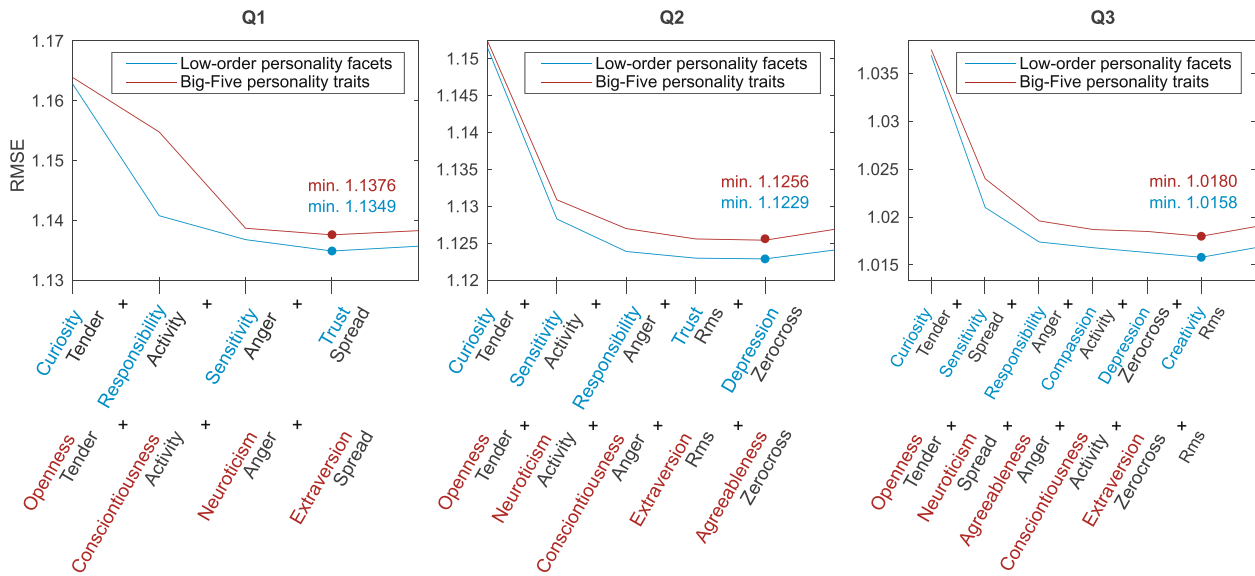
Fig. 7 The recommendation quality when gradually adding consecutive the best pairs to the previous ones. They were selected in two ways: only the best results belonging to the low-level personality facets (blue colors) and only the best results belonging to the Big Five personality domains (red color). This gradual improvement method revealed the moment (the subset of personality domains and audio features), marked by dots on each graph, after which the errors started to grow

results (1.192 vs 1.163). The reason is that the similarity of the audio features expresses the actual nearness of the songs (taking into account their audio content) better than the similarity of rating vectors. On the other hand, we have not observed the improvement when replacing ratings with the personality traits' similarities in the case of the user-based approach (1.326 vs 1.365). This result may suggest that people with similar personalities might not share similar musical tastes with the same strength as people with similar song ratings. However, in [10] the authors have shown that combining the personality similarity with a rating-based CF can bring improvement in rating prediction, compared to predictions based on rating data only. Therefore, we think we could not get a lower error because we used personality similarity alone, without the similarity of rating data. To confirm whether this combination will improve the results, as stated in [10], there is a need to combine personality and ratings similarity in the future work.

It is worth noting that only the similarity of Intellectual Curiosity gives a lower error than the similarity of all personality traits together (1.359 vs 1.365). It confirms the findings of Braunhofer et al. [87] who have shown that exploiting even a single personality trait may lead to a considerable improvement in recommendation accuracy. Still, even if the improvement was observed for a single personality trait (Curiosity), the error (1.359) is still higher than user-based CF with rating similarity (1.326). Therefore, additional experiments with more data that

combine the similarity of ratings and personalities are needed in the future.

