LOW-LEVEL FEATURES AND TIMBRE CHARACTERIZATION

2ND LAB ASSIGNMENT

♪ **André de Azevedo Barata**
Dep. of Electrical and Computer Engineering
Faculty of Engineering of the University of Porto
Porto, Portugal
up201907705@up.pt

♪ **Marija Jocic**
Dep. of Electrical and Computer Engineering
Faculty of Engineering of the University of Porto
Porto, Portugal
up202202484@up.pt

November, 2023

Keywords Audio classification, Instrument recognition, Audio Processing, Machine Learning

1 Introduction

This report explores the development of an audio classification system for identifying both instrument excitation (percussive vs. non-percussive) and instrument type from isolated musical instrument note samples. It employs low-level audio descriptors, such as RMS/energy, zero crossing rate, and spectral centroid, to capture the acoustic properties of the sounds. The distribution of these descriptors is analyzed to understand their relationship with the classification task. Classification models are implemented and evaluated using relevant metrics, demonstrating the effectiveness of the proposed approach in musical instrument recognition.

2 Sound Descriptors

Sound descriptors play a crucial role in the analysis and manipulation of audio signals. These numerical or categorical representations capture various aspects of a sound's characteristics, including its pitch, loudness, timbre, and temporal structure. In this section, we measure several descriptors related to audio signals, including: RMS, Energy, Zero Crossing Rate (ZCR), Log-attack time (LAT), Temporal centroid (TC), Effective duration (RF), Spectral centroid (SC), Spectral spread (SS), Spectral variation (SV), Spectral flatness (SF). In table 2 we can find the results of this descriptors for three cases: sine was of 100 Hz, sine wave of 1250 Hz and White Gaussian Noise (WGN). The generated signals are visible on figure 1.

Descriptor		Sinusoidal 100 Hz	Sinusoidal 1250 Hz	WGN
Time-domain	ZCR (Hz)	200	2500	22620
	Energy (dB)	1100	88.0	2216
	RMS (dB)	0.707	0.707	1.00
	LAT (s)	2.27e–5	2.27e-5	9.09e-5
	TC (s)	0.025	2e-3	1.24e-2
	ED (s)	1.00	1.00	1.00
Freq-domain	SC (Hz)	10.50	7.745	554.5
	SS (Hz)	57.0	12.77	317.6
	SV (Hz)	5.43	1.46	0.573
	SF (dB/Hz)	1.23	1.15e-3	7.74e-4
	SD (dB/Hz)	6.26	8.05e-4	-4.11e-4

Table 1: Descriptors of 100 Hz and 1250 Hz Sine waves and WGN

Regarding Zero Crossing rate, it's expected that this rate is the double of the frequency in sine waves. When considering WGN, the value should be much higher and variable.

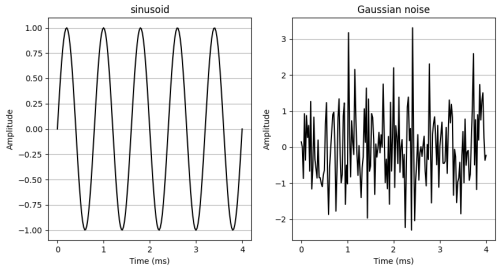

Figure 1: Sine Wave (left) and WGN (Right)

3 Data Analysis

Table 2 has the descriptors mention in 2 applied for three cases: a flute and two different guitars. The chosen audios for this experiment were: "flt_G4_12", "gui_a_G2_12" and "gui_b_G3_12".

	Descriptor	Flute	Guitar A	Guitar B
Time-domain	ZCR (HZ)	0.0402	0.0307	0.0263
	Energy (dB)	1.14e5	1.98e3	3.32e3
	RMS (dB)	0.509	0.0924	0.112
	LAT (s)	0.0.332	0.276	0.304
	TC (s)	3.57	0.497	0.537
	RD (s)	1.00 s	1.00	1.00
Freq-domain	SC (Hz)	20.2e3	6.08e3	4.34e3
	SS (Hz)	30.0e3	12.1e3	7.94e3
	SV (dB/Hz)	1.48	1.99	1.83
	SF (dB/Hz)	8.26e-7	5.54e-7	2.40e-7
	SD (dB/Hz)	-3.81e-5	-1.77e-5	-1.48e-6

Table 2: Descriptors for Flute and two Guitars

The descriptors present in table 2 are enough for percussive vs. non-percussive classification, but not for instrument recognition. Percussive sounds, like drums and cymbals, have high energy and a short duration, while non-percussive sounds, like piano and guitar, have lower energy and a longer duration. As this descriptors cover this characteristics, they are enough to accomplish both classifications.

4 Applications

This descriptors could be used for Sound Event detection purposes. We could identify and categorise various sounds in an audio recording. This task that involves detecting and classifying sounds in various settings, including noisy environments and those with overlapping sounds. Sound Event Detection has a wide range of applications, such as environmental monitoring, audio surveillance, audio tagging, music information retrieval, and human-computer interaction.

5 Classification - Percussive / Non-percussive

The binary classification of the instruments was done by using two descriptors, which were attack time and effective duration. This was considered the best option since the attack time and duration of percussive sounds are smaller then of non-percussive. The provided dataset had percussive sounds noted with "pizz" so the evaluation could have been done easily. The obtained accuracy was 80%. This was due to the fact that some sounds had similar descriptor values. The improvement would be possible by using more descriptors or by using another combination of descriptors. In the file "instruments_characterisation" attached to the report you can find the implementation of the used algorithm.

6 Classification – Instrument Recognition

Classification of the instruments by their type was done by using the same descriptors as in section 5 with using one additional descriptor called Spectral Centroid. This technique is important because it provides information about the "center of mass" or the average frequency of a sound signal. This was considered important since the classification had to be done for 12 different instruments. With the use of these three descriptors the resulting accuracy was 57%. Since the dataset is small and the classification of the musical instruments with using only the descriptors is not an easy task, this accuracy is considered satisfying. The implementation of the used algorithm can be found in file "instruments_characterisation".

7 Classification using Machine Learning

We developed classified the instruments by using Support vector machine - SVM. SVM is a powerful supervised algorithm that works best on smaller datasets. As this is our case, since we only have 66 audio samples from 12 different instruments (on average, less than 6 audios for each instrument!). As so, we used the descriptor defines in section 2 as features for the SVM and the instrument name as labels.The objective is to predict the instrument name by the evaluation of it's descriptors. The results of this algorithm would vary due to the small dataset. In fact, we used a random separation of the data to be used in train and test, therefor, as the data set is already small, the result are highly dependent on the ones that are used in the test. We got a maximum accuracy of around 73% and a minimum of around 21%. In the file "instruments_characterisation" attached to file report you can find implementation of the SVM algorithm used.

In order to predict if a sound is or not percussive, we could use the same algorithm but with only two labels. This labels were assign in section 5.

8 Conclusion

To summarise, this work has let us explore audio classification by using different descriptors and also machine learning method. It lead us to conclusion that descriptors are a good enough tool when it comes to binary classification of audio sounds, but that machine learning methods still provide better results for multi-class type of classification.

References