Personal and Ubiquitous Computing
https://doi.org/10.1007/s00779-023-01721-4

ORIGINAL ARTICLE

ENSA dataset: a dataset of songs by non-superstar artists tested with an emotional analysis based on time-series

Yesid Ospitia-Medina[1,2] · José Ramón Beltrán[3] · Sandra Baldassarri[3]

Received: 31 March 2022 / Accepted: 19 March 2023

Abstract

This paper presents a novel dataset of songs by non-superstar artists in which a set of musical data is collected, identifying for each song its musical structure, and the emotional perception of the artist through a categorical emotional labeling process. The generation of this preliminary dataset is motivated by the existence of biases that have been detected in the analysis of the most used datasets in the field of emotion-based music recommendation. This new dataset contains 234 min of audio and 60 complete and labeled songs. In addition, an emotional analysis is carried out based on the representation of dynamic emotional perception through a time-series approach, in which the similarity values generated by the dynamic time warping (DTW) algorithm are analyzed and then used to implement a clustering process with the *K-means* algorithm. In the same way, clustering is also implemented with a *Uniform Manifold Approximation and Projection* (UMAP) technique, which is a manifold learning and dimension reduction algorithm. The algorithm *HDBSCAN* is applied for determining the optimal number of clusters. The results obtained from the different clustering strategies are compared and, in a preliminary analysis, a significant consistency is found between them. With the findings and experimental results obtained, a discussion is presented highlighting the importance of working with complete songs, preferably with a well-defined musical structure, considering the emotional variation that characterizes a song during the listening experience, in which the intensity of the emotion usually changes between verse, bridge, and chorus.

Keywords Musical datasets · Non-superstar artists · MRS (Music Recommender Systems) · MER (Music Emotion Recognition) · Popularity bias · Time-series approach

1 Introduction

In the last 20 years, the music industry has shown important changes in its business model, generating new challenges for all the stakeholders involved in this industry [18]. One of the most important changes is related to the music distribution process, which has presented a transition from physical storage media to fully digital media hosted through a cloud computing approach. According to the 2021 Global Music Report which summarizes statistical data from 2020, the global recorded music market grew by 7.4%, and global streaming revenues increased 18.5% [24]. This transition implies a reorganization of the music market, creating a much closer relationship between artists and listeners, as well as new consumption possibilities.

All this evolution in the music market has generated a special interest in the interaction between end-users and music streaming services, in order to understand how the end-users search, select, classify, and evaluate new songs indicating if they like or dislike them [30]. Moreover, there is currently a special interest in improving the interaction processes to generate effective recommendations. One of the main challenges in music recommender systems (MRS) is the cold start, in which the system does not know what to recommend to new users because they have not rated anything yet [37]. In this way, MRS appear as a field of research that studies

✉ Yesid Ospitia-Medina
yesid.ospitiam@info.unlp.edu.ar

José Ramón Beltrán
jrbelbla@unizar.es

Sandra Baldassarri
sandra@unizar.es

[1] University of La Plata, La Plata, Argentina

[2] Universidad Icesi, Cali, Colombia

[3] University of Zaragoza, Zaragoza, Spain

Published online: 05 June 2023

and proposes various strategies to facilitate and improve the listener's experience.

The understanding of the listener's emotional perception is one of the fields of study most explored by MRS since music has a close relationship with emotions [15, 41]. Music information retrieval (MIR) researchers and cognitive psychologists consensually understand that certain features of music trigger several transformation processes, one of them, related to emotions [2]. Taking into account this background, much of the research has focused on the recognition of emotions in music (MER), in which many models based on artificial intelligence are currently designed to recognize emotions in musical pieces and then represent these emotions through affective models. These models can be dimensional or categorical. In dimensional models the different emotions are described by a value in valence and arousal. Instead, in categorical models, emotions are described by the specific adjectives that define them [51].

Despite the significant development of MRS today, many artists, especially the non-superstar artists, are dissatisfied with the actual chances of their songs being recommended in a way that can boost their commercial artistic career [4, 23]. The non-superstar artists are musicians who compose original songs, and who have very limited popularity in the music market because in many cases they are known only locally. Similarly, sometimes listeners also have a perception of dissatisfaction with the MRS [41], mainly because the recommended songs do not appeal to them, or because the MRS usually recommends the same set of songs, preventing the listeners from discovering new musical content. This artificial bias applied to data and users' preferences revealed by some recommender systems constitutes a research topic of great interest to the current MRS research community [36].

The objective of this work is to analyze the problem of biases in the MRS field and to propose some strategies to mitigate them. The main contributions are:

- A literature review to identify the main attributes of the most commonly used recommendation strategies, highlighting important findings and limitations.
- A review of some of the most commonly used datasets in the fields of MIR and MER with an understanding of their limitations from a biases analysis perspective.
- The design of a new song dataset with novel features such as: complete musical structure, artists who are not superstars, labeling as like/dislike, musical genre, and emotion labeling evoked by each artist about their own songs.
- A strategy based on emotion recognition that takes into account the time variable and the song structure, which is tested over the new dataset.

This paper is organized as follows: initially, Section 2 offers an overview of MRS, describing some of the most important recommendation strategies, a theoretical background for biases, as well as an analysis of biases in some selected datasets. Based on the drawbacks found in the analysis of these datasets, Section 3 presents the development of a new dataset, detailing the design process and its specifications. Section 4 exposes the proposed strategy for song recommendation through an emotional analysis based on time-series. Section 5 includes some experiments regarding similarity metrics and clustering strategies over the new dataset. Finally, Section 6 highlights the conclusions obtained, followed by Section 7, which presents the limitations and the proposed future work.

2 Music recommender systems

The main objective of MRS is to suggest new songs to the user, and their effectiveness is measured according to the acceptance degree (likes/dislikes) expressed by the user with respect to the recommended song. In order to achieve this, a variety of strategies are usually implemented. Following, Section 2.1 describes the most well-known and currently used recommendation strategies, Section 2.2 presents a theoretical background for biases, while Section 2.3 presents an analysis of the limitations of some datasets, considering how these limitations can generate biases in MRS.

2.1 Recommendation strategies

Some of the most recent work has focused on the development of recommendation strategies based on content-based filtering (CBF), emotion-based filtering (EBF), personalized approach (PA) and user context (UC) [17, 37, 44, 54]. Many of these strategies subsequently allow the generation of metadata associated with the domestic consumption of music, which is a key aspect for recommendations, especially those related to personalized music [9]. The strategies are briefly explained below.

CBF recommendation strategies extract and analyze the internal characteristics of the songs, and then apply different criteria to perform a classification process. For example, after automatically extracting the sound characteristics of a song, and from previously trained models, it is possible to determine the musical genre or the evoked emotions, among others [33, 37, 44]. EBF techniques are based entirely on the recognition of emotions in music and often involves emotional labeling processes to classify music by emotions and to make recommendations based on the emotion the listener wants to perceive [10, 25, 37]. PA focuses on suggesting very specific music considering the particular preferences of each user avoiding the different biases that can generate a recommender system based on a general community and the popularity of the artists [37, 54]. Finally, UC analyzes all the users' contextual variables, such as their current activity, the day of the week, weather, among others, and consequently determines the recommendations [25, 28].

Moreover, the literature review allows to detect some important findings that should be considered in order to develop music recommender systems:

- There is a tendency to design hybrid strategies for recommendation processes, in which personalization and filtering based on content and emotions have recently become predominant. According to [37], creating a personalized user experience is the best way to recommend songs because it reduces human effort in searching, classifying, and discovering new music.
- It is a fundamental fact that emotions have a close relationship with musical features [42], this argument is considered from the listener's and composer's perspective, as well as from several psychological studies [26]. Music is considered emotionally significant, and mechanisms such as brain stem reflex, emotional contagion, episodic memory, and musical expectancy explain in some way the emotional reactions that music generates in listeners [27]. These facts motivate to deepen emotion-based filtering to improve the recommendation strategies.
- Based on interviews with artists, they suggest that the musical structure of a song is comparable to a journey in which the intensity of emotions may fluctuate, establishing a relationship of influence between the musical structure and the perceived emotional expressions [20, 46]. In [3] the time-course of emotional responses during listening sessions is studied finding that listeners require 8.31 s of music before initiating emotional judgments. Similarly, [50] highlights that emotion data averaged across listeners and over time to infer the emotion expressed by a piece of music is considered a limitation that may restrict the effectiveness of MER systems. Thus, summarizing a song to a particular emotion may be convenient from a technical simplicity perspective, but it is not enough from an artistic perspective. This discussion suggests going deeper into the structure of the song, analyzing the variability of emotional intensity in each part of the song structure.
- There are a very limited number of works dedicated to analyze the case of non-superstar artists, therefore these artists are severely affected by the bias imposed by the majority of recommender systems.