When analyzing the recommendation quality of the hybrid model, and using the similarity of a single personality trait and a single audio feature, we can see that Intellectual Curiosity and Tender (emotion-based audio feature of music) result in the lowest error, see Fig. 6. Furthermore, this hybrid model slightly outperforms the item-based CF that considers all the personality and audio features dimensions (RMSE=1.1630 for CF vs 1.1628 for this particular hybrid model).

Prediction using the similarity of personality facets yields a lower error for all Qs than prediction based on the similarity calculated for any combination of the main Big Five personality domains (see Fig. 7). Error reduction is relatively small, but always exists. However, the main gain results from the reduction of the set of personality facets (together with the appropriate set of audio features) applied in the similarity calculations. We found that Intellectual Curiosity, Responsibility, Aesthetic Sensitivity and Trust yielded the lowest recommendation error for Q1 (see Fig. 7). In this case, the RMSE error was reduced from 1.1628 to 1.1349. The characteristic of the above set of personality facets is as follows: people of high Intellectual Curiosity desire to acquire general knowledge about the world, such as on how systems work, about mathematical relationships, what objects are composed of, etc. Responsible people are being accountable or blamed for something. Therefore, they feel a moral

obligation to behave correctly, so other people usually perceive them as reliable. Aesthetic Sensitivity describes the ability to detect and appreciate beauty wherever it exists.

We used miremotion library [83] to calculate all the audio features, including the description of music-evoked emotions, based on the analysis of the audio signal of the recordings. These emotions have been described using two representations: 1) a discrete model with five basic emotions: happy, sad, tender, anger, and fear, 2) a three-dimensional model, where these five basic emotions can also be placed: with the following dimensions: activity (energetic arousal), valence (a pleasure-displeasure continuum), and tension (or tense arousal). From Fig. 7, we can see that the similarity of activity of the tender and anger emotions evoked in music contributed most to the reduction of the recommendation error. This conclusion can also be drawn from an item-based approach, with single audio features used in similarity calculations (see Fig. 5).

It is also worth noting that, among other features, the indicator of how a spectrum is spread in the frequency (spread) contributed to reaching the minimum RMSE for Q1 and Q3. In addition, the global energy of the signal (rms) and its noisiness (zero-crossing rate) also contributed to reaching the minimum RMSE for Q2 and Q3.

Since the prediction highly depends on the similarity measure, further experiments may incorporate dimensionality reduction techniques (such as Singular Value Decomposition (SVD) or Principal Component Analysis (PCA)) together with clustering algorithms (such as k-means of Self Organizing Maps (SOM)) to infer similar users or items [88]. SOM produce clusters in an unsupervised manner from multi-dimensional data. Since the prediction could also depend on Q2 and Q3 values, the SOM can be used to group similar users or items, based on 3 rating types and also other available observations together (personality and audio features). The clusters obtained in this way can bring improvements in rating prediction [89]. Additionally, further analysis is required to investigate how motivational (Q2) and social (Q3) components of the attitude towards the song influence the cognitive-emotional (Q1) component, which may depend on the personality. This can be inferred from the dataset by employing appropriate statistical analysis. Another interesting hypothesis to check, similar to the described in [52], is the existence of the difference between features rated low and high, which may depend on the level of personality traits. The authors of this paper leave this (and also others) hypotheses to investigate by other researchers.

The recommendation quality does not depend on the prediction accuracy only. The prediction is needed for the recommender systems to build the list of songs for which the highest prediction of Q1 is obtained, indicating that the user would probably like to listen to these songs. Therefore, the songs are added to the recommendation list in the order of increasing RMSE for Q1 prediction. However, the user may actually prefer listening to other songs at the moment.

In our opinion, it seems reasonable to select the song for which the predictions of Q1 and Q2 were both high. This means that the system could recommend songs similar to those the user would like to listen to in the future (Q2) and reject those for which Q2 was low. It presumably would increase user satisfaction with the recommended items. As far as Q3 rating is concerned, the system could favor the songs that received high Q3 ratings from the user's friends. However, the link between "being the nearest neighbor" (used in the recommendation algorithm) and "being a friend" is unclear. Another idea is to formulate a confidence measure that will tell the system how trustworthy a particular prediction is. This measure would need to incorporate additional knowledge about the interactions between three rating types, number of ratings in neighbors and maybe other factors. The authors leave these issues to be investigated in the future, and keep it as open research questions, to be discussed by other researchers.

