Data article

Encontros Bibli

LIRA – INTERMODAL LANGUAGE FOR AFFECTIVE RECOGNITION: A MULTIMODAL DATABASE FOR MUSIC EMOTION RECOGNITION

LIRA – Linguagem Intermodal de Reconhecimento Afetivo: uma base de dados multimodal para reconhecimento de emoções musicais

Paulo Sergio da Conceição Moreira
Universidade Federal do Paraná
Curitiba, PR, Brasil
psxm54@gmail.com
https://orcid.org/0000-0002-3168-2177

Denise Fukumi Tsunoda
Universidade Federal do Paraná
Curitiba, PR, Brasil
dtsunoda@ufpr.br
https://orcid.org/0000-0002-5663-4534

Marília Nunes-Silva
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brasil
marilianunespsi@gmail.com
https://orcid.org/0000-0003-1329-2682

A lista completa com informações dos autores está no final do artigo

ABSTRACT

Objective: this paper presents LIRA — Intermodal Language for Affective Recognition, a multimodal dataset designed to advance research in music emotion recognition. It addresses limitations in existing databases by providing rich emotional annotations alongside diverse feature representations across five modalities.
Method: LIRA contains 1,412 thirty-second song segments, each labeled with one of four discrete emotions: joy, anger/fear, serenity, or sadness. The dataset encompasses five modalities: audio, chords, lyrics, symbolic features, and voice. Feature extraction was performed using tools including Librosa, Essentia, music21, and Spleeter.
Results: a total of 171 features were extracted: 67 from audio, 58 from voice, 25 from chords, 12 symbolic, and 9 from lyrics. Emotional and structural data are systematically organized in a reusable format. All data and processing scripts are publicly accessible via Mendeley Data and GitHub.
Conclusions: LIRA is a publicly available, multimodal, affectively annotated database that fosters robust and reproducible research in music emotion recognition. Its multimodality and standardized structure enable comprehensive exploration of emotional responses to music and support the development of more expressive computational models.
KEYWORDS: Database. Multimodal. Music emotion recognition. Music feature extraction. Information retrieval.

RESUMO

Objetivo: este artigo apresenta a LIRA - Linguagem Intermodal de Reconhecimento Afetivo, uma base de dados multimodal desenvolvida para apoiar pesquisas em reconhecimento de emoções musicais. A LIRA preenche lacunas de bases existentes ao oferecer anotações emocionais e representações ricas em cinco modalidades.
Método: a base é composta por 1.412 segmentos de 30 segundos de músicas, cada um rotulado com uma das quatro emoções discretas: alegria, raiva/medo, serenidade ou tristeza. A LIRA inclui cinco modalidades: áudio, acordes, letras, atributos simbólicos e voz. A extração das características foi realizada com ferramentas como Librosa, Essentia, music21 e Spleeter.
Resultados: foram extraídas 171 características no total: 67 do áudio, 58 da voz, 25 dos acordes, 12 simbólicas e 9 das letras. Os dados emocionais e estruturais estão organizados em formato reutilizável. Todo o material e os scripts estão disponíveis publicamente no Mendeley Data e GitHub.

Conclusões: a LIRA é uma base de dados multimodal e anotada afetivamente, disponível publicamente, que favorece pesquisas robustas e reprodutíveis em reconhecimento de emoções musicais. Sua diversidade modal e formato padronizado permitem uma exploração aprofundada das respostas emocionais à música e apoiam o desenvolvimento de modelos computacionais mais expressivos.
PALAVRAS-CHAVE: Banco de dados. Multimodal. Reconhecimento de emoções musicais. Extração de características musicais. Recuperação da informação.

1 INTRODUCTION

Compared to other stimuli, music can induce a deeper and longer-lasting emotional experience and arouse a wide range of emotions (Rajesh; Nalini, 2020). However, while organizing songs according to their emotional impact is intuitive for humans, it presents significant challenges for computers (Yang; Dong; Li, 2017), such as the variability of musical content between genres, the cultural context of listeners, and the reliability of the data used (Fan *et al.,* 2017).

Regarding data use, obtaining songs properly associated with a set of labeled emotions presents a significant challenge for music emotion recognition (MER) development systems. For example, research data is rarely made publicly available due to copyright restrictions, limiting its reuse in new studies (Chaturvedi *et al.,* 2021). These limitations also hinder direct access to and manipulation of the used music files, complicating the replication of procedures for extracting these features.