2.2 Biases in MRS

In its most general use, the term bias could be interpreted like a slant to a specific issue. In this way, the term bias also involves, in most cases, a discussion related to fair treatment, because it can benefit some stakeholders, as well as negatively impact others in a specific context. This unequal treatment promotes damaging conditions that even currently motivate moral discussions [19]. According to Friedman and Nissenbaum, biases in computer science can be classified into 3 categories [19]: preexisting, technical, and emergent.

Preexisting biases are generated by social institutions, practices, and attitudes. This kind of bias is promoted by society, it has a direct relationship with culture, and it can be exercised explicitly or implicitly way by customers, system designers, and other stakeholders. Technical biases have their roots in technical constraints or technical considerations. This kind of bias arises from technical limitations, which may be present in hardware, software, and peripherals. In the case of software, it is very important to analyze and deal with decontextualized algorithms, which promote unfair data processing. And finally, emergent biases can only be detected in a real context of use. This kind of bias appears sometimes after a design phase is completed, as a result of changing societal knowledge, population, or cultural values.

One of the most important causes of biases in MRS is related to datasets due to their design process (which may include preexisting biases), and the recommendation algorithms used over them (which may include technical biases) [1]. In view of this fact, the next section presents a review of 14 datasets available for conducting research in MIR with the main purpose of understanding in-depth its limitations.

2.3 MRS dataset analysis

The criteria considered for the comparison between datasets correspond to the most common ones according to the literature reviewed, in which the main focus regarding the annotations is set on the listeners. The following is a detailed explanation of the criteria analyzed for each dataset in Table 1:

- **Audio:** If the audio files are available or not.
- **Musical structure:** If the dataset includes or does not include metadata related to the complete identification of the musical structure of a song, typically composed by introduction, verse, chorus, and solo. The musical structure allows performing experiments based on the similarity of the parts of the structure, which is very useful considering that a song is an emotional experience that changes over time.

Table 1 Datasets review *C: Categorical, D: Dimensional*

Dataset	Year	Audio	Musical structure	Affective model	Emotional labeling by listener
GTZAN [49]	2002	✓	X	–	X
Ballroom [22]	2006	✓	X	–	X
MagnaTagATune [29]	2009	✓	X	–	X
Last.fm [8]	2010	X	X	–	X
Million Song Dataset [5]	2011	X	X	C	✓
UrbanSound8k [40]	2014	✓	X	–	X
ESC-50 [38]	2015	✓	X	C	X
TUT Acoustic Scene [31]	2016	✓	X	–	X
Mediaeval [47]	2016	✓	X	C & D	✓
AudioSet [21]	2017	X	X	C	X
PMemo [53]	2018	✓	Chorus	C & D	✓
Spotify Million Playlist [52]	2018	X	X	D	✓
Jamendo [7]	2019	✓	X	C	✓
WCMED-CCMED [16]	2020	✓	X	D	✓

- **Affective model:** The affective model used, which can be categorical (C) or dimensional (D).
- **Emotional labeling by listener:** If the listener has emotionally labeled the songs or not.

A very important finding to highlight is that in 7 of the 14 reviewed Datasets, GTZAN [49], Ballroom [22], MagnaTagATune [29], AudioSet [21], TUT Acoustic Scene [31], UrbanSound8k [40], ESC-50 [38], the duration of the sound files varies between 1 and 30 s, limiting the possibilities for experimentation. In most cases, audio files with real songs are not available, instead the audio files correspond to ambient sounds or perhaps small sound fragments with a little bit of musical content. Last.fm datasets [8] include data related to listeners' habits, such as user, artist, song, playing time and total plays. In the case of the Million Song Dataset [5], the average length of the files is not very clear, and although it is a dataset of songs, these correspond to covers of famous songs and not original songs of non-superstar artists. In addition, the audio files are not available. The Mediaeval dataset [47] is very complete due to a large number of emotional annotations. However, there are no annotations for the different parts of the song structure (introduction, verse, chorus), and there is no data referring to a deeper analysis of the artist's perspective. The emotional recognition is performed over time by giving valence and arousal coordinates in a dimensional affective model every 500 milliseconds in [34]. Then, these coordinates are used in [35] to achieve an emotional classification in 4 quadrants, finding that the dataset is unbalanced with respect to the distribution of songs by those classes. This fact represents a problem when implementing machine learning algorithms because this kind of algorithms tend to recognize with more accuracy the majority class, whereas they give a low accuracy rate with the minority classes, in this particular case, there are many more songs classified with positive emotions such as happy, and excited, rather than negative emotions such as alarmed and angry. PMemo dataset includes dynamic and static emotional annotations on the chorus of real songs [53]. It is important to note that these songs include artists with an intermediate or even high level of fame in the music industry. Spotify Million Playlist dataset [52] is a large repository of playlists, with the possibility of accessing different labels on songs through the Spotify web API, such as energy and valence. On the other hand, even though Jamendo dataset includes full tracks, these songs are not labeled by their musical structure, neither the *like* tag is included and the level of fame of artist is not detailed [7]. In general, there is no annotation of any kind by artists/composers that would allow a deeper analysis from a musical artist perspective.

Popularity is one of the most important causes of preexisting and technical biases in MRS, which not only affects listeners but also has a crucial impact on artists, especially in the case of non-superstars. MER and MRS might benefit from being focused on a content preference, either in perception or audio parameter, over popularity in order to better suggest from the complete range of available music. Taking into account the above considerations for the case of musical recommender systems, the following particular limitations are identified in the existing datasets:

- In most cases, they include very short audio clips which are not exactly songs, and consequently do not have a complete musical structure.
- There is no in-depth analysis from the point of view of the field of music that involves the composer, which would be

really important to design new and better musical descriptors.

- The analyses are not focused on original songs by non-superstar artists, which generates a popularity bias and is, nowadays, one of the main objections of non-superstar artists regarding the performance of many music streaming services [4].
- There is no information available about the degree of balance of the data for the different classes defined through labeling processes.
- There is no *like* or *dislike* information tags available about the songs in the datasets.

3 ENSA: dataset

Considering the limitations presented above, a new dataset[1] of 60 musical pieces has been designed to implement experiments that take into account the musical structure of the songs. The following sections describe the preparation of the dataset (Section 3.1), the contents (Section 3.2), and then an analysis of matches in the labeling processes with respect to emotions, musical genres, and likes/dislikes (Section 3.3).

3.1 Dataset preparation

The following steps were carried out for the preparation of the dataset:

- Invitation for artists: To collect the original songs, an invitation was shared via email and some social networks asking non-superstar artists from the Valle del Cauca in Colombia if they would like to be part of this research. The artists interested in participating provided their original songs under a Non Commercial Creative Commons License which allows free access to the metadata and audio of the songs on a non-profit basis. In addition, before starting the interviews, data collection, and the different experiments, the artists were informed about the purpose of the research and asked for their consent.

- Interviews with artists: Individual interviews were conducted with the artist in order to understand their compositional strategies and the musical structure that they defined in their songs. The information obtained from these interviews was very important because it allowed us to understand that not all artists use the same composition methods, some of them apply very organized methodologies while others make use of improvisation, but in all cases, the starting point is the personal inspiration which has a direct connection with the emotions that the artists want to evoke with their compositions.

Regarding song structure, artists emphasize that with it they can create the emotional experience they want to communicate [32]. Many of them use a well-known structure that involves an introduction, bridge, verse, bridge, chorus, verse, solo, chorus. This highlights the importance of understanding the structure of a song and analyzing the emotional changes between parts, and for this reason, the artists were asked to indicate the beginning and end of the verse and chorus of their song, as well as their emotional intention for the verse, the chorus and an approximation for the whole song. The affective model proposed for this experiment was previously explained to the artists in order to avoid misunderstandings related to the meaning of each emotion.

The artists used 8 adjectives to label their emotional intention (excited, happy, alarmed, angry, depressed, boring, calm, satisfied) adopted from James A. Russell's circumplex model of affect [39], which presents a dimensional model (2 dimensions, valence and arousal) that divides the space into 4 quadrants (Q1, Q2, Q3, and Q4), each one respectively with 2 adjectives. The selected emotions are usually the basic ones and the most used in emotional labeling processes [43]. In addition, labeling processes that include a moderate amount of emotions decrease the difficulty that a user may experience in differentiating emotions correctly. The model adopted for the experiment is shown in Fig. 1.

