Sound Signal Analysis with Understanding of Musical Data

Özlem Kılıçkıran
Department of Information Systems and Technologies
University of Bilkent
Çankaya, Ankara
ozlem.kilickiran@ug.bilkent.edu.tr

Abstract*— This paper examines basic sound visualizations, such as waveforms and spectrograms, to understand sound data. The study comprises two parts: In Part I, fundamental concepts of sound processing are introduced on sound files along with descriptions of basic audio features. After providing basic information on sound analysis and audio features, Part II conducts a musical data analysis of 240 song tracks from 15 Depeche Mode albums. The paper aims to create an understanding in sound and music analysis through basic sound processing techniques, illustrated by an example of musical data analysis.

Keywords—sound, waveform, spectrogram, musical data, digital signal processing, musical data, audio signals, beat, tempo.

I. INTRODUCTION

In this modern world we are surrounded by all kinds of signals in various forms. Audio signals are the representation of sound. How do we see audio signals and then make a meaningful understanding of music that is comprised of audio signals?

Some of the signals are natural, but most of the signals are handmade in an engineering context, signals are carriers of information, both useful and unwanted. Therefore, extracting or enhancing useful information from a mix of conflicting information is the simplest form of signal processing [7].

In audio technology, innovations like speech recognition, 360-degree audio, and wireless audio have become pivotal forces at the forefront of the industry. Applications like Shazam uses techniques audio signal processing techniques [4]. Therefore, understanding audio and musical data is important.

II. HYPOTHESES

H1: Basic sound analysis visualizations, such as waveforms and spectrograms, can help visualize and understand the sound data.

H2: Spectrograms are effective visualizations for demonstrating sound.

H3: Analyzing musical data on top of sound analysis, with a description of basic audio features, will contribute to the overall understanding of sound and music analysis performance.

III. METHODOLOGY

This project comprises two parts to analyze audio signals and musical data. All graph visualizations will be done using R. In Part I, sound analysis visualizations, employing basic sound representation techniques such as waveforms and spectrograms, will aid in understanding the examined sound data. In Part II, alongside basic sound analysis explanations, visualizations, and descriptions of basic audio components, a musical data analysis will be performed with visualizations using 15 albums of Depeche Mode, following Spotify's defined audio features for sound analysis performance. Both parts will utilize selected sound and music files: river and nature sound in WAV and MP3 formats for Part I, and 15 albums comprising 240 songs extracted from remixes, mixes, and live performances for Part II.

PART ONE: BASIC SOUND ANALYSIS WITH R AND THE IMPACT OF HRTF ON SOUND PERCEPTION

Sound waves can be analyzed in terms of their *amplitude* and *frequency*. The loudness of a sound corresponds to the *amplitude* of the wave and is measured in *decibels*. The *frequency* of a sound wave affects the *pitch* of the sound we hear [10].

IV. DESCRIPTIVE ANALYSIS

TABLE I. REPRESENTATIONS OF SOUND SIGNALS

Representation	*Attributes*	*Brief Description*
Waveform	Time and Amplitude	Depicts the amplitude variations of a signal over time.
Spectrogram	Time, Frequency, and Amplitude	Illustrates how the frequency of a signal changes over time.

Information about the "river.wav" audio file that will be used for sound representation using tuneR library in R:

> summary(riversound)
Wave Object
Number of Samples: 333600
Duration (seconds): 6.95
Samplingrate (Hertz): 48000
Channels (Mono/Stereo): Stereo
PCM (integer format): TRUE
Bit (8/16/24/32/64): 16

Summary statistics for channel(s):

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
left	-27033	-1735	0	-0.521289	1746	27032
right	-25515	-1305	0	-0.415045	1305	19446

Information about the "naturesound-before.wav" and "naturesound-after.wav" audio files that will be used for sound representation using tuneR library in R:

> summary(naturesound-before)
Wave Object
Number of Samples: 345828
Duration (seconds): 7.2
Samplingrate (Hertz): 48000
Channels (Mono/Stereo): Stereo
PCM (integer format): TRUE
Bit (8/16/24/32/64): 16

Summary statistics for channel(s):

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
left	-32768	-309	0	2.114407	311	32767
right	-32768	-309	0	2.114407	311	32767

> summary(naturesound-after)
Wave Object
Number of Samples: 345828
Duration (seconds): 7.2
Samplingrate (Hertz): 48000
Channels (Mono/Stereo): Stereo
PCM (integer format): TRUE
Bit (8/16/24/32/64): 16

Summary statistics for channel(s):

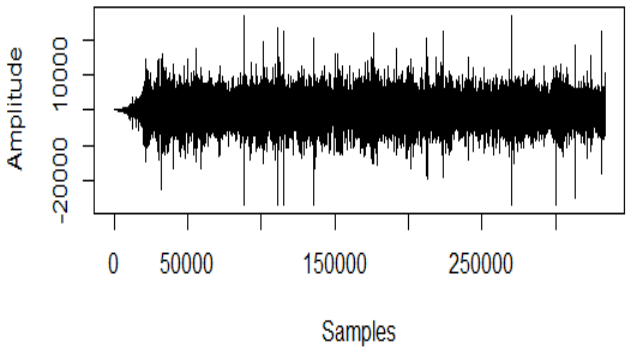
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
left	-32768	-131	0	1.71426	133	32767
right	-32768	-131	0	1.71426	133	32767

The sound files represent a stereo audio file with 2880512 sample points in each channel.
The sample rate is 48000. Each bit depth is16 bits and their durations differ.

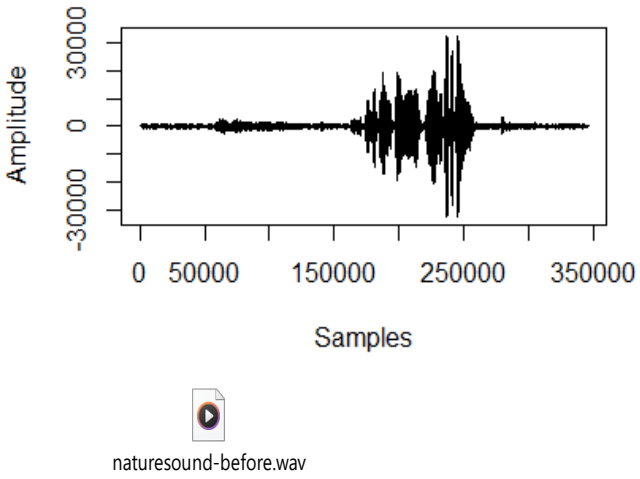
In this part, the chosen sound file will be analyzed with the written representations in the above Table I using R.

A. Plotting the Waveform

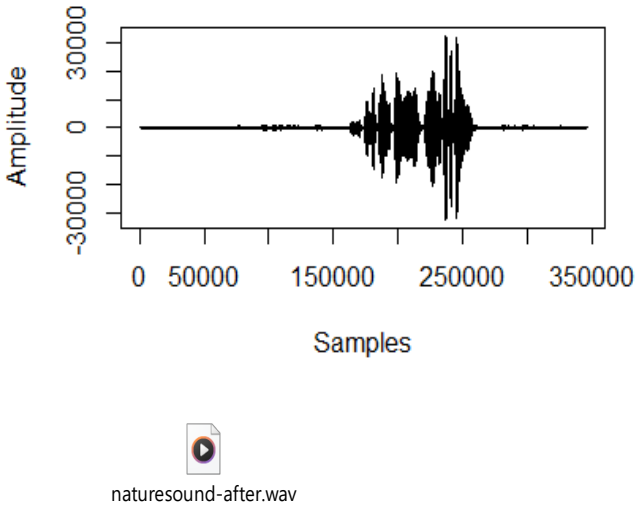
A spectrogram allows to 'see' sound, which helps to quickly review large datasets or find patterns that we don't or can't hear. Spectrograms keep time on the X axis but place frequency on the Y axis. Amplitude is also represented as a sort of heat map or scale of color saturation [3].
Waveforms serve as valuable tools for conveying fundamental information about a recording, such as identifying loud and quiet sections and gauging the overall dynamic range. For instance, when listening to an interview, a waveform provides a clear visualization of when someone is speaking. However, waveforms have limitations in revealing details about pitch, frequency, or harmonic content [13].



