Wiafe *et al.*
EURASIP Journal on Audio, Speech, and Music Processing (2025) 2025:31
https://doi.org/10.1186/s13636-025-00419-0

EURASIP Journal on Audio, Speech, and Music Processing

EMPIRICAL RESEARCH

Open Access

Emotional response to music: the Emotify + dataset

Abigail Wiafe[1*], Sami Sieranoja[1], Abedin Bhuiyan[1] and Pasi Fränti[1]

Abstract

We performed a survey with 181 volunteers who were tasked to listen to 400 musical extracts from four different genres (rock, pop, classical and electronic) and reported the emotions they perceived along with their intensity. The result is a public dataset called Emotify + with 10 different emotions. It can serve as a research tool in behavioural analysis, sentiment analysis, content analysis and automatic music creation. It can also be used for training small-scale supervised models for various machine learning tasks or simply as ground-truth data for evaluating such methods. In this paper, we provide a detailed report of the dataset and perform a statistical analysis to show the connection of emotions with music genres and other factors. Additionally, we present a baseline predictive model that uses audio features to predict the predominant emotions in a song excerpt. We evaluated two classifiers: support vector machine (SVM) and k-nearest neighbor (KNN). The KNN model significantly outperformed SVM across all performance metrics, achieving a high ROC AUC score (0.81 vs. 0.53), suggesting a more reliable classification. The findings reveal KNN as an effective baseline for music emotion classification in the Emotify dataset, particularly given the complexity of a multiclass task.

Keywords Music emotion recognition, Music, Emotions, Perceived emotions

1 Introduction

The interaction between intelligent music software agents and human users has made it possible to access numerous music genres, artists, albums and releases on streaming platforms. Cooke [1] described music as *a language of emotion*, whereas Juslin et al. [2] suggested that musical experiences aim to influence emotion. Thus, individuals use music for therapeutic purposes [3], alter their emotions or match their current situation to comfort themselves or relieve their stress. As music evokes different forms of subjective emotions in listeners, several studies have attempted to investigate the impact of music and emotions in various domains. These include machine learning, cognitive science, psychology, sociology and neuroscience [4].

Music is a pervasive social phenomenon that makes individuals to experience emotions. However, the emotions that are experienced are not entirely understood. *Music emotion recognition* (MER) aims to detect the inherent emotional expressions of the people listening to it. It also provides insight into music, manages personal music collections, and recommends music for therapeutic and emotional treatments [5]. The expression of emotion does not require any training, and these emotions can be measured using biophysical indicators, such as electroencephalography (EEG), electromyogram (EMG), heart rate variability (HRV)and electrodermal activity (EDA), or self-reporting tools, such as interviews, surveys, questionnaires, rating scales, and self-assessment.

These methods assess the affective reactions associated with the listener and the resultant emotions that are expressed. One of the major challenges is the lack of a standard method for identifying and analyzing emotions. Most existing assessments are based on psychological-emotional models, which are often ineffective. Existing

*Correspondence:
Abigail Wiafe
abiwiafe@uef.fi; wiafe@cs.uef.fi
[1] School of Computing, University of Eastern Finland, P.O. Box 111, Joensuu, Finland

methods are classified into *categorical* and *dimensional* models [6]. Categorical models are identified by words or adjectival phrases such as happy, sad, fear and anger to classify emotions, whereas the dimensional model represents emotions in a two-dimensional space using polarities such as arousal (exciting/calming) and valence (positive/negative).

With considerable effort put into analyzing genre and musical instrument classification, music emotion recognition is a topic that requires extensive evaluation due to the subjective nature of emotions in music [7, 8]. Notably, discrepancies exist in the assessment of music among listeners and the judgment of individual listeners can fluctuate over time, influenced by their emotional states [9]. Variations in emotional judgment between listeners can be ascribed to personal differences, musical style, the functional context of the music and confusion between perceived and induced emotions in music.

It is essential to distinguish between perceived and induced emotions [7, 8]. Perceived emotions are the properties of songs recognized by listeners through the assessment and interpretation of musical qualities. Induced emotions are felt by listeners and entail a physiological response to music. For example, a listener who characterizes Beethoven's Ode to Joy as a happy (perceived emotion) piece of music might feel annoyance (induced emotion) instead of happiness if hearing it too often. In this study, we focused on perceived emotions. However, there remains a possibility that individual listeners may confuse these two aspects.

We designed an experiment to investigate the emotions perceived after listening to song excerpts and created a new dataset (*Emotify+*) to capture emotional responses to music. This addresses the need for a standardized and publicly available dataset for MER research. Ten distinct emotion labels (*happy*, *sad*, *amusing*, *annoying*, *anxious*, *relaxing*, *dreamy*, *energizing*, *joyful* and *neutral*) were selected based on previous studies on music emotions [10, 11]. A few less-common emotions from Warrenburg's taxonomy, such as tension, tender and nostalgic, were excluded to maintain a focused and non-redundant set. These emotions were considered too nuanced or overlapped with the selected categories in the context of our study. The aim was to ensure that the emotions were labelled clearly and consistently.

Notably, we included joyful as a distinct label from happy to capture intense positive affect. Despite the overlap between these two emotional states, we opted to retain both to emphasize a subtle yet significant difference in emotional intensity, which is consistent with prior studies that distinguish between moderate and high-arousal positive emotions, ensuring both sensitivity and consistency in our classification of emotions. Each selected emotion was rated on a 5-point Likert scale (1 = extremely poor, 2 = bad, 3 = average, 4 = good, and 5 = excellent), reflecting the perceived intensity.

Accordingly, Kim et al. [12] and Yang et al. [13] suggest that systems incorporating more than four or five distinct emotions are more proficient in addressing the challenges related to emotion classification. In light of this finding, we employed the *Emotify* dataset, which contains 400 excerpts of 60-s songs from four genres (rock, pop, classical and electronic), to create the EF Music annotation tool and collect a benchmark dataset called the Emotify + dataset. This advancement is intended to enhance the detail and accuracy of emotion recognition in music-paving way for the emergence of more emotionally intelligent music applications.

The main contribution of this study is the introduction of the Emotify + dataset. It contains the same music data as the original Emotify dataset, but with added emotion annotations. Data was collected using a music tool.[1] The dataset is publicly available to other researchers to encourage reproducibility and open-source research. The second contribution is the music tool called EF Music, which is a web music player. The tool is available for others to better understand how annotations are provided. A listener can listen to music, select and rate their emotions. The intent was to listen to music and report the emotions perceived by the music. Most listeners were unfamiliar with the songs included. The selected songs are not mainstream music, which one would hear on the radio or find from Spotify. For example, the artist Falling You has only 108 monthly listeners on Spotify and Domased, only 7. This is a tiny fraction compared to popular artists like the Beatles, with over 33 million monthly listeners. Although we cannot guarantee that the listeners were unfamiliar with every song, the fact that these songs were not mainstream suggests that the majority were likely unfamiliar with them. The unfamiliarity of songs help to remove bias related to songs that people have listened to several times. The rest of this paper is organized as follows. Section 2 reviews related work on MER and some existing datasets. Section 3 discusses the Emotify + dataset. Section 4 presents the methodology (participants, annotation tools, and evaluation approaches). The results are presented in Sect. 5, and conclusions are drawn in Sect. 6.

2 Literature review

Early studies on MER established a significant distinction between perceived emotions, which are expressed or conveyed by music, and induced emotions, which pertain

[1] http://cs.uef.fi/ml/musicemotions/app/

to the listener's emotional experience. In practice, these aspects often overlap during annotations. Some listeners mistakenly confused perceived and induced emotions. This conceptual differentiation has led to the development of more detailed models and annotation strategies in affective computing and music psychology.

Previous research has demonstrated that musical mood is associated with various acoustic features such as tempo (rhythmic speed), harmony and timbre (tone color). For instance, sad music is often characterized by slow tempo and softer dynamics, whereas happy music is energetic and brighter.

Music emotion recognition (MER) is a research field that intends to recognize emotions evoked by music using a computational model [13]. With the application of machine learning methods, MER models can be trained to map perceived music emotions using acoustic features. This enables music retrieval, indexing and organization [14]. It is also used in music emotion variation detection [15], emotion recognition of music clips [16], classification of song lyrics [17], automatic playlist generation [18, 19], exploitation of lyrical information [18] and bimodal approaches [20]. In general, MER applications in machine learning are divided into three types: emotions, features and classification models [18].

MER is multidisciplinary, and various approaches have been implemented. Gutierrez Paez et al. [21] developed a platform to annotate the emotional content in music. Participants discovered new music based on the emotional content of the music and assisted with the gathering of ground truth data, which consisted of participants' mood, demography, language, 4721 annotations and 1161 music excerpts for MER algorithms. Zentner et al. [22] compiled a comprehensive list of emotion-coloured adjectives and asked participants to rate the frequency of the emotions perceived in their preferred music. These studies identified varied responses in terms of music genre and the types of music emotions perceived or felt. The rating of music-specificity of words, tested in both lab and concert environments, suggested that the interpretation of mood words varies between different music genres.