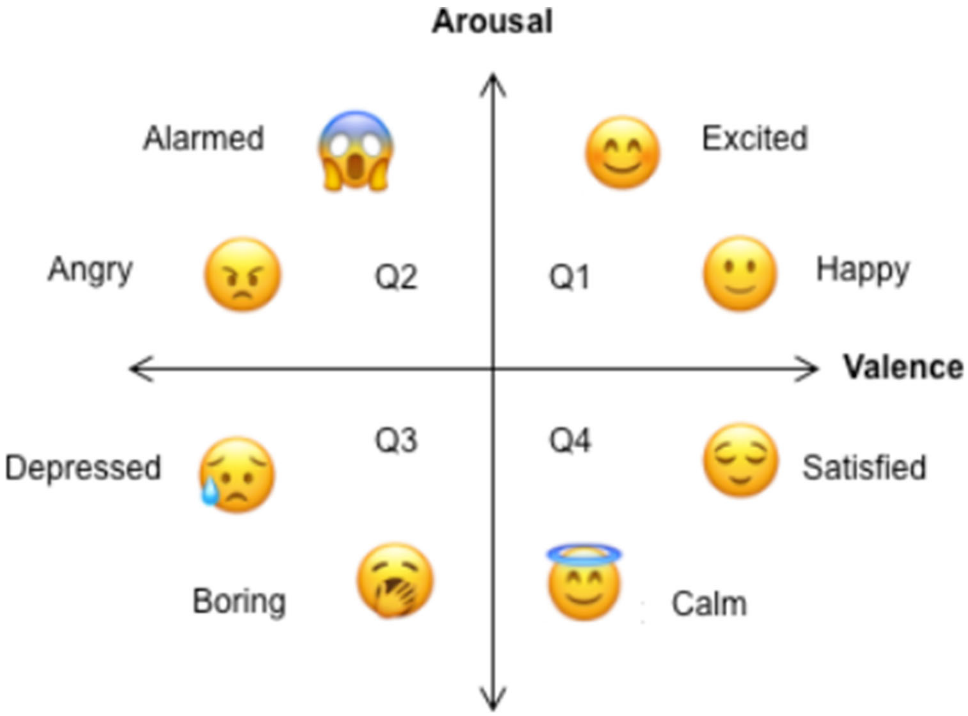
In the process of labeling songs by musical genre, artists in many cases assigned very specific genres that could be considered sub-genres. In order to generate a more general grouping of songs by musical genre, some of these genres were indicated in a more general way, as in the case of Metal, which includes subgenres such as Heavy Metal and Death Metal, allowing to reduce the number of musical genres from 21 to 16. Besides, sometimes the artist defined a mix of genres for some songs, such as Funk Blues, Blues Folk, Country Blues. For these cases, each song was classified with the predominant musical genre considering the one with the greatest influence on the different musical elements of each song.

-Questionnaire for listeners: With the purpose of validating if the emotion that the artists intend to evoke with their song is the same as the one perceived by the listeners, a web questionnaire[2] was used through which the listeners could listen a song and indicate what their emotional perception for the whole song is (with the same adjectives used by the artists). Additionally, the listeners can specify if they like or dislike the song, can select the musical genre that they consider appropriate for the song, and can as well write through a complementary field any observation or detail that they consider important to explain regarding their experience. There are also other important demographic data requested to characterize the user, such as gender, age range, and level

[1] Available at: https://github.com/yesidospitiamedina/ENSA.

[2] Available at: http://104.237.5.250/evaluacionensa/form.php.

Fig. 1 Affective model



of music knowledge (music listener, amateur practitioner, intermediate experience, advanced experience). In the same way, as with the artists, the listeners were informed about the purpose of the research and asked for their consent.

3.2 Dataset content

The dataset consists of a total of 234 min of audio, equivalent to 3.91 h or 468 30-second audio clips. From the musical point of view, the dataset includes 60 complete songs, which means that the original work of the artist is included without any cuts. The specifications and each relevant attribute of the dataset are detailed in Table 2. It is important to note that not all the songs follow a musical structure with verse and chorus; only 39 songs have this characteristic. Also, the songs with female, male and no vocals are marked.

Figure 2 presents the distribution of the songs according to the artists' emotional perceptions of their own musical works. As mentioned above, some songs define in their structure the verse and the chorus, 39 songs follow this structure and for this reason, they have labeled the emotional perception in verse, chorus, and the complete song. For the other 21 songs, the artists indicated that they do not follow any particular structure, and for this reason, they have only been labeled as a complete song. Table 3 shows the distribution of songs by musical genre, as well as specifying the type of voice, female voice, male voice, children's voice, or no voice.

In addition to the audio files, we also included metadata files containing emotional and musical genre labeling by artists and listeners, verse and chorus marking, low-level features extracted for each whole song every 500 ms through the *OpenSmile*,[3] and complementary information provided through the listener questionnaire.[4]

3.3 Dataset analysis

An analysis of the ENSA-Dataset was carried out to check the degree of coincidence between the emotions and the musical genres indicated by the artists and the listeners. With respect to emotions, the analysis was performed at the global level of the whole song, trying to identify the quadrants and semi-planes in which there are coincidences. In this comparison, it was particularly useful to also consider the emotional perception according to the bias that can generate the *Like* or *Dislike* evaluation indicated by the listener. Regarding the musical genre, the degree of coincidence between the genre indicated by the artist and the listener was analyzed taking into account that some genres can be similar and for the most common listener it is difficult to differentiate them. For this study, 106 annotations were collected through the questionnaire. The questionnaire was designed to randomly choose the song to be evaluated each time the user accesses it, avoiding the repeated selection of the song in a particular user session. In this experiment 46 different songs received between 1 and 4 evaluations by different listeners.

Table 4 presents the perfect matches of emotions labeled by the artists and the listeners from two perspectives, the

[3] http://opensmile.sourceforge.net/

[4] https://github.com/yesidospitiamedina/ENSA.

Table 2 Dataset specifications

Attribute	Detail
Audio encoder	MPEG layer 3 (MP3)
Total time	234 min
Clips of 30-second	468
Number of songs	60 songs
Number of artists	10 Non-superstar
Low-level features	260 extracted every 500 ms
Identified verses	39 verses
Identified choruses	39 choruses
Undefined musical structure	21 songs
Specified musical genres	21 genres
Grouped musical genres	16 genres
Songs with female voice	15 songs
Songs with male voice	39 songs
Songs without voice (instrumental)	4 songs
Songs with child voice	2 songs
Affective model	Categorical (8 emotions)
Emotional labeling by artist	39 verses
	39 choruses
	60 whole song
Emotional labeling by listeners	106 annotations
Like / *dislike* marking per listener	106 marking

first is the case in which the listener likes the song, and the second is the case in which the listener does not like the song. One of the most important findings that can be evidenced is that more emotional labeling matches are generated when the users evaluate songs that they like. These songs lie within Q1 and Q4.

Table 5 presents the matches by emotions, quadrant, and V/A dimensions according to the musical genres labeled by the artist in the songs that have been evaluated in this experiment. The main idea of this analysis is to highlight that for some musical genres, such as Blues, Folk, and Metal, there was a greater coincidence between the emotional evaluation process of artists and listeners (cases of likes). This suggests that some musical genres may be more closely related to certain emotional stimuli which are more clearly perceived. With respect to the matches between the music genres labeled by artists and listeners, Table 6 shows that the Metal and Blues genres are the ones with the highest number of matches. Similar to the previous analyses focused on emotional perception, the highest number of matches was also found with the songs that the user likes.

The listener's profile was also analyzed to find out if there is any relationship with the number of matches between the musical genres labeled by artists and the ones labeled by listeners. For this purpose, the results presented in Table 7 were analyzed, which show that a higher level of musical knowledge in the listener's profile increases the chances of correctly labeling the musical genre.

4 Time series approach for song recommendation

Although there are several ways to classify music, such as the musical genre, if it is instrumental, the degree of danceability, if it is performed live, if it is appropriate for a specific activity (sport, sleep, relaxation), most of these criteria are usually related to moods, or to some interest in evoking emotions [45]. Furthermore, it is important to mention that, despite the

Fig. 2 Song distribution by quadrant, emotion, and musical structure according to the artist

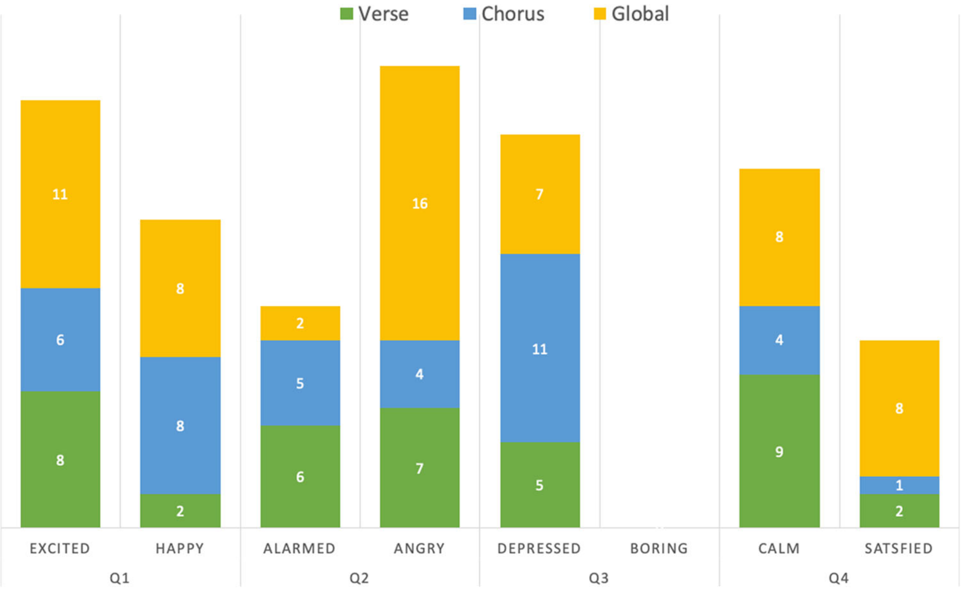


Table 3 Song distribution by musical genre

Musical genre	Female	Male	Child	No Voice
Ballad	–	4	–	–
Ballad Rock	–	2	–	–
Blues	–	4	–	–
Bossa nova	1	–	–	–
Classical music	–	–	–	1
Currulao	1	–	–	–
Country	–	1	–	–
Folk	11	–	–	–
Funk	–	2	–	–
Jazz	1	–	–	–
Latin music	–	–	1	–
Metal	–	21	–	–
Pop	–	–	–	2
Pop Rock	1	–	–	–
Rock	–	5	–	1
Swing	–	–	1	–