Waveform 1: waveform of the river.wav file



Waveform 2: waveform of the "naturesound-before" file

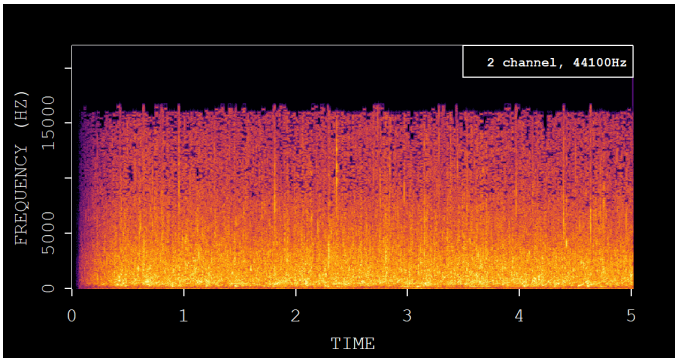


Waveform3: waveform of the "naturesound-after" file

B. Fourier Transforms and Spectrograms and Plotting the Spectrogram

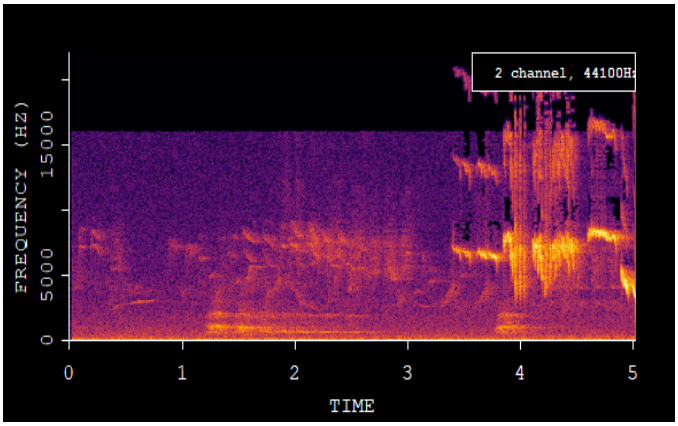
When working with signals, it is often considered whether t he project needs to work in the *time domain* or the *frequency domain*. These two different domains can provide valuable information about a signal. The time-domain signal represent s the amplitude change of the signal over time [19]. The most popular way to explore signals in the frequency domain is to use Fourier Transforms. A *Fourier Transform* is a reversibl e mathematical transform developed by Joseph Fourier in th e early 1800s [6].

Raw data of sound might not be very informative. Therefore, most analyses require that the signal is transformed into frequency data [11].
In R, the read_audio_fft function converts audio directly into frequency data using FFmpeg's built-in FFT. It returns a matrix with the time-frequency data, from which we can plot the spectrogram [20].