In this study, we identified a set of emotion labels based on the union of [10] and [11] label sets. However, not all the labels were retained. The labels were systematically evaluated for overlap, uniqueness and relevance to emotions evoked by music. After a thorough review, 'tender' was excluded since it substantially overlapped with relaxing and dreamy. The exclusions were based on clear criteria, including semantic redundancy and minimizing confusion among emotionally related categories, to ensure a more distinct and manageable label set. This technique prevents the assumption of empty intersections between the proposed emotion categories and instead insists on an explicit justification for each label's inclusion or exclusion.

Juslin et al. [23] measured listeners' musical emotion using 12 rating scales (happiness-elation, sadness-melancholy, surprise-astonishment, calm-contentment, interest-expectancy, nostalgia-longing, anxiety-nervousness, pride-confidence, anger-irritation, love-tenderness, disgust-contempt and awe-admiration) with two words denoting a unipolar scale which emphasized the interpretation of the emotion category or similarity. Their results concluded that positive emotions were more common in listeners' reactions to music than negative emotions. Music tends to induce basic and complex emotions.

Kallinen [24] performed a survey on emotion ratings by music professionals and observed that six basic emotions (joy, sadness, anger, fear, surprise and disgust) were present in Western art music. Joy and sadness were more common than fear, anger, and surprise. Vieillard et al. [25] concluded that emotions such as happiness, sadness, fear and peacefulness can be recognized in relatively short stimuli (9–16 s), and although positive emotions can be aroused and perceived by music, negative emotions are more often perceived [26].

Relying solely on self-reported emotions introduces subjectivity and inherent limitations to MER research. Although self-reports are often essential for feasibility and practicality, they can be affected by individual differences, such as mood, attention, personal biases, individual interpretation of emotional labels and cultural background. As a result, the perceived reported emotions may not accurately reflect the physiological or unconscious emotional states evoked by music. To address this, integrating physiological measures such as heart rate, skin conductance, or facial expression analysis could offer more objective and complementary insights into listeners' emotional experiences. These methods are more challenging, time-consuming and intrusive than the self-reports method. To improve the validity and depth of MER datasets, future research should use multimodal approaches that incorporate both subjective self-reports and objective physiological data.

Additionally, several datasets have been created for MER, each with its own advantages and disadvantages in terms of size and annotation depth. The DEAP dataset [27] consisted of 120 songs but only 32 participants. The MER60 [28] included 60 songs and was annotated by 99 Chinese participants. Similarly, the CH496 dataset [29] contained 496 songs, yet only three (3) expert annotators of Chinese origin. In contrast, Emotify + had 400 songs and 181 participants, resulting in 3031 emotion annotations. This provides a more extensive collection of songs than MER60 or DEAP, and has a significantly higher number of annotators than CH496, thus addressing the need for a large and diverse MER dataset. Emotify + is publicly available to encourage reproducibility and further exploration in this field. However, we note that most of the participants were volunteers from Africa, specifically Ghanaian IT students being the most prevalent, resulting in a sample that is heavily skewed towards a particular culture and gender profile (77% male). The generalizability of the findings across broader cultural or gender contexts may be limited by limited cultural and gender diversity. This bias was considered in the analysis and is also noted as a limitation to be addressed in future work.

In summary, MER studies have highlighted the importance of 'rich' feature sets and comprehensive datasets. Emotify + enhances this study by presenting a new dataset with a diverse range of emotional annotations and establishing baseline findings to guide future research.

3 Dataset

To develop and evaluate MER models, a dataset with a sufficient supply of emotion labels is required. The dataset should exhibit substantial size in terms of the number of songs, participants, and annotations per subject [30]. Furthermore, the listening and rating of songs from several participants provide 'ground truth' data for ML models [7], as ML often requires a substantial amount of data for accurate prediction [31]. Capturing a wide range of emotions and annotating songs from numerous subjects are essential for providing comprehensive and general emotional responses [28, 32]. Thus, we present the Emotify + dataset, which is an extension of the Emotify music dataset [33] consisting of 400 songs (44,100 Hz, 128 kbps, 1 min each) from four different genres: classical, rock, pop and electronic music (100 tracks per genre). Participants were instructed to select the emotion(s) that they most strongly perceived while listening to each song excerpt, selecting from a predefined set of emotion labels the emotion that best reflected their subjective experiences. Although joyful was included as a label distinct from happy to capture intense positive emotions, there was an overlap between these two emotional states. Both labels were retained to reflect a subtle nuance in emotional intensity and maintain alignment with prior studies that also distinguished between these closely related but non-identical affective states.

Although the Emotify music dataset consists of 400 songs, a detailed analysis revealed that the dataset had 390 unique songs with 10 repeated songs. The pop, classical and rock genres had two, three, and five songs, respectively, repeated twice. Furthermore, out of the 400 songs with repetition, a total of 391 songs were played at least once, 275 songs were listened to less than 10 times and the longest listening time for one participant lasted for 26 min 44 s. Table 1 provides a summary of the song analysis.

To prevent data loss and ensure synchronization between the user's listening and selection actions, a custom-built backend system was implemented to automatically validate and store each annotation in real-time. The underlying data structure was designed to aid in the analysis by linking each rating to the participants' demographics and listening context while maintaining anonymity. For each rating, the data contained the following information: song ID, song name, rating, emotion annotation, genre, listening duration, date and time of listening and personal data of the participants, including gender, educational level and nationality. The nationality of participants in the dataset was marked as either Ghana or non-Ghana. It may be interesting for some studies to focus on data from Ghana to counter the widely known bias in scientific research in general for using test subjects from Western countries [34]. Most of the participants were volunteers from Africa, specifically Ghanaian IT students being the most prevalent, resulting in a sample that is heavily skewed towards a particular culture and gender profile (77% male). The generalizabilityof the findings across broader cultural or gender contexts may be limited bylimited cultural and gender diversity.Furthermore, considering that the sample size of 181 participants may be relatively low compared to other studies using large participant pools in music psychology, the results may raise questions about the representativeness and generalizability of the findings, especially given the

Table 1 Summary of participants profile, genre, song analysis and emotions perceived

Data	Values
Participants	181 participants students = 162 (77% male, 23% female) non-students = 19 (63% male, 37% female)
Gender	140 males, 41 females
Countries	Ghana, Finland, Bangladesh and Germany
Music dataset	Emotify music dataset
Genres in dataset	Pop (100 songs, 2 repeated) Classical (100 songs, 3 repeated) Rock (100 songs, 3 repeated) Electronic (100 songs, 1 repeated)
Songs analysis	391 songs were played at least once 275 songs were listened less than 10 times A song was listened to 8 times on average Longest listening time is 26 min 44 s
Emotions studied	Amusing, annoying, anxious, dreamy, energizing, happy, joyful, neutral, sad and relaxing

subjective and culturally influenced nature of emotional responses to music.

4 Methodology

4.1 Music selection

The Emotify music dataset [33] consists of 400 songs (44,100 Hz, 128 kbps, one minute each) from four different genres: classical, rock, pop and electronic music (100 tracks per genre). The collection was selected from the Magnatune recording company (magnatune.com) because familiarity with familiar music might precondition perceived emotions [35]. The Genres were selected by the recording company. The dataset contained music from 241 different albums created by 140 performers [33]. The songs are in English, and the files are in mp3 format.

4.2 Emotion labels

Each song excerpt was annotated with ten (happy, sad, amusing, annoying, anxious, relaxing, dreamy, energizing, joyful and neutral) emotional labels. This set contained emotions that were identified in previous studies, covering a wide range of emotions from positive to negative as well as neutral. Selectively, Warrenburg's list of emotions that were uncommon and overlapped; for instance, tenderness overlapping with relaxing or joyful was excluded to simplify the annotation task. An intensity rating of 1–5 was to be assigned to all emotions perceived in the excerpts: listeners expressed the intensity of each emotion that describes the song. This enables a song to evoke multiple emotions simultaneously, with distinct levels of intensity (for instance, Fly Free song might be rated high in both happy and annoying). This tagging system provided a more precise representation of each track.

4.3 Participants and response collection

Classification of emotions relies on annotated data, which can be costly and time-consuming if performed by professionals [36]. In light of this, we opted for crowdsourcing as an alternative method, a practice previously employed for annotating emotional datasets and sentiments, because of its cost-effectiveness [37, 38].

Study approval was granted by the review committee of the university, and the experiment was announced to IT students enrolled in a multimedia class. A total of 181 volunteers (140 males and 41 females) participated in this experiment. Most of the participants were volunteers from Africa, with Ghanaian IT students being the most prevalent, resulting in a sample that is heavily skewed towards a particular culture and gender profile (77% male). However, the generalizability of the findings across broader cultural or gender contexts may be limited by limited cultural and gender diversity.

Participation was voluntary without any reward. Participants were invited to listen to the selected songs and annotate their emotions perceived and not induced while using our customized EF Music tool (a web application). Participants had the option to perform a listening task at their convenience using their device (phone or computer). Given the selected list of emotions, a listener can freely choose an emotion that he/she perceives and rate its intensity. The instructions to use the music player were as follows.

- The participants are given a link to the webpage, instructed to register as a user and then listen to any music he/she wishes to listen to. It was not required to listen to all 400 songs, but as many as the participants felt like doing.
- After listening, the participant can select any emotion from the list and rate its intensity on a Likert scale from 1 to 5. There were no limitations to the number of emotions that could be annotated to the same song.
- Participants can skip listening and go to another song or genre at any time.
- If no emotion is perceived, the participant can either skip the annotation of that song or select 'Neutral' from the list of emotions.