existence of many research papers on emotion recognition in music, most of them focus on categorical models (to classify emotionally), and a smaller amount on dimensional models [39]. In the case of classification, adjectives are used for specific emotions; regarding dimensional models, a coordinate in a two-dimensional plane is used to establish the emotional valence (positive or negative), and the arousal (energy level).

It is also important to note that, from an engineering sound perspective, songs are analyzed as digital signals, and from this approach, some works such as [4, 14, 17, 37], and [6] refer in some way content-based filtering strategies based on low-level features of sound. This suggests that the artistic perspective of music may have some bias in current research because although music may be treated as sound, artists see it as an art that is based on their emotions and described through music theory. One way to transmit the artist's emotion is through the song structure, considering that listening to music is an emotional experience that happens over time [32]. Artists create emotional experiences that can fluctuate in valence and arousal. In this way, the listener can experience a changing energy when the song moves from the verse to the bridge, and then to the chorus.

Table 4 Matching analysis based on specific emotions

Quadrant	Emotion	Matches with Like	Matches with Dislike
Q1	Excited	5	1
	Happy	7	0
Q2	Alarmed	0	0
	Angry	2	1
Q3	Depressed	1	1
	Boring	0	0
Q4	Calm	7	0
	Satisfied	1	0

Table 5 Analysis of matches between musical genre (specified by the artist), emotions, quadrants and V/A dimensions. Only cases of likes

Musical genre	Emotion	Quadrant	Valence	Arousal
Ballad	–	–	2	2
Ballad Rock	–	1	1	2
Blues	5	7	7	9
Bossa nova	–	1	2	1
Classical music	-	–	3	–
Currulao	1	1	3	1
Country	–	–	–	–
Folk	6	12	17	15
Funk	2	2	3	2
Jazz	–	–	–	–
Latin music	–	–	–	–
Metal	4	12	14	15
Pop	1	1	6	1
Pop Rock	–	–	–	1
Rock	2	4	6	5
Swing	2	2	2	3
Total	23	43	66	57

Table 6 Analysis of coincidences by musical genre

Music genre	Matches with likes	Matches with dislikes
Music genre	**Matches with likes**	**Matches with dislikes**
Ballad	1	–
Ballad Rock	2	1
Blues	4	1
Bossa nova	–	–
Classical music	2	1
Currulao	3	–
Country	–	–
Folk	1	–
Funk	–	–
Jazz	–	–
Latin music	–	–
Metal	14	5
Pop	–	–
Pop Rock	–	–
Rock	2	–
Swing	1	–
Total	30	8

Table 7 Analysis of matches for musical genre (Likes & Dislikes) by listener profile

Profile	Matches	Total evaluations	% Matches
Listener	27	77	35%
Amateur	6	18	33%
Intermediate	3	7	43%
Advanced	2	4	50%
Total	38	106	36%

Considering the relevance of representing the emotional perception over time in music, it is proposed to use the time-series approach, in which a time-series will be designed to represent the emotional perception of each important part of the structure of a song. In this way, the artist provides as inputs the audio recording, as well as annotations of the musical structure with time stamps, and then obtain as outputs two arrays: the valence array, and the arousal array. This process and its main phases can be seen in Fig. 3 and are explained below:

1. **Song structure analysis:** The structure of the song is very important because the composer uses the structure to create the musical experience through which the listener perceives emotions with different levels of intensity. Thus, one of the most well-known structures involves an introduction, bridge, verse, bridge, chorus, verse, solo, chorus. Although in general one emotion usually predominates (especially in modern musical productions), this emotion could present a different intensity for each part of the structure and there could appear exceptional cases in which even the valence is different. In this initial phase of our system, with the assistance of each artist, a CSV file is constructed in which the different parts of the structure of their original songs are defined and used as Metadata, indicating the moment in which each one begins and ends in the timeline. This kind of information is not usually available in most of the existing musical datasets; in contrast, this information is available in ENSA.
2. **Signal processing analysis:** In this phase a sound feature extraction tool is applied. For this work, *OpenSMILE*[5] has been used to extract 260 low-level features every 500 ms. The extraction intervals are determined by the phase that defines the structure of the song, so depending on the part to be analyzed, the extraction time must be adjusted. This phase generates as output a two-dimensional array of M x 260, where M will be the number of time-windows determined by the total duration of the part of the song structure of interest to be analyzed.
3. **Emotion recognition:** For emotion recognition, two predictive neural networks[6] previously trained with low-level features in [34] are used, one for valence recognition and the other for arousal recognition. The output of this phase includes two arrays, in which each one presents values between -1 and 1, and respectively indicates the valence and arousal level.
4. **Time series representation:** The variability of both valence and arousal over time is plotted on a two-dimensional plane, allowing an emotional view of the part of the structure of a particular song to be analyzed.

The final arrays obtained through the previously described process can be interpreted as emotional descriptors of the song oriented to its different parts of musical structure. The interest in representing these descriptors as time-series is mainly due to two reasons. On the one hand, the time-series can be analyzed through similarity techniques which for practical purposes are very useful to determine the closeness level of a valence or arousal array of a particular song compare to other arrays of different songs. On the other hand, the same similarity metrics can be used to implement clustering techniques, allowing to find groups of songs that are close in valence and arousal. The definition of these groups allows to implement a recommendation strategy based on the closeness of the songs located within the same group (cluster).

5 Experiments

In this section the analysis and implementation of similarity metrics are presented in detail (Section 5.1), followed by the clustering strategies and experimental results on the ENSA musical dataset (Section 5.2).

5.1 Similarity metrics

In general, there are two popular metrics for analyzing similarity in time-series: the Euclidean matching distance and the dynamic time warping matching (DTW) distance. It is important to highlight that the use of these similarity metrics must be applied between parts of the song structure in order to make this comparison meaningful.

1. **Euclidean distance:** The comparison between time-series is made point-to-point following the order in which these points are presented sequentially over time. Additionally, it has an important restriction regarding the fact that both time-series must have the same size (number of measurements), which from the perspective of comparing verses, or choruses, is not convenient because the

[5] http://opensmile.sourceforge.net/

[6] Available at: https://github.com/yesidospitiamedina/ENSA.

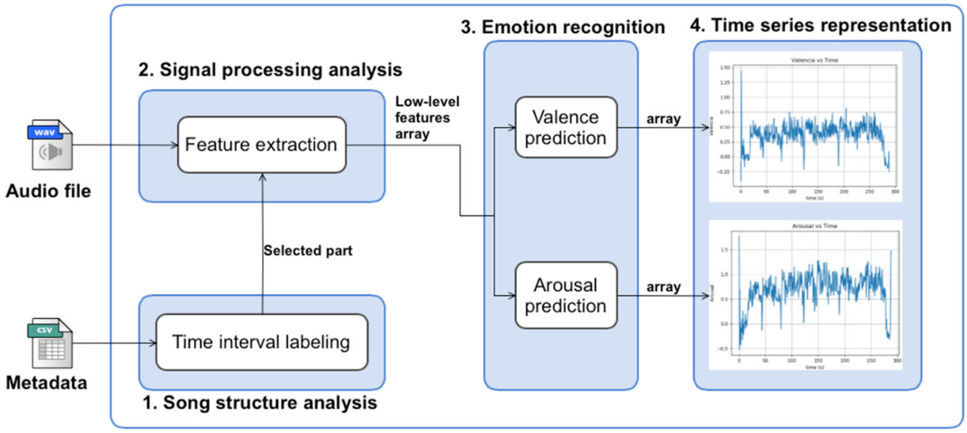


Fig. 3 Time-series approach song analysis process

durations are usually different between songs, and the variations in emotional perception will hardly be presented perfectly synchronized over time.

2. **DTW:** Its algorithm initially calculates the best alignment path between two time-series, this is called elasticity capacity, this allows aligning points out of phase over time, which also allows comparing time-series of different sizes.