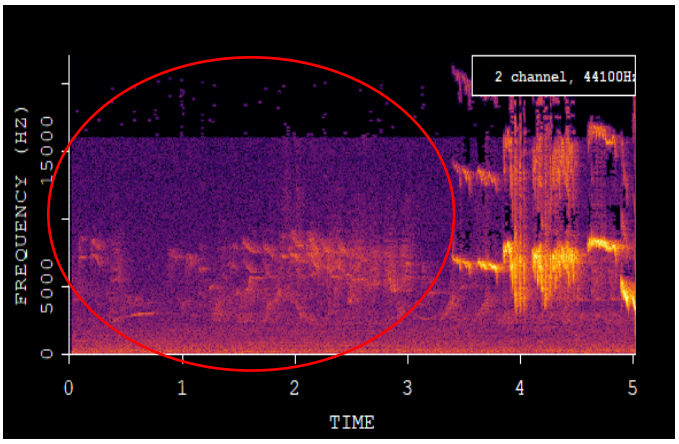


Spectrogram 1: spectrogram of the river.mp3 file's first 5 second

For a better understanding of what does spectrogram do and what kind of information one be gathered by looking at a spectrogram, two audio files will be used, and their spectrograms will be plotted, which first one is a nature sound including bird whistling, dog barking and an environment noise. The second one is a sound that is compromised of the same sounds except the dog barking sound [16].



Spectrogram 2: naturesound-before.mp3 spectrogram's first 5 second (including dog barking)



Spectrogram 3: naturesound-after.mp3 spectrogram (excluding dog barking)

Various methods can be used to examine audio signals, such as waveforms and spectrograms, using R. Upon analyzing the waveforms of three different sound files, the signals in Waveform 1 appear relatively regular over time. Waveform 2 differs from Waveform 3, notably at the beginning of the first 150,000 samples and at the end of the last 100,000 samples. This difference is attributed to the presence of a dog barking sound.

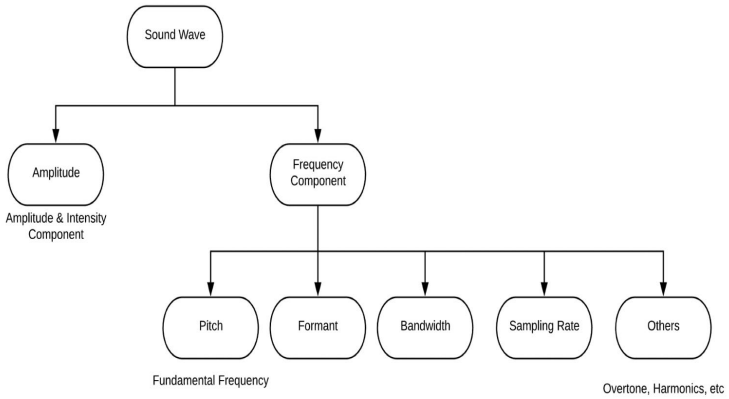
Upon closer inspection of the spectrograms, the red-circled shaped object parts highlight the dog barking effect on the spectrogram.

PART TWO: MUSICAL SOUND ANALYSIS AND REPRESENTATION

TABLE II. BASIC COMPONENTS OF SOUND

Feature	***Description***
Tempo	The speed or pace at which a musical piece is performed.
Beat	Regular, recurring pulse in music that establishes the underlying rhythm.
Rhythm	The pattern of sounds and silences in music, organized in relation to time.
Tone	Refers to the perceived frequency of a sound in music.

The musical aspects of tempo, beat, rhythm and tone play a fundamental role for the understanding of and the interaction with music [2]



Some Basic Frequency Components

Literature provides a variety of frequency components, for the purpose of this article we will talk about the pitch, sampling rate, format, and bandwidth.

Sample rate (or sampling frequency) is the number of samples per second in a Sound. For example: if the sampling rate is 4000 hertz, a recording with a duration of 5 seconds will contain 20,000 samples.

Pitch is the frequency of the fundamental component in the sound, that is, the frequency with which the waveform repeats itself. Pitch depends on the frequency of a sound wave. Frequency is the number of wavelengths that fit into one unit of time.

Formant is a concentration of acoustic energy around a particular frequency in the speech wave. Thus, they are the peaks that are observed in the spectrum envelope.

Bandwidth is the range of frequencies within a given band, that used for transmitting a signal [23].

MUSICAL DATA ANALYSIS ON DEPECHE MODE ALBUMS

This part of this paper aims to give an insightful analysis using Spotifyr package of R. Spotifyr is an R wrapper for pulling track audio features and other information from Spotify's Web API.