To ensure data reliability, multiple quality control measures were implemented during data collection. The participants were required to register with a valid email address to prevent duplicate accounts. Listening durations were automatically recorded for each annotation, enabling the monitoring of participant engagement and the filtration of the listening time(s) where participants selected emotions within an unrealistically short time (e.g. less than 5 s) unless verified by the system. In addition, annotations that were not in line with reasonable listening behavior (such as multiple annotations within a short time) were flagged and checked manually, but no instances of fraudulent behavior were discovered.

Additionally, to understand the emotions identified, we gathered data on the perceived emotions of the 400 songs and processed them further using personal characteristics to form group-based annotations and analyze the agreement of participants belonging to these groups. The recruitment instruction urged participants to annotate honestly based on their immediate perceptions instead of musical preferences or familiarity with the song. To encourage genuine engagement with the task, the participants were not required to complete a minimum number of annotations. This goal was to emphasize annotation quality and sensitivity over volume, resulting in more reliable emotional labelling. The student participants were keen to learn about our research, which explains their relatively large participation and quality of annotations (no dummy annotations were detected). Moreover, the participants voluntarily agreed that their self-reported perceived emotions could be used for research purposes and to comply with data protection regulations, no personal data were stored.

Although data collection has practical benefits, there are significant methodological limitations that affect the validity and generalizability of the findings. Despite the practical advantages of the data-collection process, there are significant methodological limitations that affect the validity and generalizability of the findings. The use of a self-reported and predominantly student sample (mainly IT students) introduces sampling bias, which limits the representativeness of the broader population. Additionally, the absence of randomization in the song selection order may have introduced hidden biases related to genre despite the participants' freedom to skip tracks. Participants' engagement patterns (number of songs annotated per session and time of day) were observed to uncover any potential confounding factors, such as fatigue or mood swings. However, a detailed examination of these factors was not within the scope of this study.

To understand the emotions identified, we gathered data on the perceived emotions of the 400 songs and processed them further using personal characteristics to form group-based annotations and analyze the agreement of participants belonging to these groups. Figure 1 shows the data gathered using a web music player.

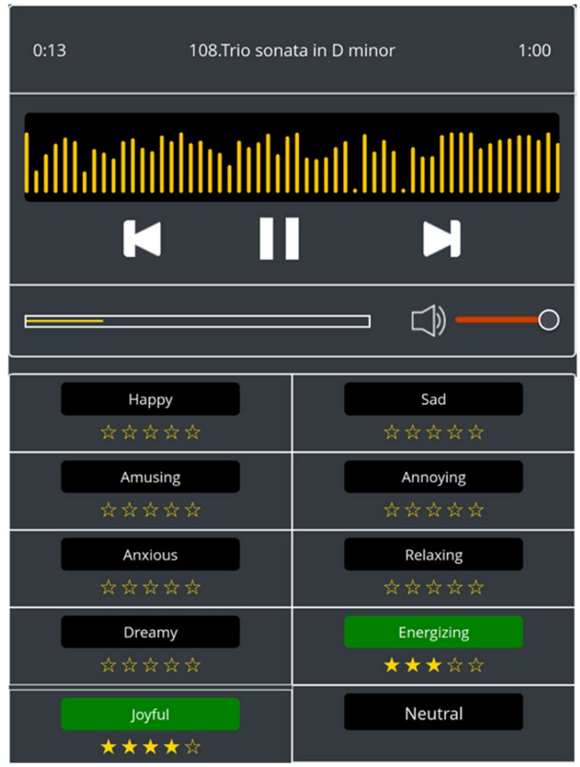

Fig. 1 Data was gathered by using a web music player

4.4 EF music annotation tool

To collect the annotations, we implemented a web-based music player called the EF Music tool, which includes basic features, such as playlist, autoplay, play/pause and next/previous buttons. The tool provides a user-friendly interface for users to listen to and rate perceived emotions. Figure 2 illustrates the system architecture of the EF Music annotation tool.

The front-end was created using standard web technologies, such as HTML, CSS and JavaScript. It was designed to run in the user's web browser, supporting both desktop and mobile environments and providing audio playback controls and rating inputs. The HTML <audio> element was utilized for playback and the JavaScript library was used to visualize waveforms. HTML forms were used for rating inputs, while the interface was tested with modern browsers and adjusted to be mobile friendly (i.e. provide a responsive layout for smaller screens). An About page in the app provides background information and instructions. In addition, the back-end application features a PHP-based web server that connects to the MySQL database. It handles the user's authentication (login/register) and serves the list of songs with metadata, such as the title and genre. When ratings are submitted, the server validates and stores the data in the database, creating a connection between each rating, the user and the song selected. User accounts, song ratings and song details were stored. The music files (in mp3 format) were streamed from the server to the client for playback upon request. This design allows users to log in from anywhere and record their annotations centrally. To maintain consistency, a decision to sort tracks by genre in the user interface was implemented in the server; thus, the playlist API returned songs grouped by genre (i.e. pop first, Classical, Rock and Electronic) rather than in a completely random order.

Tracks were ranked according to their genres using their original order. We considered two possibilities for the order: (1) randomization and (2) sorting by genre. The second choice was chosen because it makes the music collection easier to navigate, and users can better find the type of songs they prefer to listen to. This method of navigation is closer to how users navigate in music listening applications such as Spotify, and it is considered more motivating. The drawback is that earlier songs are annotated more often than later songs. This may have caused

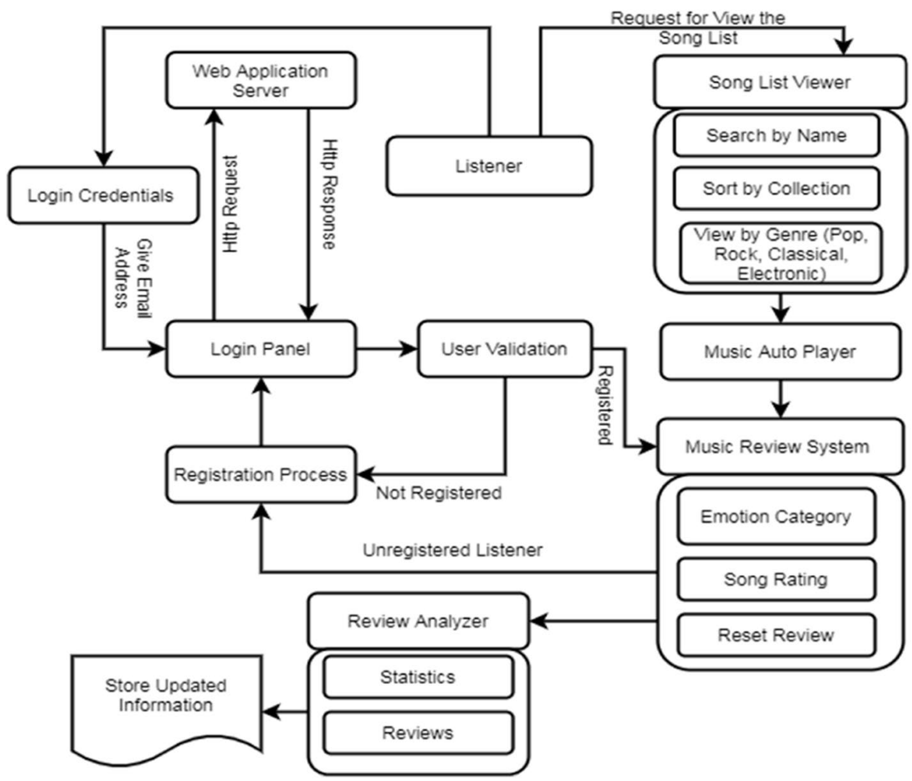

Fig. 2 System architecture of the EF Music annotation tool

some hidden bias in the results, but it mostly affected the number of times the songs were annotated. A song raising happy feelings would most likely do so no matter whether it was annotated five or twenty times. This research focused on the emotions perceived from music rather than on the songs themselves. However, it is vital to understand that although users were encouraged to listen in a 'free order' the tracks were grouped by genre to simply navigate and ensure that listeners could opt for the music they felt like enjoying at any given time. This provided participants with a personalized listening experience while preserving the structure of the dataset. It is also important to recognize that organizing by genre, rather than randomizing, could have created further subtle biases, aside from the annotation frequency. Specifically, the emotional responses reported could have been affected by order effects or participant fatigue. As participants navigate through the playlist, their emotional sensitivity might diminish, potentially reducing the number or intensity of emotional annotations for songs they experience later on. Conversely, earlier songs may provoke stronger or more diverse emotional responses because of participants' engagement at the start of the listening session. While users had the option to skip songs, the non-randomized structure could have influenced the emotional profile of the annotations systematically. Randomizing the order, even within the genres, would have decreased this risk by spreading evenly distributed fatigue effects across the dataset and ensuring that the emotional responses were less influenced by the position of each song. To enhance the internal validity of the data collection process, future implementations should consider randomizing the song order.

When a song is played, its audio waveform is displayed, and the user can select the emotions perceived at any time. A listener can select any song from the app by using the song name or genre. Users were encouraged to listen to songs in a free order of their own.