In addition to access difficulties, an analysis of studies that compiled databases for music emotion recognition (MER), such as those by Panda (2019) and Gómez-Cañón *et al*. (2021), revealed factors that hinder the reuse of these databases in subsequent research. In their thesis, for example, Panda (2019) examined 17 databases generally considered in MER studies, 11 of which have fewer than 1,000 records. Among databases analyzed by (Panda, 2019), only two have more than two modalities: a) Greek Music Dataset (GMD), composed of audio, lyrics, and Musical Instrument Digital Interface (MIDI); and b) Multimodal MIREX-like, also composed of audio, lyrics, and MIDI. However, these databases do not have an equal number of records for all the modalities considered (Panda, 2019).

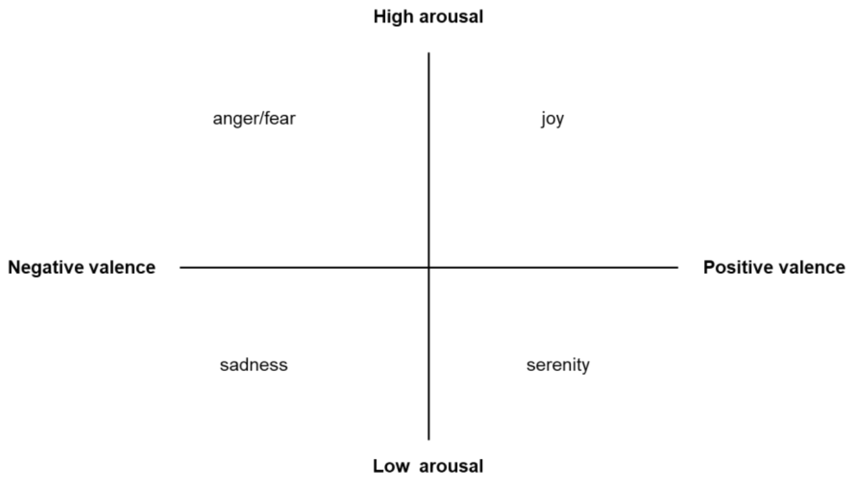
Similarly, in their surveys, Panda (2019) and Gómez-Cañón *et al.* (2021) compiled a set of databases and made their lists available in an online repository. Of the 16 databases analyzed by the authors, eight are included in the study by Panda (2019), seven of which contain fewer than 1,000 records. In addition to these seven databases, (Gómez-Cañón *et al*., 2021) added five more databases also with less than 1,000 records. Notably, none of the databases compiled by the authors present more than one modality.

Based on the research and the challenges identified in existing databases, we recognize the necessity of developing a multimodal database for recognizing emotions in music, intending to facilitate its reuse by new researchers. Our main goal is to establish a multimodal database focused on emotion recognition in music, following specific requirements: a) selection of emotion-related songs aligned with an emotion model from the area of musical cognition; b) inclusion of English lyrics for all records; c) standardization of the duration range of songs for segmentation; d) direct manipulation of music files to extract relevant features; e) implementation of a third-party review process to ensure data quality and consistency.

2 METHODS AND INSTRUMENTS

There is no consensus on the precise taxonomy or exact number of emotional labels that a user might experience (Russo *et al.,* 2020). Thus, we adopted a model that enables a clear distinction between different emotions, a requirement for approaches based on discrete models (Russo *et al*., 2020). For this purpose, we used the model proposed by Bigand *et al.* (2005), which defines the emotions of joy, anger/fear, serenity, and sadness (Figure 1).

Figure 1 - Discrete model of emotions selected



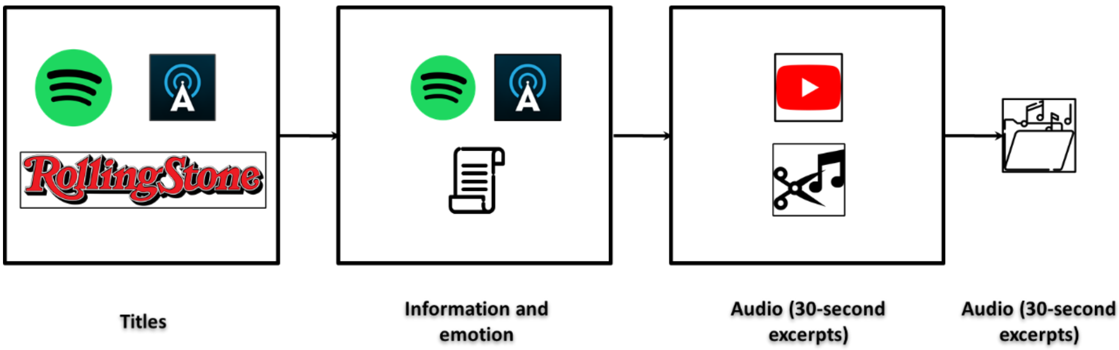
Source: adapted from Nunes-Silva *et al*. (2016).