	Album Name	Number of Tracks
1	A Broken Frame (Deluxe)	13
2	Black Celebration (Deluxe)	19
3	Construction Time Again (Deluxe)	16
4	Delta Machine	17
5	Exciter (Deluxe)	17
6	Memento Mori	12
7	Music for the Masses (Deluxe)	16
8	Playing the Angel (Deluxe)	16
9	Some Great Reward (Deluxe)	11
10	Songs of Faith and Devotion (Deluxe)	11
11	Sounds of the Universe (Deluxe)	33
12	Speak And Spell (Deluxe)	15
13	Spirit (Deluxe)	14
14	Ultra (Deluxe)	17
15	Violator (Deluxe)	13

Fig. 1.

Above 15 albums, in total 240 songs, will be analyzed according to their musical data. Spotify defines certain audio features for music analysis as follows: *acousticness danceability, energy, instrumentalness, liveness, speechiness and valence.*
A full list of these, along with their verbal definitions, can be found on Spotify's page for developers. The Spotifyr package reduces the process of pulling data for these audio features from Spotify's web API to just a few lines of code.
For this paper, especially, the valence value of Spotify and tempo will be examined.

A. Valence

Spotify uses the word "valence" to measure whether a song is likely to make someone feel happy (higher valence) or sad (lower valence). The metric is measured on a scale from 0.0 to 1.0. [10].

track_name	album_name	valence
What's Your Name?	Speak And Spell (Deluxe)	0.990
(Set Me Free) Remotivate Me	Some Great Reward (Deluxe)	0.983
Told You So	Construction Time Again (Deluxe)	0.969
Something to Do	Some Great Reward (Deluxe)	0.966
Pleasure, Little Treasure	Music for the Masses (Deluxe)	0.963

Fig. 2. The top 5 valence values of Depeche Mode tracks gathered from chosen 15 albums.

The lowest valence value belongs to the track named What's Your Name with the valence value of 0.999.

The following visualization shows the valence values according to albums.

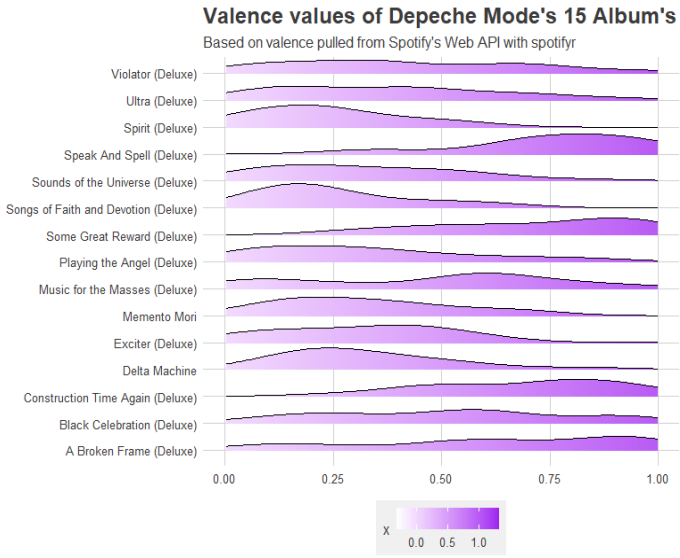


Fig. 3. Valance visualization of Depeche Mode albums.

	valence	album_name	track_name
1	0.0327	Exciter (Deluxe)	Easy Tiger

Upon looking at both the Valance visualization of each song in 15 albums and 240 songs, and according to Spotify's explanation of valance feature that is implemented using R, the saddest song is the track named Easy Tiger from Exciter album.

After finding the saddest track, visualizing a joyplot might be beneficial.