The graphical comparison of both metrics can be seen in Fig. 4. Taking into account the above arguments, DTW is a more convenient metric because it allows comparing particular parts of musical structure (such as a verse) between different songs considering also the fact that they could have different lengths.

In this work the library *TSLEARN*[7] is used for implementing similarity algorithms in time-series [48]. *TSLEARN* runs on *Python* language and allows implementing the Euclidean distance, DTW, and softDTW, the last one is a variation of DTW that includes a hyperparameter (value between 0 and 1) through which it is possible to soften the process of determining the best alignment between two time series [12]. This approach is suitable for those special cases in which the application of DTW generates unsatisfactory jumps (according to the study context) to associate the points of two time-series and determine their optimal alignment.

To analyze the practical contribution of similarity measures, an experiment has been designed to recommend songs by non-superstar artists that are emotionally close to a well-known song by a superstar artist, considering that emotionally close is related to the degree of similarity between two songs, according to two criteria: the part of the song (verse or chorus), and the emotional perception of that part of the song (previously recognized by a neural network trained with the MediaEval dataset). In this experiment the song *Hysteria* by the band *Def Leppard*[8] is taken as a reference and from this song the original songs of non-superstar artists are ordered to determine which of these songs are the closest. Taking into account the interviews with the artists, this song was selected because it inspired the composition of the song *Hasta hoy*, which is part of the ENSA dataset. Tables 8 and 9 show the similarity distances (for the verses) based on the DTW algorithm for the commercial song, and its 10 most similar songs of non-superstar artists from a total set of 39 (those that include the verse in their structure). Note that a smaller number means that it is closer, so it can be observed that the song *Hysteria* has a distance of 0 with respect to itself, with respect to valence the closest non-famous song is *Pensarás en mí* and with respect to arousal, the closest is *El swing de Caperucita*. If the first priority is set for valence, it might be a good choice for this case to suggest the songs *Pensarás en mí*, *Esperando por ti*, and *Doncella de virgo*, besides the fact that they are included in the subset of the closest songs by arousal (Table 9).

5.2 Clustering strategies

Clustering strategies belong to the field of machine learning and are particularly characterized by the implementation of non-supervised training, in which the data have not been previously labeled. Clustering algorithms have the objective of separating elements into different groups that have certain features in common, and for this purpose, the intra-cluster distance is minimized and the inter-cluster distance is maximized. The *K-means* algorithm is one of the most widely used and simplest to implement. The algorithm initializes with K randomly generated centroids and iteratively calculates the membership of the patterns to each group based on the distance to the centroid. Then, the centroid is adjusted by moving it to the median point.

[7] https://tslearn.readthedocs.io/en/stable/

[8] https://www.defleppard.com.

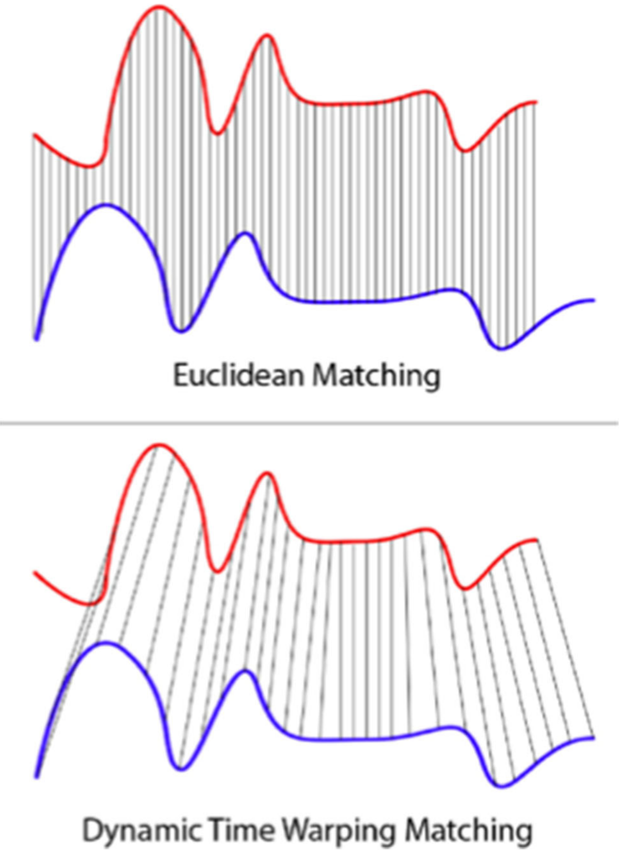


Fig. 4 Comparison of Euclidean Distance Measurements and DTW (extracted from [11])

Typically, clustering algorithms are used with datasets that relate two features for each element of the set, so it is possible to graphically observe the convergence of these algorithms in a two-dimensional plane. This *K-means* condition generates a challenge in its traditional implementation on the emotional perception over time in music, in which for each element (song) there are many measurements. However, there is a variant of this algorithm developed specifically for time-series implementations and is available in the *TSLEARN* library.

Table 8 The 10 non-commercial songs closest to *Hysteria* according to DTW applied by verse and valence

Song	Hysteria (DTW for Valence)
Hysteria	0
Pensarás en mí	0.61
Esperando por ti	0.69
Doncella de virgo	0.74
Bienvenidos	0.87
Intento de florecer	0.88
Currulao de nieves	0.95
Bajo mi piel acústico	1
Peste de silicio	1.10
Estático ser	1.10
El swing de Caperucita	1.10

Table 9 The 10 non-commercial songs closest to *Hysteria* according to DTW applied by verse and arousal

Song	Hysteria (DTW for Arousal)
Hysteria	0
El swing de Caperucita	1.0
A construir	1.20
Bienvenidos	1.30
Esperando por ti	1.40
Doncella de virgo	1.40
Estático ser	1.50
Currulao de nieves	1.50
Bajo mi piel	1.60
Eterno abrazo	1.60
Pensarás en mí	1.60

With the purpose of showing the contribution of implementing the library, clustering results are shown below, following the same experiment of the previous section. Again the focus of analysis will be the famous song *Hysteria*, the *TimeSeriesKMeans* algorithm will be run and the cluster in which the song is located in both valence and arousal will be identified and analyzed. For the case of clustering by valence (Fig. 5), the song *Hysteria* was located in the cluster 1 together with the song *Doncella de virgo*, which is consistent with the results of Table 8, in which both songs appear. With respect to clustering by arousal (Fig. 6), the song *Hysteria* was located in the cluster 5 together with the songs *Currulao de nieves*, *El swing de Caperucita*, and *A construir*. This result is consistent with the results of Table 9 which also includes these songs.

The training for the clustering processes by valence and arousal was parameterized with $K=7$ and, as can be seen in the Figs. 5 and 6, the level of clustering by valence was more adjusted (fewer songs) with respect to arousal. The value of K was increased one by one looking for a value that would best fit the cluster in which the song *Hysteria* would be placed, preferably with the least possible number of songs (the closest ones). It is important to emphasize that one of the major challenges in this experiment is to find the value of K, which requires many executions of the same experiment and a manual review of the results in order to determine the most convenient value according to the interests of the study. Moreover, although with TSLEARN is possible to generate the different clusters of time-series, it is not possible to generate a two-dimensional graph to visualize the dispersion of the songs in an easier way.

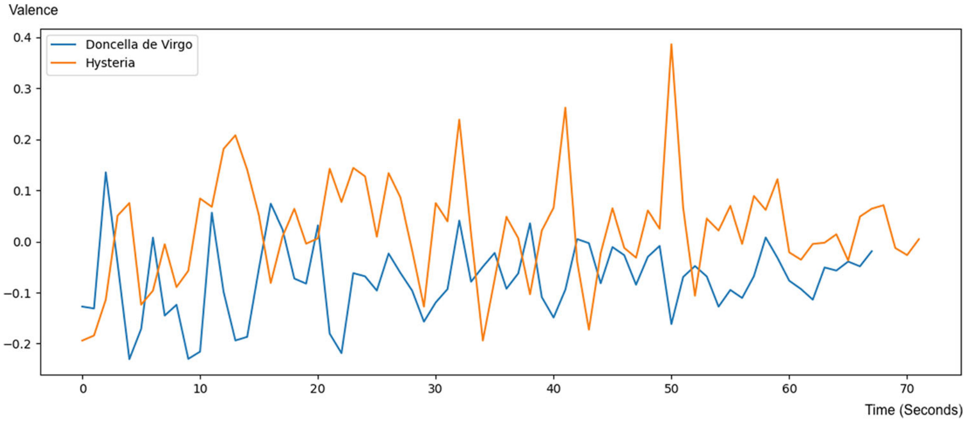


Fig. 5 Valence value of the time series for the songs in cluster 1

Considering the previous limitations of *TSLEARN*, a new clustering experiment is designed but this time using two new libraries: *UMAP*[9] and *HDBSCAN*.[10] The first one allows performing a dimension reduction of the dataset, and the second one finds the optimal number of clusters through unsupervised learning techniques. Figures 7 and 8 show the clustering of songs by valence and arousal respectively. The values that can be seen in both the x-axis and y-axis correspond to the two components resulting after applying the dimension reduction with the *UMAP* library. For a more detailed analysis, it would be useful to review the similarity value tables presented in the previous section in order to establish a relationship between the values of the similarity metrics calculated and the intra-cluster and inter-cluster distances. The entire database (60 songs) was considered to generate the two figures, and the song *Hysteria* was also included. The clustering has been carried out by verse, and for this, the songs that do not follow a structure that defines a verse were analyzed by sampling a representative part of the song in order to include it in the experiment. As can be seen in Figs. 7 and 8, the *HDBSCAN* library determined 4 clusters for both clustering cases.