The segmentation strategy employed involved a semiautomatic selection of the first 30 seconds of songs containing lyrics. We considered this approach the most efficient for ensuring that all recordings included vocal and textual content, which is essential to maintaining the multimodal nature of the research. The choice of 30-second segments relies on studies that discuss the optimal duration for recognizing emotions in music, both in computational context and in research determining the minimum time required to evoke an emotional response in volunteers (Russo *et al.* 2020).

To obtain 30-second segments of each song, the first step was to identify manually when the lyrics begin. Next, we developed a script with the Pydub library [1] to add 30 seconds from that starting point and perform the cut automatically. This strategy ensured homogeneity in segment length, maintaining all of them at 30 seconds. However, it is necessary to recognize the limitations of this strategy. The methodology adopted may introduce abrupt breaks in the continuity narrative of the music and lyric context, potentially affecting the emotional perception conveyed. Additionally, by emphasizing the uniformity of the segments, the emotional information present in the song introductions was not contemplated.

To identify the songs used in this experiment (Figure 2), we considered three primary sources: the AllMusic platform, the Spotify streaming service, and the songs from the 500 best albums in history, according to the 2020 edition of Rolling Stone Magazine.

Figure 2 - Summary of the song selection process



Source: prepared by the authors (2025).

We chose AllMusic due to its use in previous studies (Panda, 2019) and because it is an environment with emotional adjectives controlled and assigned by professional editors and experts (Panda, 2019), which minimizes noise compared to platforms such as Last.FM, where users do labeling freely. However, it is relevant to note that the AllMusic process does not allow a detailed critical analysis of its stages (Panda, 2019).

[1] Available at: https://pypi.org/project/pydub/. Accessed on 11 jul. 2025

We chose Spotify for its extensive music catalog, which offers access to a wide range of genres and artists, and it also provides detailed metadata such as genre, artist, and album information, making it easier to select tracks that meet the requirements of this study. The inclusion of Rolling Stone's list of the 500 best albums of 2020 is justified by its timeliness and comprehensiveness, which predominantly features English-language albums spanning diverse genres and popular artists from the 20th and 21st centuries, enhancing the representativeness and variety of the musical sample used in the research.

After defining the sources, we developed Python scripts using the BeautifulSoup and Selenium libraries to collect song links from AllMusic pages based on 51 emotional adjectives (Juslin; Laukka, 2004; Yang; Lee, 2009), and the synonyms suggested by AllMusic. We included the following data: a) song titles; b) artist name; and c) all the emotions associated with the songs.

This process resulted in 5,757 songs. To expand the number of records in the database, we collected song titles corresponding to the emotions described in the model in Bigand *et al.* (2005) from 29 playlists on Spotify's official profile. We selected the playlists using English terms for each emotion, which resulted in 2,966 songs. The song information, such as song titles and artist names, was extracted using a Python script with the Spotipy library. Finally, the audio files were obtained using the pytube library, ensuring the necessary audio excerpt was collected for the screening stage.

After collecting the metadata, we conducted a screening process based on the following exclusion criteria: repeated songs, songs without emotional adjectives (AllMusic), and songs with emotional adjectives not listed in the proposed terms (AllMusic). We also excluded live songs, instrumentals, songs with lyrics not recorded in English, and songs shorter than 45 seconds. After this stage, 4,797 songs were retained for the initial list in this study.

To label the emotions associated with the songs retrieved from both AllMusic and Spotify, we discarded the emotional adjectives from these platforms. Instead, we determined the "final emotion" for each segment based on the annotations from a set of evaluators. The evaluation was conducted blindly, meaning the participants were unaware of the song titles or artist names, and there was no interaction between them during the process.

We selected the participants by disseminating the research through undergraduate and postgraduate courses in various fields of knowledge in Brazil. This approach of recruiting volunteers, regardless of their academic background, aligns with recommendations in the related literature, which underscores the importance of including specialists and nonspecialists to develop systems that apply to a diverse audience (Gómez-Cañón *et al.,* 2021).