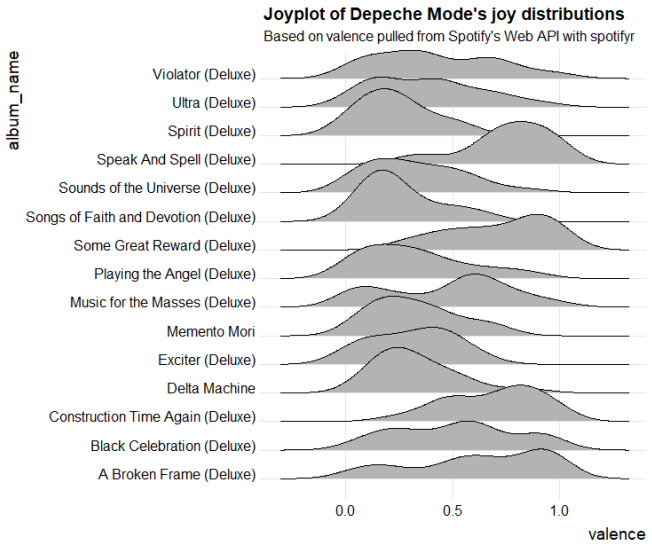


Fig. 4.

According to the joyplot, with the calculations for average valence of each album made in R, identifying the album with the minimum average valence is Spirit.

	album_name	avg_valence
	<chr>	<dbl>
1	Spirit (Deluxe)	0.23

This suggest that according to the means of each album's vale nce, Spirit is the saddest one with the 0.23 average valence val ue, while Speak And Spell is the happiest with the average val ence 0.75.

album_name	mean(valence)
Speak And Spell (Deluxe)	0.7563333
Some Great Reward (Deluxe)	0.7135455
Construction Time Again (Deluxe)	0.6866875
A Broken Frame (Deluxe)	0.6197462
Black Celebration (Deluxe)	0.5136842
Music for the Masses (Deluxe)	0.4694187
Violator (Deluxe)	0.4210615
Ultra (Deluxe)	0.3776588
Delta Machine	0.3286471
Playing the Angel (Deluxe)	0.3268250
Sounds of the Universe (Deluxe)	0.3251333
Memento Mori	0.3237833
Exciter (Deluxe)	0.3109765
Songs of Faith and Devotion (Deluxe)	0.2571455
Spirit (Deluxe)	0.2330643

Fig. 5. Means of valences of 15 albums of Depeche Mode.

B. TEMPO

The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration [11].

After checking the tempo of each 240 songs from 15 albums, tempo plot shows that there are some songs that are considerably in a low tempo. This visualization demonstrates that some tracks have relatively smaller beats per minute.
Minimum Tempo Track: Condemnation
Minimum Tempo: 58.0
Maximum Tempo Track: One Caress
Maximum Tempo: 200.909

Creating the tempo plot, by checking the standard deviation of the tempo. If it's 0, this means that all the values are identical . The bigger it is, the more the values vary. If we filter to track s with SD greater than 0 (so any variation at all), we see that m ost tracks have a little variation. However, if we filter to tracks with an SD greater than 1, we see a few songs with slightly di fferent tempo, and a few with wildly different tempo [11].

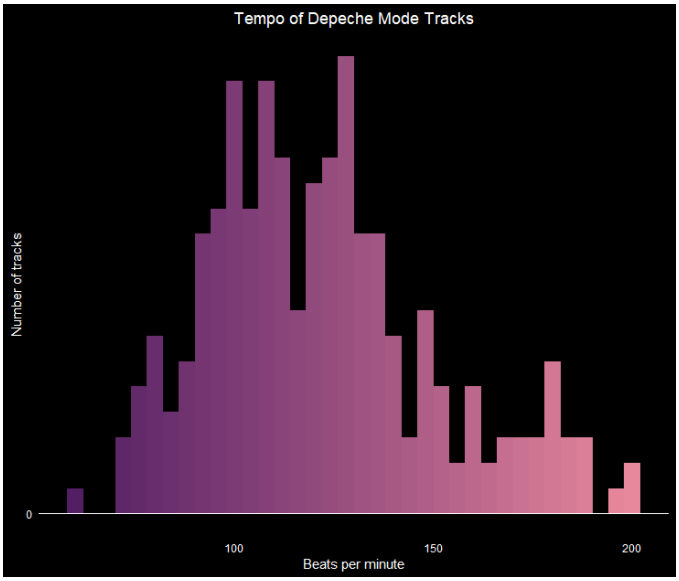


Fig. 6. The plot that is showing BPM of 240 tracks.