One of the limitations of our experiments is that we used the random selection of music from the magnatune.com website. It offers both Eastern and Western music to download. As it is stated in [90], in terms of BFI, only the preferences for Western music are universal across 53 countries but we do not know whether it is the true for Eastern music as well. Additionally, the range of music genres was limited, and a more elaborate genre taxonomy would allow us to compare the results with other researchers [49, 50, 53], in terms of the preferences to genres by personality traits. In the future study, there may also be a control question that verifies answers related to personality types. Additionally, there is still an opened question, how our findings correlate with real world scenarios and how the preferred use of music (relaxing or jogging) influences the way participants rate the music in the experimental controlled environment with the use of headphones. However, the most important conclusion from the experiments performed in this paper is that utilizing BFI-2 (instead of BFI) is worth considering with every rating type.

9 Conclusions

This article describes the effect of utilizing BFI-2 personality domains in the music recommendation systems on the recommendation error. The BFI-2 allowed performing the analysis with more granularity due to

the availability of low-order facets of Big Five personality domains. We collected the personality profiles and three music rating types (related to cognitive, motivational and social components of the attitude towards the music) from 279 users of the newly developed Music Master application. In addition, 29-dimensional vectors of audio features were incorporated into the analysis. To the best of our knowledge, a dataset with BFI-2 personality profiles, three rating types, and audio features has never been published before.

The experiments with our hybrid recommendation model showed interesting interactions between personality domains and audio features. It turned out that only the several low-order personality facets were enough to obtain the lowest recommendation error. The Intellectual Curiosity, Responsibility and Aesthetic Sensitivity decreased the error significantly for predicting all three rating types. It is essential to note, when using memory-based methods, any combination of Big Five personality traits produced a higher error than lower-order personality facets. However, there is still an open question whether the results scale to the real world scenarios or to model based methods.

The experiment also revealed the subset of audio features that contributed most to obtaining the lowest error. These features refer to the activity of tender and anger emotions (i.e., two basic emotions, tender and anger, as represented along the activity axis in 3-dimensional space) evoked in music. These features were calculated based on the analysis of the audio contents of the recordings. More details about the predictive models of emotions can be found in [83].

We performed our experiments on a small dataset (5278 ratings from 279 users) and a relatively simple recommendation model based on user or item similarity. Unfortunately, our initial trials with training Singular Value Decomposition (SVD) caused overfitting with the dataset due to its relatively small size. Therefore, a more extensive setup and even live system, working in real time, are required to prove that the reported subset of personality domains scales well with different recommendation algorithms. Nevertheless, the proposed simple hybrid model allowed a detailed analysis, based on the similarity of users and the similarity of songs.

An additional conclusion is that, instead of implementing the complete BFI-2 questionnaire, it is more practical and more effective to implement only a small subset of its questions. We observed that the best trade off between the performance and the number of questions is to have the following three personality traits: Intellectual Curiosity, Aesthetic Sensitivity and Responsibility, and the following three audio features: tender, anger, and activity (see Fig. 7). When adding additional ones, the error improvement is negligible. Therefore, instead of 60 questions (4 questions per personality facet), only the 12 of them would result in a better recommendation performance and higher user satisfaction than a full questionnaire.

The authors hope that other researchers will find the data set practical and stimulating to design other experiments, and prove other hypotheses that relate to three aspects of ratings (Q1, Q2, and Q3), recommendation models, and personalities.

Authors' contributions
MK implemented the Music Master application, gathered and analyzed the data. AW and KS interpreted the results, substantively revised them and improved the language and descriptions. WS provided the idea of incorporating BFI-2 into the research and the results validation from the psychological point of view. All authors read and approved the final manuscript.