Despite the extensive dissemination strategy, participation in the survey was limited, initially resulting in only two volunteers to evaluate all the records in a database of 4,797 songs. To ensure a complete analysis, we recruited three additional volunteers, each responsible for labeling approximately 1,599 records. Thus, at the end of the process, each song in this study was evaluated by three different volunteers. From a descriptive point of view, the participants ranged in age from 23 to 65 years (mean 45.20 years, standard deviation 19.60).

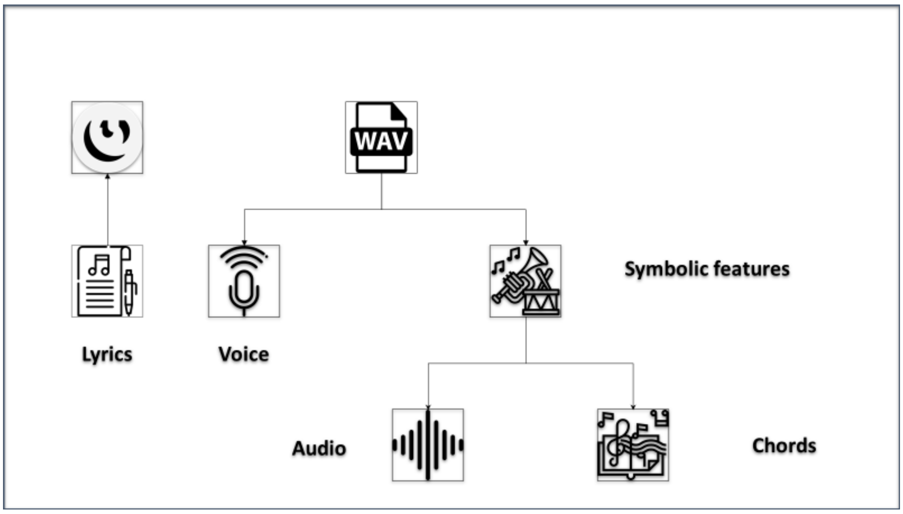
Regarding gender, two volunteers identified as male and three as female. The evaluation process adopted for the experiment was conducted remotely, with the audio files made available to evaluators so they could take notes conveniently regarding time and location. This took place between 08/04/2022 and 10/07/2022. It is noteworthy that the volunteers were native Brazilian Portuguese speakers but claimed to have at least an intermediate proficiency in English. The new evaluation process included the following steps: a) presentation of an Informed Consent Form; b) detailed instructions about the experiment; c) collection of demographic data (gender, age, presence of hearing problems, and musical training); and d) ongoing support for the evaluators throughout the evaluation period.

As a criterion for deciding which songs would build the database, we used the agreement among the evaluators. We assessed the reliability of this agreement using Fleiss' kappa statistic (Fleiss, 1971), which is suitable for measuring agreement levels in experiments with more than two evaluators and categorical variables. To assess inter-rater reliability, Fleiss' Kappa statistic was used, revealing mostly reasonable agreement among evaluators, with one subset of raters achieving moderate agreement. We can observe that agreements were mostly reasonable, except for the result achieved for the records evaluated by volunteers one, two, and five, whose Fleiss' Kappa value indicates moderate agreement. All observed values are statistically significant, considering a 95.00% confidence interval.

One of the steps in building MER systems involves extracting features related to musical content (Russo *et al*., 2020). This phase aims to transform the selected set of songs into features that can fully describe them, including acoustic and musical components such as timbre, rhythm, and harmony, as well as the lyrics. However, feature extraction requires a detailed musical content analysis. Before this stage, two preprocessing procedures were necessary: (a) standardizing the data, and (b) defining the duration of the recordings (Yang; Dong; Li, 2017).

The features used in this study encompass five modalities: chords, audio, lyrics, symbolics, and voice (Figure 3). Only features related to musical content were considered, excluding metadata-based resources. To obtain chords and symbolic features, we used specific algorithms and libraries (Mauch, 2010), which allowed us to avoid relying on external sources such as ciphers and MIDI files, ensuring the integrity of the records. We consider the estimation of features such as predominant melodic lines relevant, albeit subject to improvement, since they complement information not included in traditional MER experiments.

Figure 3. Process of extracting the modalities



Source: prepared by the authors (2025).