As it is mentioned before, each track has several audio feature or ratings: danceability, energy, speechiness, acousticness, inst rumentalness, liveness, and valence according to Spotify. Plea se, check the meanings of these audio features [10].

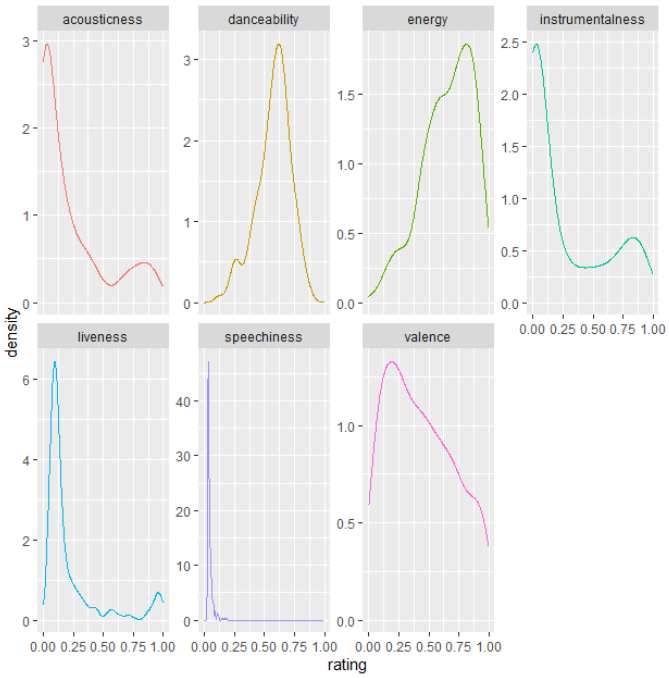


Fig. 7. Spotify's audio features: acousticness danceability, energy, Instrumentalness, liveness, speechiness and valence ratings for the 15 albums of Depeche Mode as follows from the starting to end:

Fig.7. shows these audio features ratings for the 15 albums of Depeche Mode in the following order:

Memento Mori Spirit (Deluxe)
Delta Machine Sounds of the Universe (Deluxe)
Playing the Angel (Deluxe)
Exciter (Deluxe)
Ultra (Deluxe)
Songs of Faith and Devotion (Deluxe)
Violator (Deluxe)
Music for the Masses (Deluxe)
Black Celebration (Deluxe)
Some Great Reward (Deluxe)
Construction Time Again (Deluxe)
A Broken Frame (Deluxe)
Speak And Spell (Deluxe)

IV. CONCLUSION

Part I one of the papers enabled a general understanding of description of sound features with sound analysis using basic audio signal representation techniques implemented in R. Spectrogram, Part II begins giving information of most important aspects of musical data followed by a musical data analysis on top of basic sound analysis made in Part I. In Part II, a comprehensive analysis of music is done by analyzing the 15 albums of Depeche Mode using Spotify API to examine various musical aspects that are predefined by Spotify for musical data analysis works of albums belonging to same band, Depeche Mode. Both Graphs for Part I and Part II visualized using RStudio.