One more time our focus will be on analyzing the cluster in which the *Hysteria* song has been located in both the valence and arousal clustering, and for this purpose, a comparison between the different approaches to the previously described experiments is presented in Table 10. Regarding the case of valence (Fig. 7) the song *Hysteria* has been located in cluster 4 together with 18 songs of non-superstar artists. From these songs our interest is to highlight the following 7 songs: *Esperando por ti*, *Bajo mi piel acústico*, *Doncella de virgo*, *Intento de florecer*, *Pensarás en mí*, *Currulao de nieves* and *Bienvenidos*. These 7 songs are also present in Table 8, as well as the song *Doncella de virgo* is part of the cluster 1 (Fig. 5) calculated with *TSLEARN* for valence. Concerning the clustering results for arousal (Fig. 8), the song *Hysteria* has been located in cluster 1 together with 10 songs of non-superstar artists. From these songs our interest is to highlight the following 5 songs: *Esperando por ti*, *Doncella de virgo*, *Estático ser*, *Pensarás en mí* and *Currulao de nieves*. These 5 songs are also present in Table 9, as well as the song *Currulao de nieves* is part of the cluster 5 (Fig. 6) calculated with *TSLEARN* for arousal.

This analysis shows an important coincidence of songs between the different results obtained from the similarity value tables, clustering by *TSLEARN*, and clustering with *UMAP* and *HDBSCAN*.

6 Conclusions

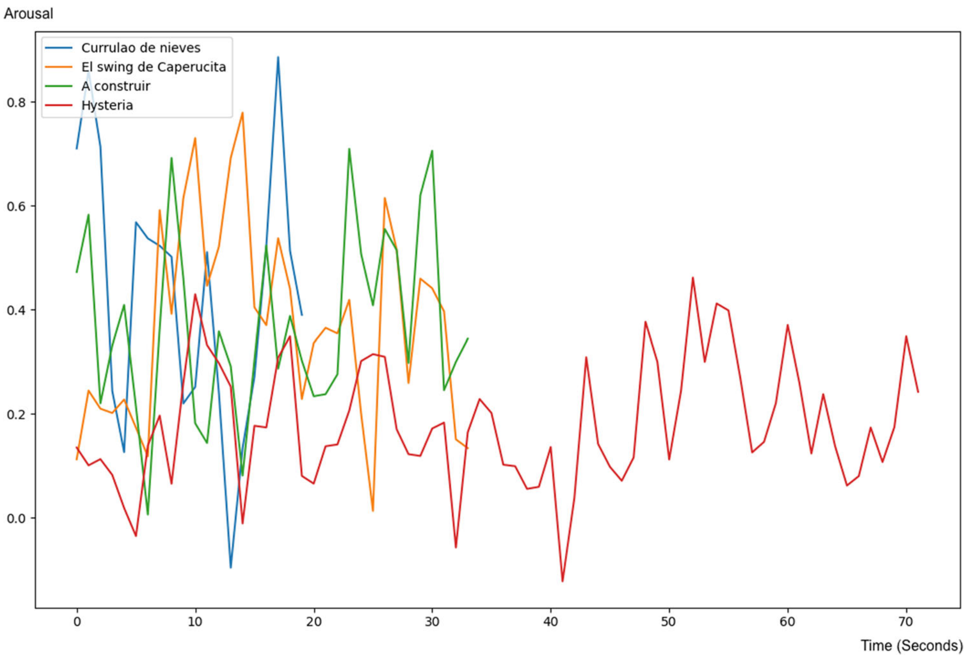
This paper has focused on the design of a new dataset of songs considering the limitations found in some of the most commonly used datasets in the fields of MIR, MER, and MRS. It is important to emphasize that many of these limitations are related to preexisting biases, in particular, biases related to the popularity effect that is often generated because non-superstar artists do not participate. Moreover, these limitations also include the lack of many labels such as musical genre and the emotional perception of each part of the song structure. In view of these findings, the ENSA dataset was designed with the participation of non-superstar artists as a key factor to obtain important information such as their emotional perception, their musical genre classification, and the identification of the musical structure for each song. In the same way, novel information was collected from the listeners, such as the listener's expertise in music and the labeling of *likes*.

Based on the information collected through the labeling processes, an analysis of the new dataset was carried out, which revealed some relationships between *like* labels and emotional labeling matches between artists and listeners, as well as other relationships between the listener's musical profile and musical genre labeling matches between

[9] https://umap-learn.readthedocs.io.

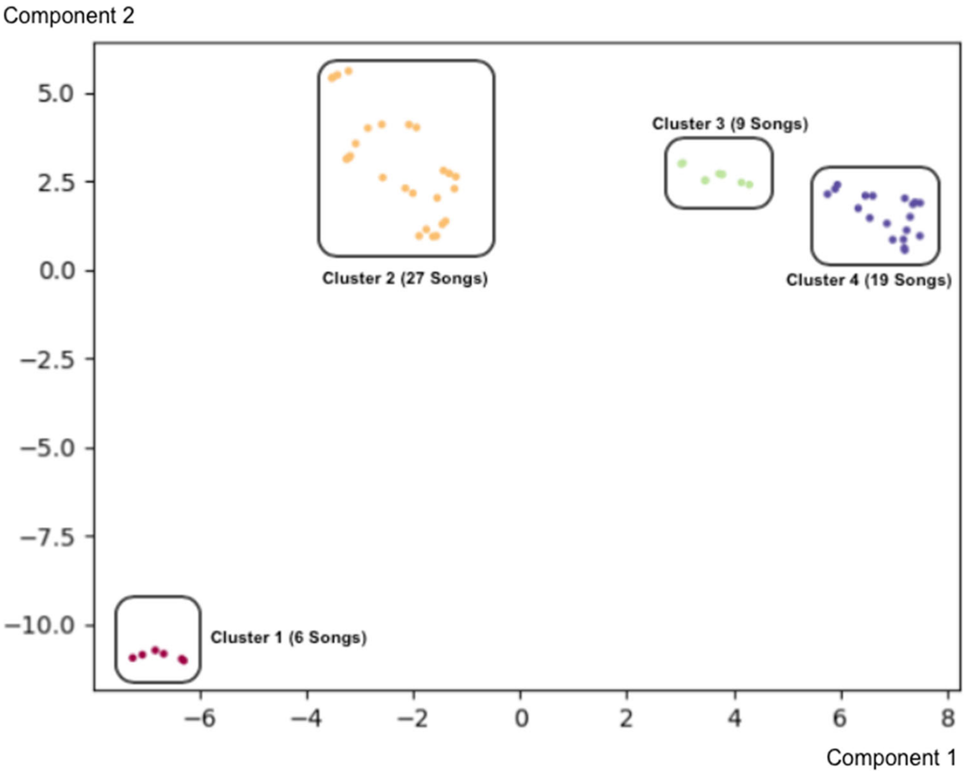
[10] https://hdbscan.readthedocs.io.

Fig. 6 Arousal value of the time series for the songs in cluster 5



artists and listeners. Moreover, the dataset was tested by doing an emotional analysis based on time-series, considering the importance of the temporal evolution of music from the emotional perception of the listeners and the availability of the musical structure in the metadata of each song. The experiments conducted on the dataset allowed to calculate the degree of similarity between the verses of the songs using the DTW algorithm, as well as implementing clustering strategies, obtaining results in which the songs generally coincide between the different clustering experiments.