To obtain the lyrics, a Python script was used to collect them directly from the Genius website, as limitations in automatic transcription made it impossible to automate this process. The vocal modality was acquired using the Spleeter library[2], which allowed us to split the vocal and instrumental parts of the songs. All the audio files were standardized to 30 seconds in length, using the WAV format, a sampling rate of 22,500 Hz, a sample size of 16 bits, and a bit rate of 1,411 kbps.

[2] Available at: https://github.com/deezer/spleeter. Accessed on 11 jul. 2025.

3 SPECIFICATION TABLE

Table 1 - Dataset specifications

Area of knowledge	Information treatment for information services
Specific subject area	Multimodal Music Information Representation for Music Emotion Recognition.
Language	English
Data type	Number
How the data was acquired	Song titles and metadata were collected from AllMusic and Spotify using Python scripts. Audio was sourced from YouTube and segmented into 30-second clips starting from the first lyrics. Feature extraction covered five modalities (audio, lyrics, chords, voice, and symbolics) using libraries such as librosa, Essentia, Spleeter, and music21.
Data state	• Filtered • Analyzed
Parameters for data collection	Emotional criteria. On AllMusic, 51 predefined emotional adjectives and their synonyms were used to filter tracks. On Spotify, songs were selected from 29 curated playlists labeled with emotion terms defined by Bigand *et al.* (2005).
Description of data collection	The data collection process involved web scraping and data extraction techniques implemented in Python. Libraries such as BeautifulSoup and Selenium were used to parse and automate retrieval of song metadata and emotion labels from various sources. Audio data was collected via YouTube links to support multimodal analysis. This automated pipeline enabled the construction of a comprehensive dataset for music emotion recognition research.
Data source location	The final curated dataset is publicly available on Mendeley Data. This repository contains all processed data, including extracted features and emotion labels. The original raw data was collected from AllMusic, Spotify, and YouTube, as detailed in the methodology section.
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/9zdww6wnyx.3 URL: https://data.mendeley.com/datasets/9zdww6wnyx/3

Source: Research data (2025).

3.1 Dataset description

From an initial concordance analysis of 4,797 records, we retained 1,412 songs for further study based on evaluator agreement and exclusion criteria. Among the emotional categories, joy and sadness were predominant, together representing just over 60% of the database (Table 2).

Table 2 - Emotion distribution in analyzed songs

Emotion	Frequency (%)
Joy	442 (31.3%)
Anger/fear	236 (16.7%)
Serenity	305 (21.6%)
Sadness	429 (30.4%)

Source: Research data (2025).

Regarding the total number of features extracted from each modality (Table 3), each one contributes a specific number of features, reflecting the diversity of attributes analyzed in the study. The audio modality has the highest number of extracted features, followed by voice, while lyrics have the lowest number.

Table 3 - Total features extracted for each modality

Modality	Number of features
Chords	25
Audio	67
Lyrics	9
Symbolics	12
Voice	58
Total	**171**

Source: Research data (2025).

For chord estimation, we used the chord-extractor library, a Python implementation of the Chordino plugin. Chordino is based on the strategy proposed by Mauch (2010), which employs the Non-Negative Least Squares (NNLS) algorithm to enhance chord recognition (NNLS Chroma) **(**Mauch, 2010). Chords were extracted considering only the instrumental parts of the songs, applying the standard values recommended by the Chordino documentation. Each estimated chord was classified by its respective type (e.g., Fdim, C7, C6) (Table 4).

Table 4 - Chord features and descriptions

Feature – Description
quantidadeAcordes – Total number of chords in the piece
outros – Chords classified as other (non-standard or uncommon)
setima – Chords with seventh (general)
setimaMaior – Major seventh chords
setimaMenor – Minor seventh chords
setimaMenorQuinta – Minor seventh chords with diminished fifth

sexta – Chords with sixth (general)

sextaMenor – Minor sixth chords

triadeAumentada – Augmented triad chords

triadeDiminuta – Diminished triad chords

triadeMaior – Major triad chords

triadeMaior – Major triad chords

triadeMenor – Minor triad chords

acordesUnicos – Number of unique chords

proporcaoUnicos – Proportion of unique chords relative to total

propTriadeMenor – Proportion of minor triads

propTriadeMaior – Proportion of major triads

propTriadeDim – Proportion of diminished triads

propTriadeAum – Proportion of augmented triads

propSextaMenor – Proportion of minor sixth chords

propSexta – Proportion of sixth chords (general)

propSetima – Proportion of seventh chords (general)

propSetimaMaior – Proportion of major seventh chords

propSetimaMenor – Proportion of minor seventh chords

propSetimaMenorQuinta – Proportion of minor seventh chords with diminished fifth

propOutros - Proportion of other chords

Source: Research data (2025).