We selected the Emotify music dataset [33] which consists of 400 songs (44,100 Hz, 128 kbps, 1 min each) from four different genres: classical, rock, pop and electronic music (100 tracks per genre). The collection was obtained from *Magnatune* recording company (magnatune.com). We chose it because of its unfamiliarity with the listeners. Familiar songs may have preconditioned emotions and hidden biases.

The objective of this study was to identify the emotions perceived after listening to music. To ascertain an emotion perceived, a listener needs to register, log in and listen to any of the 400 song excerpts from the Emotify music dataset.[2]

An essential consideration in designing the data tool was determining which emotions users should have the option to select. According to psychological theories on emotions in music [39] and the arguments posited by Chiang et al. [40], there are no standardized rules for the selection of emotions to use.

We aimed to maintain a nuanced (and precise) understanding of emotions while avoiding overwhelming users with an abundance of choices. Therefore, we selected ten emotion annotations aligned with prior research where the emotions varied from 4 [41] to 13 [10]. The selection criteria focused on emotional uniqueness and interpretability, focusing on emotions that were easy for non-expert listeners to understand and recognize across various musical genres. The selected emotions closely resemble those utilized in Warrenburg [11] albeit with the exclusion of a few based on two key reasons: (1) conceptual overlap with other selected emotions, resulting in redundancy (for instance, tender was considered to overlap (or as a synonym) with relaxing and dreamy), and (2) short (60 s) music clips were less likely to be salient or difficult to perceptually identify certain emotions (such as triumphant or tense), as these emotions may need more context to be perceived. We selected the top three emotions (sad, happy and neutral) most commonly studied in music [11] and added them to this list (dreamy, relaxing, joyful, annoying, energizing, anxious and amusing) based on the emotions used by Cowen et al. [10]. Thus, the full list of selected emotions was happy, sad, amusing, annoying, anxious, relaxing, dreamy, energizing, joyful and neutral.

4.5 Emotions annotated

In total, 3031 responses were gathered from 181 participants. Participants listened to multiple songs and noted the emotions they perceived while listening to a song. We decided to keep the data collection process simple to guarantee that sufficient annotations would be obtained. Additional questions about listeners' moods and personal music tastes might have provided additional insight, as demonstrated by Flexer et al. [9]. However, we did not want to burden the listeners too much, so our primary goal would be to receive enough data. The total number of emotion annotations identified in this study is shown in Table 2 and an example of the recorded data is shown in Table 3.

We processed the data by calculating the distribution of the emotions for each song. For example, the song Fly Free was described as relaxing (23%), amusing (15%), dreamy (11%), sad (11%), happy (10%), annoying (8%) and 21% as some other emotions. Table 4 shows data examples from the top-3 songs with the highest emotions identified. Songs with the highest emotion annotations were listened to more than 20 times. Songs with the highest neutral emotions were listened to the least.

Table 2 Number of participants and emotion annotations identified

Emotion	Participants	Annotations
Relaxing	113	399
Energizing	111	319
Amusing	106	370
Dreamy	105	290
Happy	103	335
Neutral	102	218
Sad	101	260
Anxious	95	302
Joyful	95	252
Annoying	87	286

Participants could perform the annotation at any time while listening to the 60-s song extract. Additional investigation was conducted to examine annotation behaviour over time, focusing on how song listening timing correlated with various emotion selections. The behavioural insights were intended to inform future improvements in annotation protocols for music-emotional studies. Figure 3 shows how fast participants reported their emotions with respect to time while listening to music. It was observed that neutral emotion was most frequently selected as the longest (50–60 s) duration before selecting other emotions. Furthermore, neutral was the least (1%) selected when participants listened to music in less than 10 s. An opposite trend was observed for relaxing (14%) and amusing (14%), as these two highest emotions were identified most often within the first 10 s. Relaxation was used most frequently within the first 30 s.

4.6 Evaluation: baseline emotion classification

In this section, we present a baseline predictive model designed to predict the predominant perceived emotions of a song based on audio features. We explored support vector machines (SVM) and K Nearest neighbors (KNN). The selection of these two models is justified by their proven effectiveness in classification tasks, particularly in pattern recognition, feature importance analysis and data nonlinearity management. The reason for utilizing these methods is their ability to manage diverse feature sets, including the spectral and temporal characteristics of music. SVM works well in scenarios where clear decision

[2] http://www3.projects.science.uu.nl/memotion/emotifydata/

Table 3 Examples of collected data

Song ID	Genre	Song Name	User ID	Rating	Feeling	Gender
9	Pop	A warning to giants	87	1	Energizing	F
9	Pop	A warning to giants	122	4	Sad	M
16	Pop	Fly Free	51	3	Happy	F
16	Pop	Fly Free	94	3	Annoying	M
245	Rock	Juego	78	3	Joyful	M
245	Rock	Juego	102	5	Dreamy	F
305	Electronic	Beneath the skin	37	4	Dreamy	M
305	Electronic	Beneath the skin	128	2	Annoying	F
396	Electronic	Easter	117	4	Amusing	M
396	Electronic	Easter	102	5	Annoying	F

boundaries are required, whereas KNN is also effective for MER, where emotions can be clustered in the feature space based on similarity. This demonstrates the strength and complementarity of these methods in achieving high classification accuracy and efficiency.

4.6.1 Audio features classification

Audio features were extracted from 400 song excerpts using the jMIR tool (McKay and Fujinaga, 2009). Standard acoustic features that correlates with musical emotions were extracted. Specifically, we extracted and computed 13 Mel-Frrequency Cepstral Coefficients (MFCC) to capture the trimbral texture, spectral features such as Spectral Flux, Spectral Centroid (brightness) and Spectral Rolloff, tempo (Strongest Beat) measured as beats per minute to reflect the rhythmic speed, and dynamic (RME energy, zero crossing rate) because volume and articulation have a significant impact on emotions. Fifteen features were extracted from the audio files, with each feature's overall average and standard deviation. Each feature was summarized by computing the average and standard deviation across the entire duration of the excerpt, effectively reducing high-dimensional time series into compact, fixed-size features. Computations were performed using Python's Librosa audio analysis library.

These extracted audio features were converted into numerical values (essentially, a list of numbers representing a song's audio profile). Non-audio metadata and emotion annotations were excluded from analysis. The emotion label for each song was determined by selecting the dominant label (i.e., the highest score) from a set of binary-coded emotion tags. This ensured that each excerpt was assigned a distinct emotional category. These categorical labels are then transformed into a numeric format using a Label Encoder.

To assess the model performance and generalizability, we employed Stratified K-fold cross-validation, a method well suited for classification tasks involving imbalanced classes. The dataset was partitioned into ten folds, maintaining a proportional class distribution in each fold. In this method, the dataset was repeatedly split into training and testing sets across ten folds, with 80% of the data used for training and 20% used for testing in each fold. Stratification ensured that the proportion of emotion categories was preserved across all folds, thus reducing the risk of bias introduced by uneven label distributions. The cross-validation framework provided a robust estimate of the model performance and reduced the variance associated with the random train-test split. Analyzing all available data in both the training and testing roles leads to a more comprehensive evaluation than a single static partition. Both the SVM and KNN models were compared using the same ten-fold cross-validation strategy.

4.6.2 Classification

We experimented with two classic algorithms: support vector machine (SVM) and k-nearest neighbor (KNN). The SVM model incorporates a scaling step to normalize the feature set before applying a radial basis function (RBF) kernel. Various hyperparameters, including regularization strength (C) and kernel coefficient (gamma), were fine-tuned to enhance the performance of the model. Similarly, the KNN classifier was also applied to the scaled features, with a grid search conducted to determine the optimal number of neighbors (n_neighbors) and weighting mechanism (weights), which facilitated the performance evaluation. With models developed to fit particular data, it is necessary to test their performance using these data. Thus, part of the data (training data) was used to train and validate the model, whereas the test set was used to test the accuracy of the model. Both models were

Table 4 The top-3 most frequently rated songs for each emotion. The number of times a song was rated for a specific emotion is shown on the right-hand side of the table. The relative value is calculated by dividing the absolute value by the total number of times a song was rated

Emotion	Song name	Genre	Emotion count	
			Absolute	Relative
Relaxing	Fly free	Pop	14	23%
	Something about eve	Pop	10	18%
	A warning to giants	Pop	9	17%
Amusing	Fly free	Pop	9	15%
	Sooner or Later	Pop	7	18%
	Donata for violoncello	Classical	7	24%
Happy	Pais Oceano	Pop	9	30%
	A warning to giants	Pop	7	13%
	Shining Star Games	Pop	7	16%
Energizing	Frontliner Blues	Pop	16	30%
	Mercurial Girl	Pop	14	33%
	Modern Anguish	Pop	10	29%
Anxious	Something about eve	Pop	8	15%
	A warning to giants	Pop	6	11%
	Fly free	Pop	5	8%
Dreamy	Fly free	Pop	7	11%
	Something about eve	Pop	7	13%
	A warning to giants	Pop	7	13%
Annoying	Calling	Pop	7	18%
	Frontliner Blues	Pop	6	11%
	Shining Star Games	Pop	6	14%
Neutral	Give me love	Pop	2	17%
	Brilliant Day	Pop	2	67%
	Beyond Late	Pop	1	17%
Sad	Something about eve	Pop	10	18%
	A warning to giants	Pop	10	19%
	Fly free	Pop	7	11%
Joyful	Frontliner Blues	Pop	8	19%
	Shining Star Games	Pop	5	11%
	Sooner or Later	Pop	5	13%

analyzed based on the ROC AUC and F1-score as the performance metrics. The SVM achieved an ROC AUC score (0.526), indicating that it is slightly better than random guessing, and an F1-score (0.131), indicating poor performance. KNN achieved an ROC AUC score (0.805) suggesting good class separability, and an F1-score (0.294) indicated a more reliable dataset. Although both models indicate some improvement, the KNN model significantly outperforms the SVM across all metrics. KNN demonstrates a promising baseline for automatic emotion recognition, especially given the complexity of multiclass tasks.