Fig. 7 Clustering by valence with UMAP and HDBSCAN



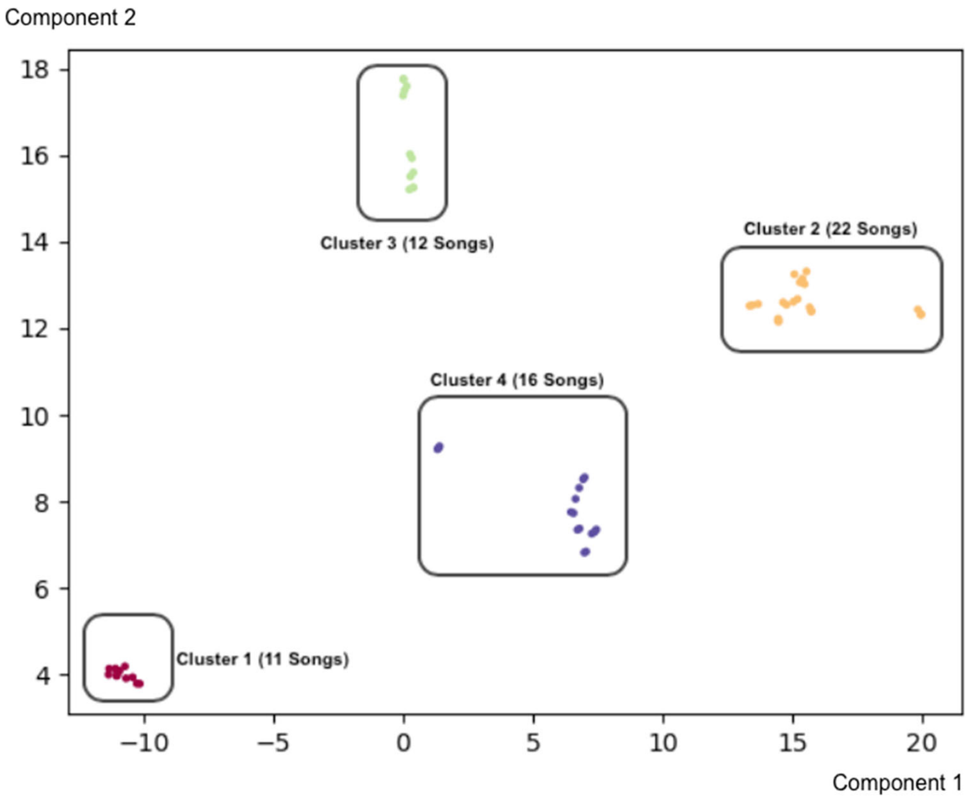


Fig. 8 Clustering by arousal with UMAP and HDBSCAN

Table 10 Comparison of song clustering matches in the different experiments

Algorithm	Valence	Arousal
DTW	Hysteria (0.0)	Hysteria (0.0)
	Pensarás en mí (0.61)	El swing de Caperucita (1.0)
	Esperando por ti (0.69)	A construir (1.20)
	Doncella de virgo (0.74)	Bienvenidos (1.30)
	Bienvenidos (0.87)	Esperando por ti (1.40)
	Intento de florecer (0.88)	Doncella de virgo (1.40)
	Currulao de nieves (0.95)	Estático ser (1.50)
	Bajo mi piel acústico (1.0)	Currulao de nieves (1.50)
	Peste de silicio (1.10)	Bajo mi piel (1.60)
	Estático ser (1.10)	Eterno abrazo (1.60)
	El swing de Caperucita (1.10)	Pensarás en mí (1.60)
		Hysteria (Cluster 5)
TimeSeriesKMeans	Hysteria (Cluster 1)	Currulao de nieves (Cluster 5)
with TSLEARN	Doncella de virgo (Cluster 1)	El swing de Caperucita (Cluster 5)
		A construir (Cluster 5)
UMAP	Hysteria (Cluster 4)	
	Esperando por ti (Cluster 4)	Hysteria (Cluster 1)
	Bajo mi piel acústico (Cluster 4)	Esperando por ti (Cluster 1)
	Doncella de virgo (Cluster 4)	Doncella de virgo (Cluster 1)
	Intento de florecer (Cluster 4)	Estático ser (Cluster 1)
	Pensarás en mí (Cluster 4)	Pensarás en mí (Cluster 1)
	Currulao de nieves (Cluster 4)	Currulao de nieves (Cluster 1)
	Bienvenidos (Cluster 4)	

7 Limitations and future work

Based on the depth degree of the experiments carried out and the discussion of their results, some limitations were identified in this work. The labeling processes only consider a single-label model which simplifies the recommendation strategy, but at the same, it limits the possibility for listeners to consider alternative labels. Moreover, some songs were emotionally labeled as a complete unit because they did not follow a musical structure that considers verses and choruses, which makes it difficult to ensure a clustering strategy based on the same kind of musical structure. The preliminary analysis of the dataset suggests the existence of some relationships between different data features, but these relationships are not yet very clear and it would be interesting to study them in depth since they could be very useful to improve the recommendation strategy, especially if they correspond to correlations. All clustering experiments were based on a single similarity metric, DTW. It would be interesting to repeat the experiments with the soft version of DTW and to explore other metrics that are applicable so that the results between different clustering experiments with different similarity metrics can be evaluated. Finally, although the inclusion of biases has been avoided, it is possible that the dataset includes a cultural bias considering that all the artists are from the same country.

In the near future, the analysis of the similarity and clustering of songs through the chorus and other parts of the musical structure will be studied. It will also be very important to design additional in-depth experiments to validate the emotional perception of end-users regarding the emotional intention of the artist and the results generated by the clustering strategies. For this purpose, it is important to extend the ENSA dataset, in order to include more songs and new labels and to increase the participation of different and more artists and listeners.

Further, there are some other interesting research lines that could be explored, such as the implementation of automatic analysis of the musical structure in order to avoid manual labeling of the song structure, the annotation of musical phrases on the dataset to perform experiments with song parts that have a duration shorter than a verse or chorus, the transition from a single-label model to a multi-label model in the labeling processes, as well as the consideration of other similarity metrics in new clustering experiments, like the *Mahalanobis* distance [13].

Acknowledgements We express our special gratitude to each one of the artists and bands that made this research possible. Bajo Cuerda, Psicophony, Madriguera, Skaparate, Kimberly Aguiar, Víctor Roll, Resistencia al Olvido, Lina Cardona Herrera, Denisse Rocío García Lozano, and Atadura.

Funding This research has been partially supported by the Spanish Ministry of Science, Innovation and Universities through project RTI2018-096986-B-C31 and the Aragonese Government through the AffectiveLab-T60-23R project. It has also been partially supported by the Computer Science School of the National University of La Plata (UNLP) through the Ph.D. program in Computer Science. This work is part of the project Technological Ecosystem for the MOod Recognition in cardiac rehabilitation patients (TEMOR), TED2021-130374B-C22, funded by MCIN/AEI/ 10.13039/501100011033 and by European Union NextGenerationEU/PRTR.

Declarations

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Abdollahpouri H, Mansoury M (2020) Multi-sided exposure bias in recommendation http://arxiv.org/abs/2006.15772, 2006.15772
2. Aucouturier J, Bigand E (2012) Mel cepstrum & ann ova: the difficult dialog between MIR and music cognition. In: Gouyon F, Herrera P, Martins LG, Müller M (eds) Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012, Mosteiro S.Bento Da Vitória, Porto, Portugal, October 8-12, 2012, FEUP Editorial, 2012, http://ismir2012.ismir.net/event/papers/397-ismir-2012.pdf pp 397–402
3. Bachorik JP, Bangert M, Loui P, Larke K, Berger J, Rowe R, Schlaug G (2009) Emotion in Motion: Investigating the Time-Course of Emotional Judgments of Musical Stimuli. Music Perception 26(4):355–364. https://doi.org/10.1525/mp.2009.26.4.355
4. Bauer C, Kholodylo M, Strauss C (2017) Music recommender systems challenges and opportunities for non-superstar artists. In: Digital Transformation - From Connecting Things to Transforming Our Lives, University of Maribor Press, Bled, pp 21–32, 10.18690/978-961-286-043-1.3
5. Bertin-Mahieux T, Ellis DP, Whitman B, Lamere P (2011) The million song dataset. In: Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)
6. Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. Knowledge-Based Systems 46:109–132, DOI: 10.1016/j.knosys.2013.03.012
7. Bogdanov D, Won M, Tovstogan P, Porter A, Serra X (2019) The MTG-Jamendo dataset for automatic music tagging
8. Celma O (2010) Music recommendation and discovery in the long tail. Springer, Barcelona. https://doi.org/10.1007/978-3-642-13287-2
9. Chamberlain A, Crabtree A (2016) Searching for Music: Understanding the Discovery, Acquisition, Processing and Organization of Music in a Domestic Setting for Design. Personal Ubiquitous Comput 20(4):559–571. https://doi.org/10.1007/s00779-016-0911-2
10. Chen J, Ying P, Zou M (2019) Improving music recommendation by incorporating social influence. Multimedia Tools and Applications 78(3):2667–2687. https://doi.org/10.1007/s11042-018-5745-7
11. Costa BG, Freire JCA, Cavalcante HS, Homci M, Castro ARG, Viegas R, Meiguins BS, Morais JM (2017) Fault classification on transmission lines using knn-dtw. In: Gervasi O, Murgante B,

Misra S, Borruso G, Torre CM, Rocha AMA, Taniar D, Apduhan BO, Stankova E, Cuzzocrea A (eds) Computational Science and Its Applications - ICCSA 2017. Springer International Publishing, Cham, pp 174–187