Authors' information
M.K, MSc, endlessly fascinated by technology and music, he has found curiosity driving his research in the direction where music and technology meet together. He conducts research on music processing by deep neural networks, while working to complete his doctoral dissertation. Mariusz is also a full-stack web developer. What excites him more than anything else is the prospect of combining all of his skills and interests together: web development, machine learning, music, and sound processing.
A.W. PhD, DSc is a computer scientist, specializing in multimedia. She is presently an Associate Professor and the Head of Multimedia Laboratory at the Polish-Japanese Academy of Information Technology (PJATK), Warsaw, Poland. Additionally, she is also an associate member of the Graduate Faculty at the University of North Carolina at Charlotte. She has always been interested in music, and graduated from the F. Chopin State School of Music (Second Level) in Gdansk. Her scientific interests include multimedia, music and audio information retrieval, human-computer interaction, as well as automated identification of emotions from various signals, data mining, and computer graphics. She has co-authored over 100 scientific works.
K.S (PhD, DSc) is a computer scientist. He received his PhD in 2009 in the area of human-computer interaction. His PhD thesis was on a multimodal speech synthesis system, the first non-commercial system of this kind in Poland. In 2020, he earned his DSc. His areas of expertise include voice quality issues, as well as classification and the digital processing of speech signals using electroglottography. Krzysztof Szklanny is the author and co-author of many papers, and he has also participated in several national and international research projects. Dr. Szklanny is also a professional photographer.
W.S., PhD, is a personality psychologist specializing in research on personality structure. In particular, he is interested in basic dimensions and higher-order factors of personality, as well as in personality disorders and optimal functioning. Together with Jan Cieciuch and Tomasz Rowiński, he developed the Circumplex of Personality Metatraits - a synthetising model of personality.

Availability of data and materials
The dataset analyzed during the current study are available in the figshare repository, https://dx.doi.org/10.6084/m9.figshare.19678962.

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
All authors have approved the paper for being published.

Competing interests
The authors declare that they have no competing interests.