- By comparing Spectrogram 2 and Spectrogram 3, one can observe that a spectrogram enables the identification and removal of unwanted noises from the sound file, since spectrograms demonstrate various patterns within a sound file, allowing for the selective elimination of undesired elements.
- It is recognized that mp3 format audio files are better for plotting spectrograms, specifically for the spectrograms, as sounds significantly appear clearer with mp3 format than wav format.
- According to Spotify's predefined music analysis, each track has several ratings: danceability, energy, speechiness, acousticness, instrumentalness, liveness, and valence.
  In that context, 15 albums of Depeche Mode and 240 songs were analyzed with different visualizations. Using Spotify's valence value and joyplot, the saddest song is determined as "Easy Tiger" from the "Exciter" album. However, the saddest album obtained from the results of the joyplot, by calculating the average valence value for each album, is determined to be the "Spirit" album. This showcases that the saddest song does not belong to the

saddest album of Depeche Mode, according to the algorithm of valence value specified by Spotify.

- "Easy Tiger" from the "Exciter" album is chosen as the saddest track from the 15 albums and 240 songs. After listening, it could be said that Spotify's valence feature mostly relies on the mood and tempo of the song, as lyrics were not used as a feature for this paper. Defining the saddest song by looking at joyplot and "valence" factors can be deceptive, as the lyrics behind a song immensely define the overall sadness/happiness of a song. Relying solely on the feature of valence might not be one hundred percent true for some deeper analysis. For a better understanding of the sadness of a song, lyrics analysis might also be done, as it is done in other searches using the Spotify API.

REFERENCES

[1] Cvetkovic, James Johnston, and Julius O Smith III. IEEE Signal Processing Magazine 34(3).

[2] Perceptual Spatial Audio Recording, Simulation, and Rendering: An overview of spatial-audio techniques based on psychoacoustics. Huseyin Hacihabiboglu, Enzo De Sena, Zoran.

[3] J. Hansen and T. Johnson, "Spectrograms in R," HansenJohnson.org, Available: https://hansenjohnson.org/post/spectrograms-in-r/. Accessed on 12 30 2023.

[4] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 6. Institute of Electrical and Electronics Engineers (IEEE), pp. 1088–1110, Oct. 2011. doi: 10.1109/jstsp.2011.2112333.

[5] P. M. Grosche, "Signal processing methods for beat tracking, music segmentation, and audio retrieval." Universität des Saarlandes, Apr. 04, 2013. doi: 10.22028/D291-26471.

[6] J. W. Doser, A. O. Finley, and Vince, "R Programming for Data Sciences." Michigan State University, MI, USA, 2022.

[7] V. K. Ingle and J. G. Proakis, Digital Signal Processing Using Matlab: A Problem-Solving Companion. Cengage Learning, 2016.

[8] Toptal, "Shazam It! Music Processing, Fingerprinting, and Recognition" Toptal Blog. https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition (accessed Jan. 10, 2024).

[9] R. J. Hijmans, R. Bivand, K. Forner, J. Ooms, and E. Pebesma, "Spatial data analysis," R package version, 2022.

[10] Spotify. "Get Audio Features," Spotify for Developers. [Online]. Available: https://developer.spotify.com/documentation/web-api/reference/get-audio-features. [Accessed: January 8, 2024].

[11] E. Nordmann and L. DeBruine, Applied Data Skills. Zenodo, 2022. doi: 10.5281/ZENODO.6365077.

[12] M. Müller, Fundamentals of Music Processing: Using Python and Jupyter Notebooks, 2nd ed. Springer, 2021. ISBN: 978-3-030-69807-2.

[13] Cheryl Tipp, "Seeing Sound: What is a Spectrogram?" https://blogs.bl.uk/sound-and-vision/2018/09/seeing sound-what-is-a-spectrogram.html (accessed Sep. 19, 2018).

[14] B. Wundervald and J. Trecenti, "Music Data Analysis in R" Unpublished, 2019, doi: 10.13140/RG.2.2.16222.48966.

[15] T. O. Otuokere, A. L. Imoize, and A. A.-A. Atayero, "Analysis of Sonic Effects of Music from a Comprehensive Datasets on Audio Features," ELEKTRIKA- Journal of Electrical Engineering, vol. 20, no. 1. Penerbit UTM Press, pp. 43–53, Apr. 30, 2021. doi:10.11113/elektrika.v20n1.23