12. Cuturi M, Blondel M (2017) Soft-DTW: A differentiable loss function for time-series. 34th International Conference on Machine Learning, ICML 2017 2:1483–1505, http://arxiv.org/abs/1703.01541v2
13. De Maesschalck R, Jouan-Rimbaud D, Massart DL (2000) The mahalanobis distance. Chemometrics and intelligent laboratory systems 50(1):1–18
14. Deshmukh P, Kale G (2018) A survey of music recommendation system. In: International Journal of Scientific Research in Computer Science, vol 3, p 27
15. Eerola T, Vuoskoski JK (2013) A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli. Music Perception 30(3):307–340. https://doi.org/10.1525/mp.2012.30.3.307
16. Fan J, Yang YH, Dong K, Pasquier P (2020) A comparative study of Western and Chinese classical music based on soundscape models. In: 45th International Conference on Acoustics, Speech, and Signal Processing, IEEE, Barcelona
17. Fessahaye F, Perez L, Zhan T, Zhang R, Fossier C, Markarian R, Chiu C, Zhan J, Gewali L, Oh P (2019) T-RECSYS: a novel music recommendation system using deep learning. In: 2019 IEEE International Conference on Consumer Electronics (ICCE), IEEE, YILAN, pp 1–6 https://doi.org/10.1109/ICCE.2019.8662028
18. Frejman AE, Johansson D (2008) Emerging and conflicting business models for music content in the digital environment. In: eChallenges e-2008, IOS Press, Stockholm
19. Friedman B (1996) Bias in computer systems. ACM Transactions on Information Systems 14(3), 330–347, DOI: 10.1145/230538.230561
20. Gabrielsson A, Lindström E (2010) The role of structure in the musical expression of emotions. Handbook of music and emotion: Theory, research, applications pp 367–400, https://doi.org/10.1093/acprof:oso/9780199230143.003.0014
21. Gemmeke JF, Ellis DPW, Freedman D, Jansen A, Lawrence W, Moore RC, Plakal M, Ritter M (2017) Audio set: an ontology and human-labeled dataset for audio events. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 776–780 https://doi.org/10.1109/ICASSP.2017.7952261
22. Gouyon F, Klapuri A, Dixon S, Alonso M, Tzanetakis G, Uhle C, Cano P (2006) An experimental comparison of audio tempo induction algorithms. IEEE Transactions on Audio, Speech, and Language Processing 14(5), 1832–1844, DOI: 10.1109/TSA.2005.858509
23. Hesmondhalgh D (2021) Is music streaming bad for musicians? Problems of evidence and argument. New Media & Society 23(12):3593–3615. https://doi.org/10.1177/1461444820953541
24. IFPI (2021) Global Music Report 2021. Tech. rep., IFPI, London
25. Jin Y, Htun NN, Tintarev N, Verbert K (2019) Contextplay. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, ACM, New York, pp 294–302 https://doi.org/10.1145/3320435.3320445
26. Juslin P, Juslin PN, Sloboda J, Sloboda P, Frijda N (2010) Handbook of music and emotion: theory, research, applications. Affective Science, OUP Oxford, Oxford, https://books.google.com.co/books?id=t8j5pduTkboC
27. Juslin PN, Harmat L, Eerola T (2014) What makes music emotionally significant? exploring the underlying mechanisms. Psychology of Music 42(4):599–623
28. Katarya R, Verma OP (2018) Efficient music recommender system using context graph and particle swarm. Multimedia Tools and Applications 77(2), 2673–2687, DOI: 10.1007/s11042-017-4447-x
29. Law E, West K, Mandel M, Bay M, Downie JS (2009) Evaluation of algorithms using games : the case of music tagging. In: In Proc. wISMIR 2009
30. Lee JH, Downie JS (2004) Survey of music information needs, uses, and seeking behaviours: Preliminary findings. ISMIR 2004:441–446
31. Mesaros A, Heittola T, Virtanen T (2016) Tut database for acoustic scene classification and sound event detection. In: 2016 24th European Signal Processing Conference (EUSIPCO), pp 1128–1132 https://doi.org/10.1109/EUSIPCO.2016.7760424
32. Nielzen S, Cesarec Z (1982) Emotional experience of music as a function of musical structure. Psychology of Music 10(2):7–17
33. Ospitia-Medina Y, Baldassarri S, Beltrán JR (2019a) High-level libraries for emotion recognition in music: a review. In: Agredo V, Ruiz P (eds) Human-Computer Interaction. HCI-COLLAB 2018., Springer, Popayán, pp 158–168 https://doi.org/10.1007/978-3-030-05270-6_12
34. Ospitia-Medina Y, Beltrán JR, Sanz C, Baldassarri S (2019b) Dimensional emotion prediction through low-level musical features. In: ACM (ed) Audio Mostly (AM'19), Nottingham, p 4, https://doi.org/10.1145/3356590.3356626
35. Ospitia-Medina Y, Beltrán JR, Baldassarri S (2020) Emotional classification of music using neural networks with the MediaEval dataset. Personal and Ubiquitous Computing 10.1007/s00779-020-01393-4
36. Ospitia-Medina Y, Baldassarri S, Sanz C, Beltrán JR (2022) Music recommender systems: a review centered on biases (In press). Advances in Speech and Music Technology: Computational Aspects and Applications
37. Paul D, Kundu S (2020) A survey of music recommendation systems with a proposed music recommendation system. Advances in Intelligent Systems and Computing, vol 937, Springer Singapore, Singapore, pp 279–285 https://doi.org/10.1007/978-981-13-7403-6_26
38. Piczak KJ (2015) Esc: dataset for environmental sound classification. Association for Computing Machinery, New York, NY, USA, MM '15, p 1015-1018 https://doi.org/10.1145/2733373.2806390
39. Russell JA (1980) A circumplex model of affect. Journal of Personality and Social Psychology 39(6):1161–1178
40. Salamon J, Jacoby C, Bello JP (2014) A dataset and taxonomy for urban sound research. In: Proceedings of the 22nd ACM International Conference on Multimedia, Association for Computing Machinery, New York, NY, USA, MM '14, p 1041–1044 https://doi.org/10.1145/2647868.2655045
41. Schedl M, Zamani H, Chen CW, Deldjoo Y, Elahi M (2018) Current Challenges and Visions in Music Recommender Systems Research. International Journal of Multimedia Information Retrieval 7(2):95–116. https://doi.org/10.1007/s13735-018-0154-2
42. Schubert E (2004) Modeling perceived emotion with continuous musical features. Music Perception 21:561–585
43. Semeraro A, Vilella S, Ruffo G (2021) Pyplutchik: Visualising and comparing emotion-annotated corpora. PLOS ONE 16(9):1–24. https://doi.org/10.1371/journal.pone.0256503
44. Shah F, Desai M, Pati S, Mistry V (2020) Hybrid music recommendation system based on temporal effects. In: Advances in Intelligent Systems and Computing, vol 1034, pp 569–577, DOI: 10.1007/978-981-15-1084-7_55

45. Sloboda J (1986) The musical mind, oxford psy edn. Oxford University Press, New York, https://doi.org/10.1093/acprof:oso/9780198521280.001.0001
46. Sloboda J (1991) Music structure and emotional response: Some empirical findings. Psychology of music 19:110–120. https://doi.org/10.1177/0305735691192002
47. Soleymani M, Aljanaki A, Yang YH (2016) DEAM: MediaEval database for emotional analysis in music pp 3–5, http://cvml.unige.ch/databases/DEAM/manual.pdf
48. Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, Payne M, Yurchak R, Rußwurm M, Kolar K, Woods E (2020) Tslearn, a machine learning toolkit for time series data. Journal of Machine Learning Research 21:1–6
49. Tzanetakis G, Cook P (2002) Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing 10(5), 293–302, DOI: 10.1109/TSA.2002.800560
50. Yang S, Reed CN, Chew E, Barthet M (2021) Examining emotion perception agreement in live music performance. IEEE Transactions on Affective Computing pp 1–1 https://doi.org/10.1109/TAFFC.2021.3093787
51. Yang Yh, Chen HH (2012) Machine recognition of music emotion. ACM Transactions on Intelligent Systems and Technology 3(3):1–30. https://doi.org/10.1145/2168752.2168754
52. Zamani H, Schedl M, Lamere P, Chen C (2019) An analysis of approaches taken in the ACM recsys challenge 2018 for automatic music playlist continuation. ACM Trans Intell Syst Technol 10(5):57:1–57:21. https://doi.org/10.1145/3344257
53. Zhang K, Zhang H, Li S, Yang C, Sun L (2018) The PMEmo dataset for music emotion recognition. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, Association for Computing Machinery, New York, NY, USA, ICMR '18, p 135–142 https://doi.org/10.1145/3206025.3206037
54. Zheng HT, Chen JY, Liang N, Sangaiah A, Jiang Y, Zhao CZ (2019) A Deep Temporal Neural Music Recommendation Model Utilizing Music and User Metadata. Applied Sciences 9(4):703. https://doi.org/10.3390/app9040703

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