The audio features (Table 5) encompass the fundamental properties of music: dynamics, harmony, melody, rhythm, and timbre (McKay, 2010). Dynamics were analyzed by extracting RMS (average energy of a signal) and Loudness (subjective sound intensity) (McKay, 2010). Harmony was represented by the chromagram—a 12-dimensional vector where each dimension corresponds to the energy concentration for each of the 12 semitones. Melody was described by Pitch Salience, which indicates pitch perceptibility, and Predominant Pitch Melody, extracted with the MELODIA algorithm, which identifies the fundamental frequency of the melody based on tone contours and auditory cues.

Table 5 - Extracted audio properties and features

Property	Feature	Attr.	Tool
Dynamics	RMS	Mean, std. dev. (2)	librosa
	Loudness	Total (1)	Essentia
Harmony	Chromagram	Mean, std. dev. (24)	librosa
Melody	Pitch Salience	Total (1)	Essentia
	Pred. Pitch Melody	Mean, std. dev. (2)	Essentia
Rhythm	BPM	Total (1)	Essentia
	Danceability	Total (1)	Essentia
	PLP	Mean, std. dev. (2)	librosa
Timbre	MFCCs	Mean, std. dev. (26)	
	Spec. Centroid	Mean, std. dev. (2)	librosa
	Spec. Rolloff	Mean, std. dev. (2)	librosa
	ZCR	Mean, std. dev. (2)	librosa
	Spec. Flux	Total (1)	Essentia

Source: Research data (2025).

For rhythm, the following features were extracted: BPM (beats per minute), Danceability, estimated using the Detrended Fluctuation Analysis algorithm, and PLP (which captures dominant tempo even in non-percussive songs). Timbre was characterized using MFCCs (13 coefficients), Spectral Centroid, Spectral Flux, Spectral Rolloff, and Zero-Crossing Rate (ZCR) (McKay, 2010). All audio features were extracted using librosa[3] and Essentia[4], using default parameters except for MFCCs.

Regarding lyrics, since automatic transcription tools from Microsoft and Google were unsatisfactory, we developed a script to collect them from the Genius website. Lyrics were matched to the 30-second audio excerpts manually. Two categories of numerical features were selected: word frequency metrics and sentiment analysis (via the VADER model[5]) (Table 6). Frequencies included total word and character count, average words per line, and vocabulary richness. Sentiment features included positive, negative, neutral, and compound sentiment scores. All words were converted to lowercase and special characters, numbers, and tags like “(chorus)” were removed using NLTK. Stop words were not removed due to the short length of each excerpt.

[3] Available at: https://librosa.org/doc/latest/index.html. Accessed on 11 jul. 2025.
[4] Available at: https://essentia.upf.edu/essentia_python_tutorial.html. Accessed on 11 jul. 2025.
[5] Available at: https://github.com/cjhutto/vaderSentiment. Accessed on 11 jul. 2025.

Table 6 - Extracted lyric features

Feature – Description
charCount – Character count
wordCount – Word count
avgWord – Avg. words per line
uniqueWords – Unique word count
uniqueWordsProp – Unique word proportion
neg – Negative sentiment
neu – Neutral sentiment
pos – Positive sentiment
compound – Sentiment summary metric

Source: Research data (2025).

Symbolic features (Table 7) were extracted after transcribing audio into MIDI using the Omnizart library[6], which employs deep learning and U-Net architecture. Despite challenges in automatic transcription, this approach enabled full dataset coverage without external MIDI files. The resulting piano-style MIDI files (instrumental only) were processed using the music21 library[7], specifically its jSymbolic module, to extract 12 features appropriate for symbolic music analysis (McKay, 2010).

Table 7 - Extracted symbolic features

Feature – Description
Primary register – Avg. pitch height in MIDI
Avg. note duration – Avg. note duration (s)
Note duration var. – Std. dev. of note duration (s)
Max. note duration – Max. note duration (s)
Min. note duration – Min. note duration (s)
Range – Highest-lowest pitch diff. (semitones)
Note density – Avg. notes per second
Bass register – Fraction of notes (MIDI 0-54)
Middle register – Fraction of notes (MIDI 55-72)

[6] Available at: https://music-and-culture-technology-lab.github.io/omnizart-doc/. Accessed on 11 jul. 2025.