We emphasize that MER is a challenging task, and in future work, we intend to explore more feature-learning techniques to achieve high performance.

5 Results

The findings suggest that out of the 3031 emotional responses, pop songs were mostly rated as energizing, which concurs with previous findings [42] and further supports the role of investigating students' preferences for popular songs in various grades and colleges [43]. One hundred and sixteen (116) songs were listened to more than 10 times, and the most popular song was Fly Free by Artist (Curl) on the album Ultimate Station 61 times. Fly Free, Something about Eve, A Warning to Giants, Frontliner Blues, and Shining Star Games are the top five songs, which were rated 61, 55, 54, 54 and 44 times, respectively.

5.1 Genre-based difference

Regarding the relationship between gender and musical genre preferences, the findings in Fig. 4 indicate that the pop genre was rated more often by females than males (43.3% vs. 38.5%). This supports previous findings [44],Snježana [45],Snjezana [46] that females exhibit significantly higher preferences for popular music. Classical music was also rated more by females than males (17.2% vs. 14.2%). Males rated electronic music more than females (22.2% vs. 16.4%). Despite the preferences of these genres, there were no significant differences between males and females in the perception of emotions. Boer et al. [47] observed cross-cultural differences as an important explanation for differences in the use of music. However, we could not confirm such an observation, probably due to the lack of diversity in the participants (143 from Ghana and 19 elsewhere).

The emotional profiles of the genres are shown in Fig. 5. We counted the emotion ratings for each genre and divided them by the total number of ratings for that genre. This is formally represented as:

$$\mathrm{E}(\mathrm{g},\ \mathrm{e}) = \mathrm{C}(\mathrm{g},\mathrm{e})/\mathrm{T}(\mathrm{g}) \tag{1}$$

Where E(g, e) is the proportion of emotion e in genre g, C(g, e) is the count of ratings for emotion e in genre g, and T(g) is the total number of ratings for genre g. This equation is used to compute the emotional profile in Fig. 5.

Each genre had a distinct emotional profile. Classical music was considered more relaxing (17%) and amusing (16%) than other genres. Classical music among Pop was considered the least anxious (8%). Electronic music was the most anxious (13%) and annoying (11%). It was also the most neutral (12%), least joyful (8%), and least sad (6%). Pop music was considered the happiest

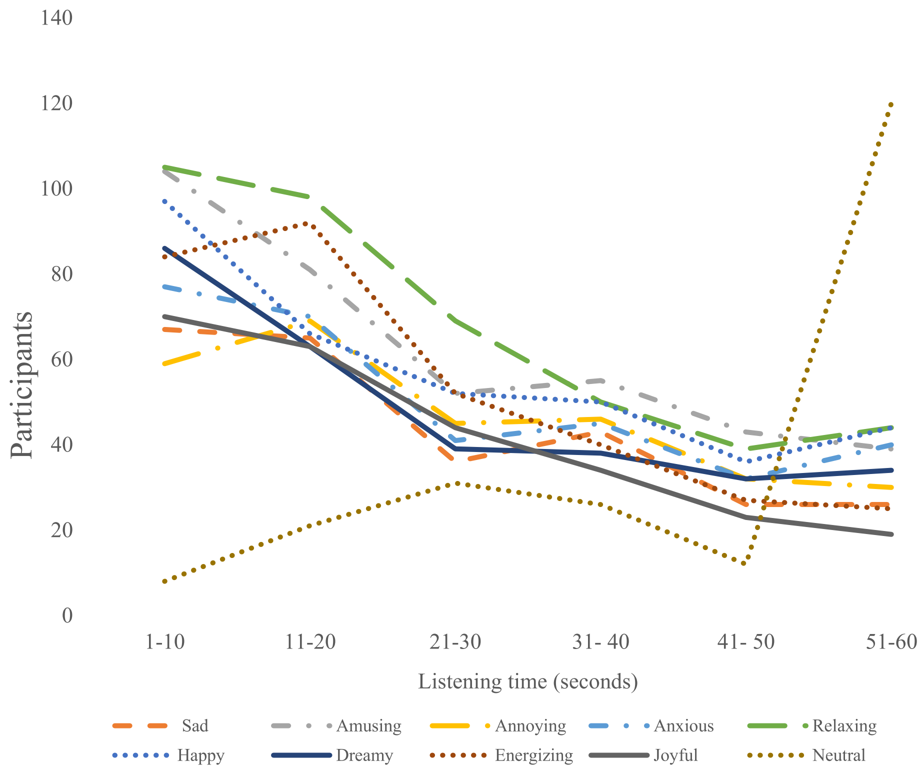

Fig. 3 Emotions and song timing frequency

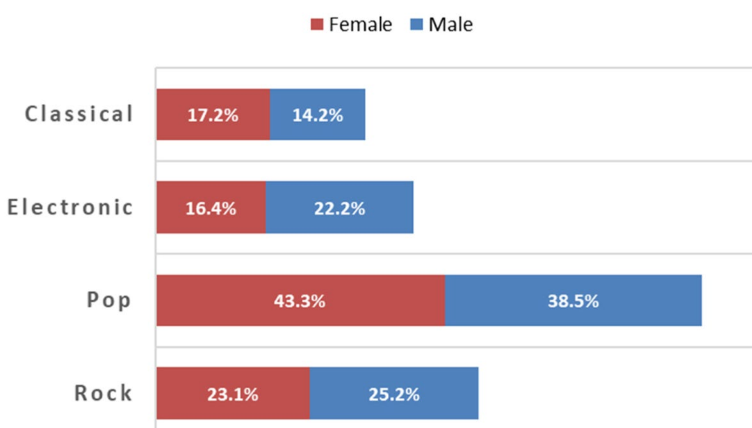

Fig. 4 Share of male and female ratings for the genres. The values represent the number of times a song from a genre was rated divided by the number of all males (or females) ratings

(12%) and had the smallest share of ratings for annoying (8%) and anxious among classical music (8%). It was also the most energizing (12%).

5.2 Gender-based differences

If we consider positive emotions (*joyful*, *happy*, *amusing*, *energizing*, *dreamy*, *relaxing*) and negative emotions (*sad*, *annoying*, *anxious*), then Classical and Pop music had the most positive emotions (11.2%, 11.3%), whereas

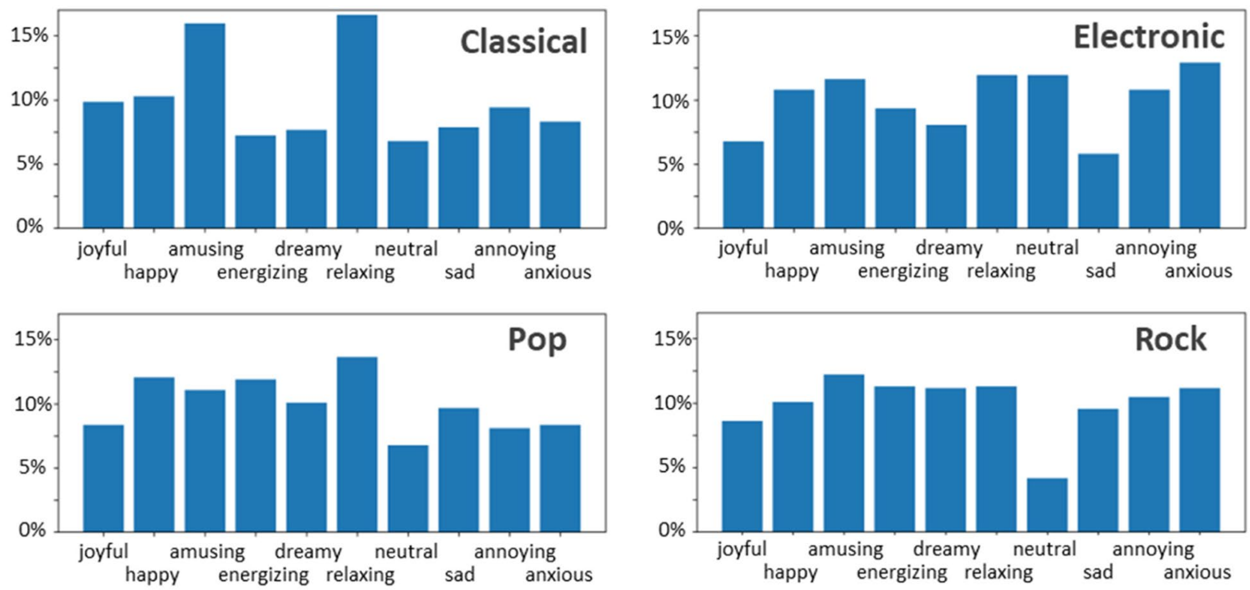

Fig. 5 Emotion profiles for each genre

Electronic and Rock had the least (9.8%, 10.8%). Electronic and Rock also had the highest mean values for negative emotions (10.4%, 9.8%), whereas Classical and Pop had the lowest (8.5%, 8.7%).