References

1. J.B. Schafer, D. Frankowski, J. Herlocker, S. Sen, in *The adaptive web*. Collaborative filtering recommender systems (Springer, Berlin Heidelberg, 2007), p. 291–324
2. J.L. Herlocker, J.A. Konstan, A. Borchers, J. Riedl, in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. An algorithmic framework for performing collaborative filtering (Association for Computing Machinery, New York, 1999), p. 230–237
3. Y. Tao, Y. Zhang, K. Bian, in *2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*. Attentive context-aware music recommendation (IEEE, Hangzhou, 2019), p. 54–61
4. F. Ricci, L. Rokach, B. Shapira, in *Recommender systems handbook*. Introduction to recommender systems handbook (Springer, Berlin, Heidelberg, 2011), p. 1–35
5. J. Herce-Zelaya, C. Porcel, J. Bernabé-Moreno, A. Tejeda-Lorente, E. Herrera-Viedma, New technique to alleviate the cold start problem in recommender systems using information from social media and random decision forests. Inf. Sci. **536**, 156–170 (2020)
6. S. Ojagh, M.R. Malek, S. Saeedi, A social-aware recommender system based on user's personal smart devices. ISPRS Int. J. Geo-Inf. **9**(9), 519 (2020)
7. L.A.G. Camacho, S.N. Alves-Souza, Social network data to alleviate cold-start in recommender system: A systematic review. Inf. Process. Manag. **54**(4), 529–544 (2018)
8. T.Z. Gizaw, H. Dong Jun, A. Oad, Solving cold-start problem by combining personality traits and demographic attributes in a user based recommender system. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **7**(5), 231–239 (2017)
9. V. Tiwari, A. Ashpilaya, P. Vedita, U. Daripa, P.P. Paltani, in *ICT Systems and Sustainability*. Exploring demographics and personality traits in recommendation system to address cold start problem (Springer, Singapore, 2020), p. 361–369
10. R. Hu, P. Pu, in *Proceedings of the fifth ACM conference on Recommender systems*. Enhancing collaborative filtering systems with personality information (Association for Computing Machinery, New York, 2011), p. 197–204
11. R.R. McCrae, P.T. Costa, *Personality in adulthood: a five-factor theory perspective*. (Guilford Press, New York, 2003)
12. M. Schedl, Deep learning in music recommendation systems. Front. Appl. Math. Stat. **5**, 44 (2019)
13. F. Fessahaye, L. Perez, T. Zhan, R. Zhang, C. Fossier, R. Markarian et al., in *2019 IEEE International Conference on Consumer Electronics (ICCE)*. T-recsys: a novel music recommendation system using deep learning (IEEE, Las Vegas, 2019), p. 1–6
14. M. Khoali, A. Tali, Y. Laaziz, in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*. Advanced recommendation systems through deep learning (Association for Computing Machinery, New York, 2020), p. 1–8
15. R.T. Irene, C. Borrelli, M. Zanoni, M. Buccoli, A. Sarti, in *2019 27th European Signal Processing Conference (EUSIPCO)*. Automatic playlist generation using convolutional neural networks and recurrent neural networks (IEEE, A Coruna, 2019), p. 1–5
16. M.F. Aljunid, M. Dh, An efficient deep learning approach for collaborative filtering recommender system. Procedia Comput. Sci. **171**, 829–836 (2020)
17. S.H. Chang, A. Abdul, J. Chen, H.Y. Liao, in *2018 IEEE International Conference on Applied System Invention (ICASI)*. A personalized music recommendation system using convolutional neural networks approach (IEEE, Chiba, 2018), p. 47–49
18. P. Knees, M. Schedl, B. Ferwerda, A. Laplante, User awareness in music recommender systems. Personalized Hum.-Comput. Interact. 223–252 (2019)
19. C. Bauer, A. Novotny, A consolidated view of context for intelligent systems. J. Ambient. Intell. Smart Environ. **9**(4), 377–393 (2017)
20. P.N. Juslin, P. Laukka, Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. J. New Music Res. **33**(3), 217–238 (2004)
21. A.C. North, D.J. Hargreaves, Situational influences on reported musical preference. Psychomusicology J. Res. Music Cognit. **15**(1–2), 30 (1996)
22. K. Bai, K. Kawagoe, in *Proceedings of the 2018 10th International Conference on Computer and Automation Engineering*. Background music recommendation system based on user's heart rate and elapsed time (Association for Computing Machinery, New York, 2018), p. 49–52
23. S. Lavanya, G. Saranya, K. Navin, in *2017 International Conference on IoT and Application (ICIOT)*. Weather based playlist generation in mobile devices using hash map (IEEE, Nagapattinam, 2017), p. 1–7
24. P. Álvarez, F. Zarazaga-Soria, S. Baldassarri, Mobile music recommendations for runners based on location and emotions: the dj-running system. Pervasive. Mob. Comput. **67**, 101242 (2020)
25. J.H. Su, H.H. Yeh, S.Y. Philip, V.S. Tseng, Music recommendation using content and context information mining. IEEE Intell. Syst. **25**(1), 16–26 (2010)
26. J. Chen, P. Ying, M. Zou, Improving music recommendation by incorporating social influence. Multimed. Tools Appl. **78**(3), 2667–2687 (2019)
27. D. Wu, in *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. Music personalized recommendation system based on hybrid filtration (IEEE, Changsha, 2019), p. 430–433
28. R. Wang, X. Ma, C. Jiang, Y. Ye, Y. Zhang, Heterogeneous information network-based music recommendation system in mobile networks. Comput. Commun. **150**, 429–437 (2020)
29. Y. Jin, N.N. Htun, N. Tintarev, K. Verbert, in *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. Contextplay: evaluating user control for context-aware music recommendation (Association for Computing Machinery, New York, 2019), p. 294–302
30. P.N. Juslin, J. Sloboda, *Handbook of music and emotion: theory, research, applications*. (Oxford University Press, Oxford, 2011)
31. S. Gilda, H. Zafar, C. Soni, K. Waghurdekar, in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. Smart music player integrating facial emotion recognition and music mood recommendation (IEEE, Chennai, 2017), p. 154–158
32. Z. Hyung, J.S. Park, K. Lee, Utilizing context-relevant keywords extracted from a large collection of user-generated documents for music discovery. Inf. Process. Manag. **53**(5), 1185–1200 (2017)
33. M. Polignano, P. Basile, M. de Gemmis, G. Semeraro, in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. Social tags and emotions as main features for the next song to play in automatic playlist continuation (Association for Computing Machinery, New York, 2019), p. 235–239
34. P.S. Lopes, E.L. Lasmar, R.L. Rosa, D.Z. Rodríguez, in *Proceedings of the XIV Brazilian Symposium on Information Systems*. The use of the convolutional neural network as an emotion classifier in a music recommendation system (Association for Computing Machinery, New York, 2018), p. 1–8
35. A.V. Iyer, V. Pasad, S.R. Sankhe, K. Prajapati, in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. Emotion based mood enhancing music recommendation (IEEE, Bangalore, 2017), p. 1573–1577
36. D. Ayata, Y. Yaslan, M.E. Kamasak, Emotion based music recommendation system using wearable physiological sensors. IEEE Trans. Consum. Electron. **64**(2), 196–203 (2018)
37. S. Kulkarni, S.F. Rodd, Context aware recommendation systems: A review of the state of the art techniques. Comput. Sci. Rev. **37**, 100255 (2020)
38. L. Xu, X. Wen, J. Shi, S. Li, Y. Xiao, Q. Wan et al., Effects of individual factors on perceived emotion and felt emotion of music: based on machine learning methods. Psychol. Music. **49**, 1069–1087 (2020)