[7] Available at: https://pypi.org/project/music21/. Accessed on 11 jul. 2025.

High register – Fraction of notes (MIDI 73-127)

Most common pitch – Fraction of most common pitch class

Pitch class variety – Number of pitch classes used

Source: Research data (2025).

For the voice modality (Table 8), the Spleeter library was used to isolate vocal signals from instrumental content, based on the U-Net architecture. We extracted features across three categories: spectral, prosodic, and voice quality. Spectral features included Formants, GFCCs (12 coefficients), and MFCCs (first 16 coefficients). Prosodic features included Duration (voiced/unvoiced segments) and F0 (fundamental frequency). Quality-related features included HNR (harmonic-to-noise ratio), Jitter (glottal cycle variability), and Shimmer (amplitude variability). Most features were extracted using openSMILE[8], Essentia, and librosa, with parameters kept at default, except for the number of coefficients in GFCCs and MFCCs.

Table 8 - Features extracted from the voice

Category	**Feature**	**Attr.**	**Tool**
Spectral	Formants	Frequency and bandwidth (mean and std)	openSMILE
	GFCC	12 coefficients (mean and std)	Essentia
	MFCC	First 16 coefficients (mean and std)	librosa
Prosodic	Duration	Voiced/unvoiced regions (mean and std)	openSMILE
	F0	Fundamental frequency (mean and std)	openSMILE
Quality	HNR	Average and std of HNR	openSMILE
	Jitter	Average and std of jitter	openSMILE
	Shimmer	Average and std of shimmer	openSMILE

Source: Research data (2025).

The constructed database covers a wide variety of musical genres, reflecting the diversity of the analyzed content. The Rock genre is the most represented in the database, comprising 26.27% of the total number of songs. Pop comes in second place with 23.87%, followed by Heavy Metal with 12.04%. The R&B and Folk genres are also significantly represented, with 11.26% and 8.29% respectively.

Genres such as Electronic, Rap, and Country have a moderate presence, with percentages between 3.40% and 5.10%. On the other hand, Punk, Jazz, and Blues appear in smaller proportions, each accounting for less than 3.0% of the total. Other genres, such as Reggae, Avant-garde, Vocal, Religious, and Easy Listening, have even smaller shares,

[8] Available at: https://github.com/audeering/opensmile-python. Accessed on 11 jul. 2025.

with less than 1.0% each. The main artists in the database include The Beatles, Frank Sinatra, Elvis Presley, Merle Haggard, Jerry Lee Lewis, Ed Sheeran, Bob Dylan, Megadeth, The Clash, and The Magnetic Fields among others. Each contributes a varying amount of songs.

All extracted features for the 1,412 annotated song segments are available in the file lira.csv. This file contains 171 features distributed across five modalities: chords, audio, lyrics, symbolic, and voice. Each row corresponds to a unique 30-second segment labeled with one of the four target emotions. In addition, all scripts used for feature extraction, preprocessing, and organization of the dataset are publicly available in a GitHub repository: https://github.com/PSCM. This ensures transparency and facilitates reproducibility of the data construction process.

REFERENCES

BIGAND, E. *et al.* Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. **Cognition and Emotion**, [*s. l.*], v. 19, n. 8, p. 1113-1139, 2005. Available at: https://doi.org/10.1080/02699930500204250. Accessed on: 11 July 2025.

CHATURVEDI, V. *et al*. Music mood and human emotion recognition based on physiological signals: a systematic review. **Multimedia Systems**, [*s. l.*], v. 28, n. 21, p. 21-44, apr. 2021. Available at: https://doi.org/10.1007/s00530-021-00786-6. Accessed on: 11 July 2025.

FAN, J. *et al.* Ranking-Based emotion recognition for experimental music. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 18., 2017, Suzhou. **Proceedings** [...]. Suzhou: ISMIR, 2017. p. 368-375. Available at: http://archives.ismir.net/ismir2017/paper/000217.pdf. Accessed on: 11 July 2025.

FLEISS, J. L. Measuring nominal scale agreement among many raters. **Psychological Bulletin**, [*s. l.*], v. 76, n. 5, p. 378-382, Nov. 1971. Available at: https://psycnet.apa.org/doi/10.1037/h0031619. Accessed on: 11 July 2025.