Positive emotions were more frequently perceived while listening to music than negative emotions (64% vs. 28%). Similar results were reported in a study by Juslin et al. [2], whose participants referred to positive emotions in 84% of the cases. Our results also showed that the participants identified relaxing emotions as the most frequent emotion (13%), see Table 1. This is consistent with the results of Juslin et al. [2], who found relaxation to be the most common motive for listening to music.

There were small differences in how females and males selected their emotions (Fig. 6). Males used *neutral* more often than females (8.0% vs. 5.3%). Males also

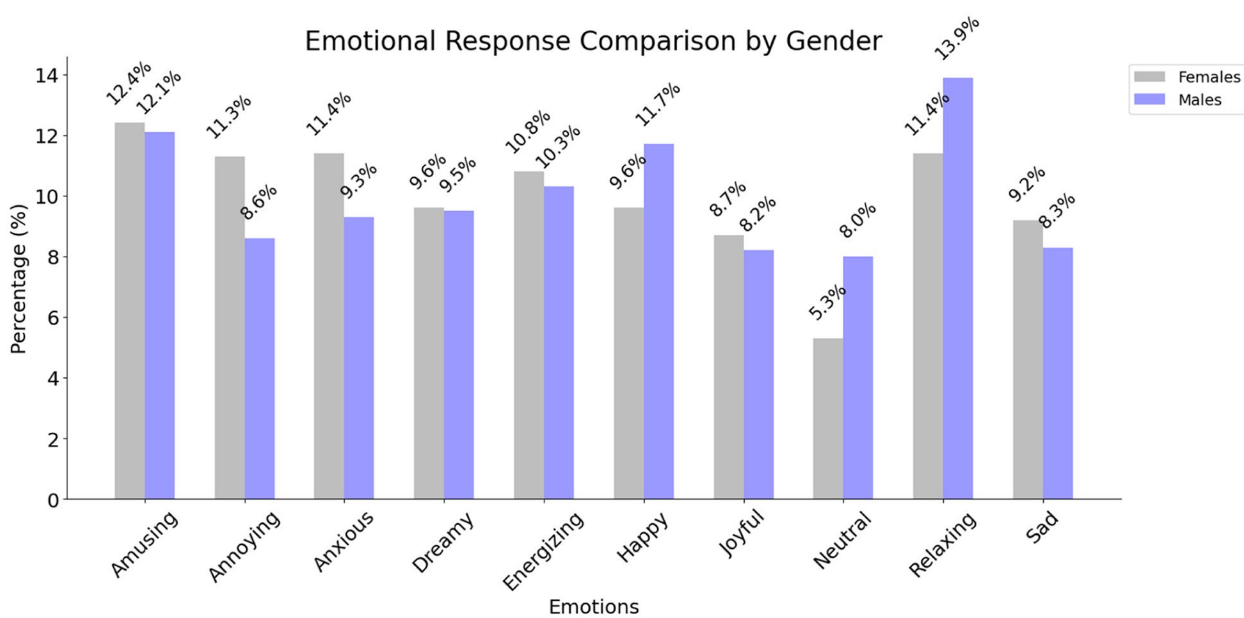

Fig. 6 Emotion profile by gender. The height of each bar is the share (%) of that emotion out of all emotion selections made by male (blue) or female (grey) participants

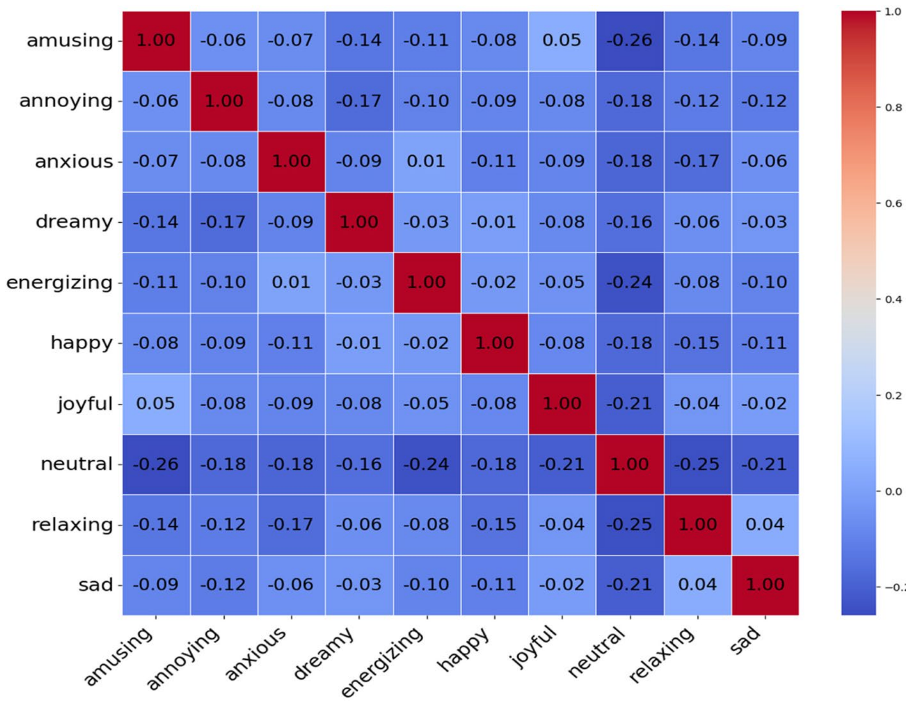

Fig. 7 Correlation matrix of emotional responses to musical stimuli

used *relaxing* more frequently (13.9% vs. 11.4%), which we consider the most neutral of the positive emotions. The actual neutral choice was selected when the participant did not have any emotional responses. This result suggests that females might be more skillful at detecting their emotions and socializing in the pop genre. Females have also slightly higher mean values for the negative emotions of *sad*, *annoying* and *anxious* than males (10.7% vs. 8.7%). A chi-square test performed on the gender difference in emotion distribution revealed that gender had a significant impact on the distribution of positive and negative emotions, the result was only marginally significant (χ^2 test, $p \approx 0.02$). There is a possibility that females are more sensitive to or are likely to express certain negative emotions when listening to music. This suggests that females may be more attuned to or more likely to express certain negative emotions when listening to music, which might result in gender differences in emotional expressions, with females being more socially encouraged or predisposed to acknowledge and report negative emotions. In addition, the results may have been influenced by potential selection bias in the dataset, as the majority of participants were male, suggesting that the music selections they chose freely reflected their preferences than females, who constituent to a small group may have encountered music that was not less aligned with their taste, leading to a more critical response, which is still a hypothesis. However, the difference observed in this study was not substantial enough to establish a significant overall trend, as further research using a larger and more diverse sample is required to investigate the gender-based distinction more thoroughly.

5.3 Correlation analysis of emotions

Additionally, a correlation analysis was performed on all emotion ratings using JASP, an open-source statistical program, to examine the relationships between the different emotional responses. This was to enhance the understanding of the emotional structure underlying listeners' experiences with music excerpts. However, the correlation analysis revealed that neutral emotion negatively correlated with most emotions, including amusing (-0.26), energizing (-0.24) and relaxing (-0.25), implying that the more neutral a state is, the less likely someone is to feel amusing, energized or relaxed. Annoying and anxious were weakly related to other emotions; annoying had a weak negative correlation with energizing (-0.10, $p < 0.05$) and dreamy (-0.17, $p < 0.001$). Anxious was moderately negatively correlated with relaxing (-0.17, $p < 0.001$) and neutral (-0.18, $p < 0.001$). The correlation between relaxing and many emotions was not strong; it showed a significant but small negative correlation with happy (-0.15, $p < 0.01$), amusing (-0.14, $p < 0.01$) and annoying (-0.12, $p < 0.05$). This implies that relaxing tends to slightly decrease happiness and amusement. Dreamy had a small negative correlation with amusing (−0.14, $p<0.01$) and energizing (−0.03, not significant). This suggests a slight tendency for dreamy to be linked to reduced amusement. The correlation matrix for the emotional responses to song excerpts is shown in Fig. 7. Overall, correlation analysis provides empirical evidence that certain emotion pairs cluster together, forming broader dimensions. For instance, happy–joyful–amusing-energizing cluster is considered positive/high-arousal, relaxing-dreamy cluster is considered positive/low-arousal, anxious-annoying cluster is considered negative/high-arousal, and sad cluster is distinctly marked as negative/low-arousal. This grouping aligns with established models that define affective states along the valence (positive/negative) and arousal (high/low) axes. Consequently, our categorical labels can be onto the 2D circumplex, for instance, happy/joyful representing positive/high-arousal, relaxing-dreamy representing positive/low-arousal, anxious-annoying reflecting negative/high-arousal, sad associated with negative/low-arousal and neutral occurrence of the center (as no emotion). This dimensional mapping improves and facilitates a logical understanding of emotional response patterns.

Furthermore, to compare multiple emotions perceived across different songs, a repeated measures ANOVA was conducted to examine the effect of listening to songs as independent variables, with expressed emotions (joyful, happy, amusing, energizing, dreamy, relaxing, sad, annoying and anxious) as dependent variables.