39. P.N. Juslin, J.A. Sloboda, *Music and emotion: theory and research*. (Oxford University Press, Oxford, 2001)

40. S. Dhelim, N. Aung, M.A. Bouras, H. Ning, E. Cambria, A survey on personality-aware recommendation systems. Artif. Intell. Rev. **55**(3), 2409–2454 (2022)

41. P.J. Rentfrow, S.D. Gosling, The do re mi's of everyday life: the structure and personality correlates of music preferences. J. Pers. Soc. Psychol. **84**(6), 1236 (2003)

42. P.G. Dunn, B. de Ruyter, D.G. Bouwhuis, Toward a better understanding of the relation between music preference, listening behavior, and personality. Psychol. Music **40**(4), 411–428 (2012)

43. P.J. Rentfrow, L.R. Goldberg, D.J. Levitin, The structure of musical preferences: a five-factor model. J. Pers. Soc. Psychol. **100**(6), 1139 (2011)

44. J. Bansal, M.B. Flannery, M.H. Woolhouse, Influence of personality on music-genre exclusivity. Psychol. Music. **49**, 1356–1371 (2020)

45. R.L. Zweigenhaft, A do re mi encore: A closer look at the personality correlates of music preferences. J. Individ. Differ. **29**(1), 45–55 (2008)

46. J. Bansal, M. Woolhouse, in *ISMIR*. Predictive power of personality on music-genre exclusivity (Proceedings of the 16th International Society for Music Information Retrieval Conference, Malaga, 2015), p. 652–658

47. S.B. Kaufman, Opening up openness to experience: A four-factor model and relations to creative achievement in the arts and sciences. J. Creat. Behav. **47**(4), 233–255 (2013)

48. T. Chamorro-Premuzic, A. Furnham, Personality and music: can traits explain how people use music in everyday life? Br. J. Psychol. **98**(2), 175–185, (Wiley-Blackwell, Hoboken, 2007)

49. B. Ferwerda, M. Tkalcic, M. Schedl, in *Proceedings of the 25th conference on user modeling, adaptation and personalization*. Personality traits and music genres: What do people prefer to listen to? (2017), pp. 285–288

50. B. Ferwerda, E. Yang, M. Schedl, M. Tkalcic, Personality and taxonomy preferences, and the influence of category choice on the user experience for music streaming services. Multimed. Tools. Appl. **78**(14), 20157–20190, (Springer, Berlin, 2019)