GÓMEZ-CAÑÓN, J. S. *et al.* Music emotion recognition: toward new, robust standards in personalized and context-sensitive applications. **IEEE Signal Processing Magazine**, [*s. l.*], v. 38, n. 6, p. 106-114, Oct. 2021. Available at: https://doi.org/10.1109/MSP.2021.3106232. Accessed on: 11 July 2025.

JUSLIN, P. N.; LAUKKA, P. Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. **Journal of New Music Research**, [*s. l.*], v. 33, n. 3, p. 217-238, 2004. Available at: https://doi.org/10.1080/0929821042000317813. Accessed on: 11 July 2025.

MCKAY, C. **Automatic music classification with jMIR**. 2010. 600 f. Tese (Doutorado) - Curso de Music Technology, Music, McGill University, Montreal, 2010. Available at: http://jmir.sourceforge.net/publications/PhD_Dissertation_2010.pdf. Accessed on: 27 July 2025.

MAUCH, M. **Automatic chord transcription from audio using computational models of musical context**. 2010. 168 f. Tese (Doutorado) - Curso de Electronic Engineering and Computer Science, University of London, Londres, 2010.

NUNES-SILVA, M. *et al.* Avaliação de músicas compostas para indução de relaxamento e de seus efeitos psicológicos. **Psicologia**: Ciência e Profissão, [*s. l.*], v. 36, n. 3, p. 709-725, July/Sept. 2016. Available at: https://doi.org/10.1590/1982-3703001672014. Accessed on: 11 July 2025.

PANDA, R. E. S. **Emotion-based analysis and classification of audio music**. 2019. 373 f. Tese (Doutorado) - Curso de Engenharia Informática, Universidade de Coimbra, Coimbra, 2019. Available at: https://hdl.handle.net/10316/87618. Accessed on: 04 Oct. 2025.

RAJESH, S.; NALINI, N. J. Musical instrument emotion recognition using deep recurrent neural network. **Procedia Computer Science**, [*s. l.*], v. 167, p. 16-25, 2020. Available at: https://doi.org/10.1016/j.procs.2020.03.178. Accessed on: 11 July 2025.

RUSSO, M. et al. Cochleogram-based approach for detecting perceived emotions in music. **Information Processing & Management**, v. 57, n. 5, p. 1-17, Sept. 2020. Available at: https://doi.org/10.1016/j.ipm.2020.102270. Accessed on: 11 July 2025.

YANG, D.; LEE, W.S. Music emotion identification from lyrics. *In*: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA, 11., 2009, San Diego. **Proceedings** [...]. San Diego: IEEE, 2009. p. 624-629. Available at: https://doi.org/10.1109/ISM.2009.123. Accessed on: 11 July 2025.

YANG, X.; DONG, Y.; LI, J. Review of data features-based music emotion recognition methods. **Multimedia Systems**, [*s. l.*], v. 24, n. 4, p. 365-389, Aug. 2017. Available at: https://link.springer.com/article/10.1007/s00530-017-0559-4. Accessed on: 11 July 2025.

NOTES

AUTHORSHIP CONTRIBUTIONS

Manuscript conceptualization and preparation: P.S.C. Moreira
Data Collection: P.S.C. Moreira
Data Analysis: P.S.C. Moreira, D.F. Tsunoda, M. Nunes-Silva
Discussion of results: P.S.C. Moreira, D.F. Tsunoda, M. Nunes-Silva
Review and approval: P.S.C. Moreira, D.F. Tsunoda, M. Nunes-Silva

PREPRINTS

The manuscript is not a preprint.

ACKNOWLEDGMENTS

The authors would also like to thank the Academic Publishing Advisory Center (Centro de Assessoria de Publicação Acadêmica, CAPA – http://www.capa.ufpr.br) of the Federal University of Paraná (UFPR) for assistance with English language translation and developmental editing.

FUNDING
This study was partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001.
This study was partly financed by the Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG - CHE - APQ-01915-18).

CONFLICT OF INTEREST
The authors declare that there are no conflicting interests.

AVAILABILITY OF RESEARCH DATA AND OTHER MATERIALS
The data is already available in reliable data repositories:
Title 1: LIRA – Intermodal Language for Affective Recognition
URL: https://data.mendeley.com/datasets/9zdww6wnyx/3

PUBLISHER
Universidade Federal de Santa Catarina. The ideas expressed in this article are the responsibility of the authors and do not necessarily represent the opinion of the editors or the university.

HISTORY
Received on: 17-07-2025 – Approved on: 29-09- 2025 – Published on: 27-10-2025