Mauchly's tests revealed that the assumption of sphericity was violated ($p<0.05$); therefore, Greenhouse–Geisser corrections were applied. The results demonstrated a significant effect of the emotions expressed ($F(9,3582)=16.439$, $p\leq0.001$, $\omega2=0.040$).

A post-hoc test using the Bonferroni correction was performed to determine the emotions that significantly differed from each other. Thus, emotions are either stronger or weaker in response to certain songs. Multiple comparisons indicated a significant difference (***) between neutral and other emotions, suggesting that it differed significantly from other emotions. A statistically significant difference was observed when comparing happy vs. neutral ($p<0.001$) and amusing vs. neutral ($p<0.001$). Joyful vs. anxious and sad vs. annoying had high p values (≥0.05), meaning that there was no statistically significant difference. Various comparisons with neutral emotions revealed significant differences (***), indicating that they differed significantly from other emotions. For instance, happy vs. neutral ($p<0.001$) and amusing vs. neutral ($p<0.001$) indicate strong statistical differences. Joyful vs. anxious and sad vs. annoying have high p values (≥0.05), meaning that there is no statistically significant difference. Likewise, comparing the mean differences revealed that neutral displayed substantial differences, which distinguished it from the other emotions.

Furthermore, a two-way ANOVA was performed to validate genre and gender differences in song excerpt ratings. Rating was used as the dependent variable, whereas genre and gender were the independent variables. The results revealed a significant effect of genre ($p=0.0078$), suggesting that song ratings vary notably across different genres, regardless of gender. Also, gender had a significant impact on ratings ($p=0.0379$), suggesting that male and female listeners tend to rate songs differently. Importantly, a significant relationship was identified between genre and gender, implying that the impact of genre on song ratings depends on the listener's gender and vice versa. Also, both genre and gender independently and interactively influence how songs are rated, highlighting the importance of considering demographic factors in music emotion research.

6 Conclusion

We developed a web platform to study the effects of music on emotion. This platform was used to collect emotional responses from 181 participants after listening to short pieces of music. The result is an extension of the Emotify dataset called Emotify+, which consists of participants' emotions and their intensity. The data are made publicly available for researchers to utilize in the fields of behavioral analysis, sentiment analysis, content analysis and automatic music creation.

We also studied the effects of gender and musical genre on the perceived emotions. Our results suggest that different genres of music have different effects on the listeners. Classical and pop music have a higher share of positive emotions (joyful, happy, amusing, energizing, dreamy and relaxing), whereas electronics and rock have relatively higher values for negative emotions (sad, annoying and anxious). Positive emotions were more frequently perceived in listening to music than negative emotions (64.8% vs. 28.0%). It is worth noting that negative emotions were slightly recorded higher in females (10.7% vs. 8.7% compared to males). Although the difference was slight, it could be a reflection of gender-related differences in emotional expression. Females are often discovered to be more attuned and willing to report negative emotions than males. This may suggest that female listeners either experience negative emotions intensively in response to certain musical features or feel more at ease sharing their emotions in a self-report context. To assess the robustness of the pattern and explore the underlying psychological or social mechanisms, it is important to conduct further investigations with larger and more diverse samples, due to the magnitude of the difference.

Relaxing was the most common (13%) emotion perceived after listening to music, even though the music listeners were unfamiliar to most pieces. This supports previous findings [48, 49] suggesting that familiarity does not have an impact on the song being relaxing.

The findings also indicate that participants were volunteers and enjoyed discovering new songs to identify the different emotions they perceived from listening to songs. Further studies should examine the limitations of some moderator variables (e.g., age, ethnicity, musical knowledge and environment) and the use of longer musical extracts, which could influence the results. It would also be interesting to investigate the possible structural differences between perceived and induced feelings in music, possibly by integrating physiological data into the analysis. Using self-reported emotion as the sole basis for emotion assessment can be subjective and influenced by various factors, such as mood, personal biases, individual interpretation of emotional labels and cultural background. Recognizing the limitations of this approach while acknowledging the practicality of self-reports is important. Physiological measures, such as heart rate, skin conductance or facial expression analysis, can be combined to provide a more objective insight into emotional responses, although these methods are more complex, time-consuming and resource-intensive. We also plan to investigate more genres and styles (as 400 musical extracts from four genres do not represent all genres and styles), participant profiles, more emotional labels (more than ten) and explore emotion ratings for songs longer than 60 s. We hope that our findings will enhance research in this field, as the study confirms the existence of gender differences and music genre preferences.

A correlation analysis of the emotion labels confirmed expected emotional relationships, such as happy, correlate closely with joyful, which are both considered positive-high arousal emotions. In addition, relaxing strongly correlated with dreamy, which reflects their shared position within the positive-low arousal dimension of emotion. These emotions are often associated with calm and peaceful musical settings, highlighting the role of serenity and low arousal in music perception. A two-way ANOVA revealed song ratings are influenced by both genre and gender, independently and interactively, highlighting the importance of demographic factors in music emotion research.

We further implemented an evaluated baseline model for music emotion classification to predict the dominant perceived emotions of song excerpts by using their acoustic features. KNN and SVM were implemented as performance benchmarks. The KNN model significantly outperformed SVM across all metrics in this evaluation. An ROC AUC score (0.81 vs. 0.53), suggests a more reliable classification performance. The results reveal that KNN is a promising baseline for music emotion classification in the Emotify dataset, particularly given the complexity of a multiclass task.

In summary, music appears to have the potential to be used as an effective emotion regulation tool to improve, alter and modulate emotions and mood, although the relationship is complex. It is worth mentioning that although the participants were mostly volunteers from Africa, the sample was strongly skewed toward Ghanaian IT students (77% male), which restricts cultural and gender diversity. This finding may not fully represent a wider and more diverse population owing to sampling bias. Additionally, the sample size of 181 participants was relatively small compared to other studies, which may further constrain the broader applicability of the results. These limitations highlight the need for caution when interpreting and generalizing these findings to diverse populations.

Acknowledgements
No acknowledgements.

Authors' contributions
AW proposed the methodology, conducted the experiment and wrote the original draft. Data analysis was performed by AW and SS. Software was conceptualized by AW and AB and implemented by AB. SS enhance visualization of figures and provided reviews. PF supervised the entire research and refined the manuscript. AW wrote the final manuscript with additional contributions from SS and PF. All authors read and approved the final manuscript.

Funding
No funding was received for conducting this study.

Data availability
The datasets generated and/or analysed during the current study are available in the [Mendeley Data] repository, [https://data.mendeley.com/datasets/wbk8zr9bd2/1].

Declarations

Competing interests
The authors declare no competing interests.

Received: 22 January 2025 Accepted: 23 June 2025
Published online: 30 July 2025

References

1. D. Cooke, *The language of music* (Oxford University Press, 1959)
2. P.N. Juslin, S. Liljeström, P. Laukka, D. Västfjäll, L.O. Lundqvist, Emotional reactions to music in a nationally representative sample of Swedish adults: prevalence and causal influences. Music. Sci. **15**(2), 174–207 (2011). https://doi.org/10.1177/1029864911401169
3. S. Dai, X. Ma, Y. Wang, R.B. Dannenberg, Personalised popular music generation using imitation and structure. Journal of New Music Research **51**(1), 69–85 (2022). https://doi.org/10.1080/09298215.2023.2166848
4. C. Laurier, P. Herrera, Automatic detection of emotion in music: interaction with emotionally sensitive machines. In: Machine Learning:

Concepts, Methodologies, Tools and Applications (pp. 9–33). IGI Global (2012). https://doi.org/10.4018/978-1-60960-818-7.ch509
5. A. Huq, J.P. Bello, R. Rowe, Automated music emotion recognition: a systematic evaluation. Journal of New Music Research **39**(3), 227–244 (2010). https://doi.org/10.1080/09298215.2010.513733
6. M. Zentner, T. Eerola, Self-report measures and models. In *Handbook of music and emotion: Theory, research, applications* Eds,(pp. 187–221). Oxford University Press (2010). https://doi.org/10.1093/acprof
7. J.S. Gomez-Canon, E. Cano, T. Eerola, P. Herrera, X. Hu, Y.-H. Yang, E. Gomez, Music emotion recognition: toward new, robust standards in personalized and context-sensitive applications. IEEE Signal Process. Mag. **38**(6), 106–114 (2021). https://doi.org/10.1109/MSP.2021.3106232
8. J.S. Gómez-Cañón, N. Gutiérrez-Páez, L. Porcaro, A. Porter, E. Cano, P. Herrera-Boyer, A. Gkiokas, P. Santos, D. Hernández-Leo, C. Karreman, E. Gómez, TROMPA-MER: an open dataset for personalized music emotion recognition. Journal of Intelligent Information Systems **60**(2), 549–570 (2022). https://doi.org/10.1007/s10844-022-00746-0
9. A. Flexer, T. Lallai, K. Rasl, On evaluation of inter- and intra-rater agreement in music recommendation. Transactions of the International Society for Music Information Retrieval **4**(1), 182–194 (2021). https://doi.org/10.5334/tismir.107
10. A. Cowen, D. Sauter, J.L. Tracy, D. Keltner, Mapping the passions: toward a high-dimensional taxonomy of emotional experience and expression. Psychological Science in the Public Interest **20**(1), 69–90 (2019). https://doi.org/10.1177/1529100619850176
11. L.A. Warrenburg, Choosing the Right Tune: a review of music stimuli used in emotion research. *Music Perception*, 240–258 (2020). https://doi.org/10.1525/mp.2020.37.3.240
12. Y.E. Kim, E.M. Schmidt, R. Migneco, B.G. Morton, P. Richardson, J. Scott, J.A. Speck, D. Turnbull, Music emotion recognition: a state of the art review. In Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, *Ismir,* 255–266 (2010)
13. X. Yang, Y. Dong, J. Li, Review of data features-based music emotion recognition methods. Multimedia Syst. **24**, 365–389 (2018). https://doi.org/10.1007/s00530-017-0559-4
14. J.C. Wang, Y.H. Yang, H.M. Wang, S.K. Jeng, The acoustic emotion Gaussians model for emotion-based music annotation and retrieval. In Proceedings of the 20th ACM International Conference on Multimedia, 89–98 (2012). https://doi.org/10.1145/2393347.2393367
15. L. Lu, D. Liu, H.J. Zhang, Automatic mood detection and tracking of music audio signals. IEEE Trans. Audio Speech Lang. Process. **14**(1), 5–18 (2006). https://doi.org/10.1109/TSA.2005.860344
16. J. Lee, J.H. Jo, H. Lim, J.H. Chae, S.U. Lee, D.W. Kim, Investigating relation of music data: Emotion and audio signals. Lecture Notes in Electrical Engineering **330**, 251–256 (2015). https://doi.org/10.1007/978-3-662-45402-2_37
17. J. Edmonds, J. Sedoc, Multi-emotion classification for song lyrics. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 221–235 (2021)
18. A. Flexer, D. Schnitzer, M. Gasser, G. Widmer, Playlist generation using start and end songs. ISMIR **2008**, 173–178 (2008)
19. G. Agarwal, H. Om, An efficient supervised framework for music mood recognition using autoencoder-based optimised support vector regression model. IET Signal Proc. **15**(2), 98–121 (2021). https://doi.org/10.1049/sil2.12015
20. R. Malheiro, R. Panda, P. Gomes, R.P. Paiva, Bi-modal music emotion recognition: novel lyrical features and dataset. In 9th International Workshop on Music and Machine Learning–MML 2016–in Conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases–ECML/PKDD (2016)
21. N.F. Gutierrez Paez, J.S. Gomez-Canon, L. Porcaro, P. Santos, D. Hernandez-Leo, E. Gomez, Emotion annotation of music : A citizen science approach. Collaboration Technologies and Social Computing (2021)
22. M. Zentner, D. Grandjean, K.R. Scherer, Emotions evoked by the sound of music: Characterization, classification, and measurement. Emotion **8**(4), 494–521 (2008). https://doi.org/10.1037/1528-3542.8.4.494
23. P.N. Juslin, L.S. Sakka, G.T. Barradas, O. Lartillot, Emotions, mechanisms, and individual differences in music listening : a stratified random sampling approach. *Music Perception*, **40**(1), 55–86 (2022). http://doi.org/10.1177/1029864911401169
24. K. Kallinen, Emotional ratings of music excerpts in the western art music repertoire and their self-organization in the Kohonen neural network. Psychol. Music **33**(4), 373–393 (2005)
25. S. Vieillard, I. Peretz, N. Gosselin, S. Khalfa, L. Gagnon, B. Bouchard, Happy, sad, scary and peaceful musical excerpts for research on emotions. Cogn. Emot. **22**(4), 720–752 (2008). https://doi.org/10.1080/02699930701503567
26. M. Zentner, S. Meylan, K. Scherer, Exploring musical emotions across five genres of music. In Sixth International Conference of the Society for Music Perception and Cognition, (Keele) (2000)
27. S. Koelstra, C. Mühl, M. Soleymani, J.S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, DEAP: a database for emotion analysis; Using physiological signals. IEEE Trans. Affect. Comput. **3**(1), 18–31 (2012). https://doi.org/10.1109/T-AFFC.2011.15
28. Y.H. Yang, H.H. Chen, Prediction of the distribution of perceived music emotions using discrete samples. IEEE Trans. Audio Speech Lang. Process. **19**(7), 2184–2196 (2011). https://doi.org/10.1109/TASL.2011.2118752
29. Y.H. Yang, X Hu, Cross-cultural music mood classification: a comparison on English and Chinese songs. *In Proceedings of the 13th International Society for Music Information Retrieval Conference,* ISMIR 2012, Ismir, (pp. 19–24) (2012)
30. Y. A. Chen, Y. H. Yang, J. C. Wang, H. Chen, The AMG1608 dataset for music emotion recognition, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. **2015**, 693-697 (2015). https://doi.org/10.1109/ICASSP.2015.7178058
31. L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In Journal of Big Data (Vol. 8, Issue 1). Springer International Publishing (2021). https://doi.org/10.1186/s40537-021-00444-8
32. E.M. Schmidt, Y.E. Kim, Prediction of time-varying musical mood distributions from audio. Proceedings of the 11th International Society for Music Information Retrieval Conference, *ISMIR* 2010, Ismir, 465–470 (2010)
33. S.M. Mohammad, P.D. Turney, Crowdsourcing a word-emotion association lexicon. Comput. Intell. **29**(3), 436–465 (2013). https://doi.org/10.1111/j.1467-8640.2012.00460.x
34. M. Nielsen, D. Haun, J. Kärtner, C.H. Legare, The persistent sampling bias in developmental psychology: A call to action. J. Exp. Child Psychol. **162**, 31–38 (2017). https://doi.org/10.1016/j.jecp.2017.04.017
35. E. Schubert, The influence of emotion, locus of emotion and familiarity upon preference in music. Psychol. Music **35**(3), 499–515 (2007). https://doi.org/10.1177/0305735607072657
36. A. Devitt, K. Ahmad, Sentiment analysis and the use of extrinsic datasets in evaluation. In Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008, 1063–1066 (2008)
37. A. Greenhill, K. Holmes, C. Lintott, B. Simmons, K. Masters, J. Cox, G. Graham, Playing with science: Gamised aspects of gamification found on the online citizen science project - Zooniverse. In GAME-ON 2014 15th International Conference on Intelligent Games and Simulation, 15–22 (2014)
38. K. Trohidis, G. Tsoumakas, G. Kalliris, I Vlahavas, Multi-label classification of music into emotions. *ISMIR* 2008 - 9th International Conference on Music Information Retrieval, 325–330 (2008)
39. A. Aljanaki, F. Wiering, R.C. Veltkamp, Studying emotion induced by music through a crowdsourcing game. Inf. Process. Manage. **52**(1), 115–128 (2016). https://doi.org/10.1016/j.ipm.2015.03.004
40. W.C. Chiang, J.S. Wang, Y.L. Hsu, A music emotion recognition algorithm with hierarchical svm based classifiers. International Symposium on Computer, Consumer and Control, 1249–1252 (2014). https://doi.org/10.1109/IS3C.2014.323
41. A.S. Cowen, X. Fang, D. Sauter, D. Keltner, What music makes us feel: at least 13 dimensions organize subjective experiences associated with music across different cultures. Proc. Natl. Acad. Sci. **117**(4), 1924–1934 (2020). https://doi.org/10.1073/pnas.1910704117
42. D.E. Wolfe, A.S. O'Connell, E.G. Waldon, Music for relaxation: a comparison of musicians and nonmusicians on ratings of selected musical recordings. J. Music Ther. **39**(1), 40–55 (2002). https://doi.org/10.1093/jmt/39.1.40
43. J.D. Boyle, G.L. Hosterman, D.S. Ramsey, Factors influencing pop music preferences of young people. Journal of Research in Music Education **29**(1), 47–55 (1981). https://doi.org/10.2307/3344679

44. A. Colley, Young people's musical taste: relationship with gender and gender-related traits. J. Appl. Soc. Psychol. **38**(8), 2039–2055 (2008). https://doi.org/10.1111/j.1559-1816.2008.00379.x
45. S. Dobrota, I.R. Ercegovac, K. Habe, Gender differences in musical taste: the mediating role of functions of music. Drustvena Istrazivanja **28**(4), 567–586 (2019). https://doi.org/10.5559/di.28.4.01
46. S. Dobrota, E. Reić, Adolescent's musical preferences with regard to some socio-demographic variables. Odgojne Znanosti **11**(2), 381–398 (2009)
47. D. Boer, R. Fischer, H.G. Tekman, A. Abubakar, J. Njenga, M. Zenger, Young people's topography of musical functions: personal, social and cultural experiences with music across genders and six societies. Int. J. Psychol. **47**(5), 355–369 (2012). https://doi.org/10.1080/00207594.2012.656128
48. D. Elliott, R. Polman, R. McGregor, Relaxing music for anxiety control. J. Music Ther. **48**(3), 264–288 (2011). https://doi.org/10.1093/jmt/48.3.264
49. H.S. Saragih, Pop music rivalry in Indonesia: past, present and future trends. Proceedings of the 2016 Global Conference on Business, Management and Entrepreneurship, 15, 664–669 (2016). https://doi.org/10.2991/gcbme-16.2016.125

Publisher's Note