51. B. Ferwerda, M. Tkalčič, in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. Exploring online music listening behaviors of musically sophisticated users (2019), pp. 33–37

52. M.B. Flannery, M.H. Woolhouse, Musical preference: Role of personality and music-related acoustic features. Music. Sci. **4**, 20592043211014016 (2021)

53. A. Dorochowicz, A. Kurowski, B. Kostek, Employing subjective tests and deep learning for discovering the relationship between personality types and preferred music genres. Electronics **9**(12), 2016 (2020)

54. I. Fernández-Tobías, M. Braunhofer, M. Elahi, F. Ricci, I. Cantador, Alleviating the new user problem in collaborative filtering by exploiting personality information. User Model. User-Adap. Inter. **26**(2), 221–255 (2016)

55. M. Atas, A. Felfernig, S. Polat-Erdeniz, A. Popescu, T.N.T. Tran, M. Uta, Towards psychology-aware preference construction in recommender systems: overview and research issues. J. Intell. Inf. Syst. **57**, 1–23 (2021)

56. R.P. Karumur, T.T. Nguyen, J.A. Konstan, in *Proceedings of the 10th ACM conference on recommender systems*. Exploring the value of personality in predicting rating behaviors: a study of category preferences on movie lens (Association for Computing Machinery, New York, 2016), p. 139–142

57. R. Liu, X. Hu, in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. A multimodal music recommendation system with listeners' personality and physiological signals (Association for Computing Machinery, New York, 2020), p. 357–360

58. T.T. Nguyen, F. Maxwell Harper, L. Terveen, J.A. Konstan, User personality and user satisfaction with recommender systems. Inf. Syst. Front. **20**(6), 1173–1189 (2018)

59. W. Wu, L. Chen, L. He, in *Proceedings of the 24th ACM conference on hypertext and social media*. Using personality to adjust diversity in recommender systems (Association for Computing Machinery, New York, 2013), p. 225–229

60. M. Onori, A. Micarelli, G. Sansonetti, in *Empire@RecSys*. A comparative analysis of personality-based music recommender systems (Association for Computing Machinery, New York, 2016), p. 55–59

61. F. Lu, N. Tintarev, in *IntRS@RecSys*. A diversity adjusting strategy with personality for music recommendation (2018), p. 7–14. CEUR-WS.org

62. O.P. John, L.P. Naumann, C.J. Soto, Paradigm shift to the integrative big five trait taxonomy. Handb. Pers. Theory Res. **3**(2), 114–158 (2008)

63. B.D.E. Raad, M.E. Perugini, *Big five factor assessment: introduction*. (Hogrefe & Huber Publishers, Cambridge, 2002), p. 1–18

64. M.G. Ehrhart, K.H. Ehrhart, S.C. Roesch, B.G. Chung-Herrera, K. Nadler, K. Bradshaw, Testing the latent factor structure and construct validity of the ten-item personality inventory. Personal. Individ. Differ. **47**(8), 900–905 (2009)

65. J. Golbeck, C. Robles, M. Edmondson, K. Turner, in *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. Predicting personality from twitter (IEEE, Boston, 2011), p. 149–156

66. G. Dunn, J. Wiersema, J. Ham, L. Aroyo, in *International Conference on User Modeling, Adaptation, and Personalization*. Evaluating interface variants on personality acquisition for recommender systems (Springer, Berlin, 2009), p. 259–270

67. C.J. Soto, O.P. John, The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. J. Pers. Soc. Psychol. **113**(1), 117 (2017)

68. O.P. John, S. Srivastava et al., The big five trait taxonomy: History, measurement, and theoretical perspectives. Handb. Pers. Theory Res. **2**(1999), 102–138 (1999)

69. H.J. Eysenck, Dimensions of personality: 16, 5 or 3?—criteria for a taxonomic paradigm. Personal. Individ. Differ. **12**(8), 773–790 (1991)

70. M.C. Ashton, K. Lee, Empirical, theoretical, and practical advantages of the hexaco model of personality structure. Personal. Soc. Psychol. Rev. **11**(2), 150–166 (2007)

71. J. Cieciuch, W. Strus, V. Zeigler-Hill, T.K. Shackelford, *Two-factor model of personality*. (Springer, Cham, 2018)

72. W. Strus, J. Cieciuch, Are the questionnaire and the psycho-lexical big twos the same? towards an integration of personality structure within the circumplex of personality metatraits. Int. J. Pers. Psychol. **5**, 18–35 (2019)

73. L.R. Goldberg, An alternative "description of personality": the big-five factor structure. J. Pers. Soc. Psychol. **59**(6), 1216 (1990)

74. M.A.S.N. Nunes, Recommender systems based on personality traits. Ph.D. thesis, Université Montpellier II-Sciences et Techniques du Languedoc (2008)

75. M. Tkalcic, L. Chen, in *Recommender systems handbook*. Personality and recommender systems (Springer, Berlin, 2015), p. 715–739

76. O. Lartillot, P. Toiviainen, in *International conference on digital audio effects*. A matlab toolbox for musical feature extraction from audio, vol. 237 (SCRIME and the LaBRI, Bordeaux, 2007), p. 244. https://dafx.labri.fr/main/dafx07_proceedings.html

77. O. Lartillot, Mirtoolbox 1.7. 2 user's manual. Oslo: University of Oslo. [Google Scholar] (2019)

78. F. Alías, J.C. Socoró, X. Sevillano, A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. Appl. Sci. **6**(5), 143 (2016)

79. G. Tzanetakis, P. Cook, Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. **10**(5), 293–302 (2002)

80. O. Lartillot, T. Eerola, P. Toiviainen, J. Fornari, in *ISMIR*. Multi-feature modeling of pulse clarity: design, validation and optimization (Proceedings of the 9th International Conference of Music Information Retrieval, Drexel University, Philadelphia, 2008), p. 521–526

81. A. Pearce, T. Brookes, R. Mason, Modelling the microphone-related timbral brightness of recorded signals. Appl. Sci. **11**(14), 6461 (2021)

82. P.N. Juslin, Cue utilization in communication of emotion in music performance: Relating performance to perception. J. Exp. Psychol. Hum. Percept. Perform. **26**(6), 1797 (2000)

83. T. Eerola, O. Lartillot, P. Toiviainen, in *Ismir*. Prediction of multidimensional emotional ratings in music from audio using multivariate regression models (Proceedings of the 16th International Society for Music Information Retrieval Conference, Kobe, 2009), p. 621–626

84. S. Zhang, L. Yao, A. Sun, Y. Tay, Deep learning based recommender system: A survey and new perspectives. ACM Comput. Surv. (CSUR) **52**(1), 1–38 (2019)

85. D. Rawlings, V. Ciancarelli, Music preference and the five-factor model of the neo personality inventory. Psychol. Music **25**(2), 120–132 (1997)

86. T. Schäfer, C. Mehlhorn, Can personality traits predict musical style preferences? a meta-analysis. Personal. Individ. Differ. **116**, 265–273 (2017)

87. M. Braunhofer, M. Elahi, F. Ricci, in *Information and communication technologies in tourism 2015*. User personality and the new user problem in a context-aware point of interest recommender system (Springer, Berlin, 2015), p. 537–549

88. F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, Recommendation systems: Principles, methods and evaluation. Egypt. Inform. J. **16**(3), 261–273 (2015)

89. M. Nilashi, O. bin Ibrahim, N. Ithnin, Hybrid recommendation approaches for multi-criteria collaborative filtering. Expert Syst. Appl. **41**(8), 3879–3900 (2014)

90. D.M. Greenberg, S.J. Wride, D.A. Snowden, D. Spathis, J. Potter, P.J. Rentfrow, Universals and variations in musical preferences: A study of preferential reactions to western music in 53 countries. J. Pers. Soc. Psychol. **122**(2), 286 (2022)

Publisher's Note