

3. Data Exploration and Storytelling:

After the data wrangling step, all data are stored in a python dictionary with keys as file IDs and values as the discrete sound channel signal. In our dataset, we have 5435 sound files categorized into ten different types or classes. In this chapter, we will choose a sample size of one file for each class to explore and visualize the sound signal within. Randomly chosen sample files from each class were 915 (dog_bark), 1873 (jackhammer), 5678 (air_conditioner), 5289 (street_music), 6729 (drilling), 522 (children_playing), 7046 (siren), 6668 (car_horn), 2099 (engine_idling), 4171 (gun_shot).

Using spectrograms and wave plots, we will visualize and explore data features that could help in classifying sound signals. We'll start with signal amplitudes in section 3.1, followed by signal power or energy analysis in section 3.2.

We will explore some feature extraction techniques of audio signal data. The purpose of this step is to explore the data, generate interesting features and insights that should help in the classification part of the project.

Extracting features requires us to operate on all files, however in this report we will visualize a subsample of the feature extraction results from the different classes. We will visualize, hypothesize and summarize the plots after.

3.1. Wave plots and Spectrograms

A wave plot is defined as a plot of the amplitude of the sound signal versus time. A spectrogram, however, is a plot of the signal's frequencies versus time. Independent axes represent time and frequency while color represents the signal's amplitude of the observed frequency at a certain time. The higher the color intensity, the stronger frequency it is present in the signal. This could be a very important feature in classifying different types of sounds. To plot a spectrogram, we need to compute the Discrete Fourier Transform of the signal to find the frequencies. Since, our signal is a time series, a window frame of a pre-specified size is used for the Fourier Transform. In our experiments, we started with the default window size of 93ms. The resulting spectrogram 2D matrix has frames as rows, frequencies as columns and matrix values are amplitudes in decibels.

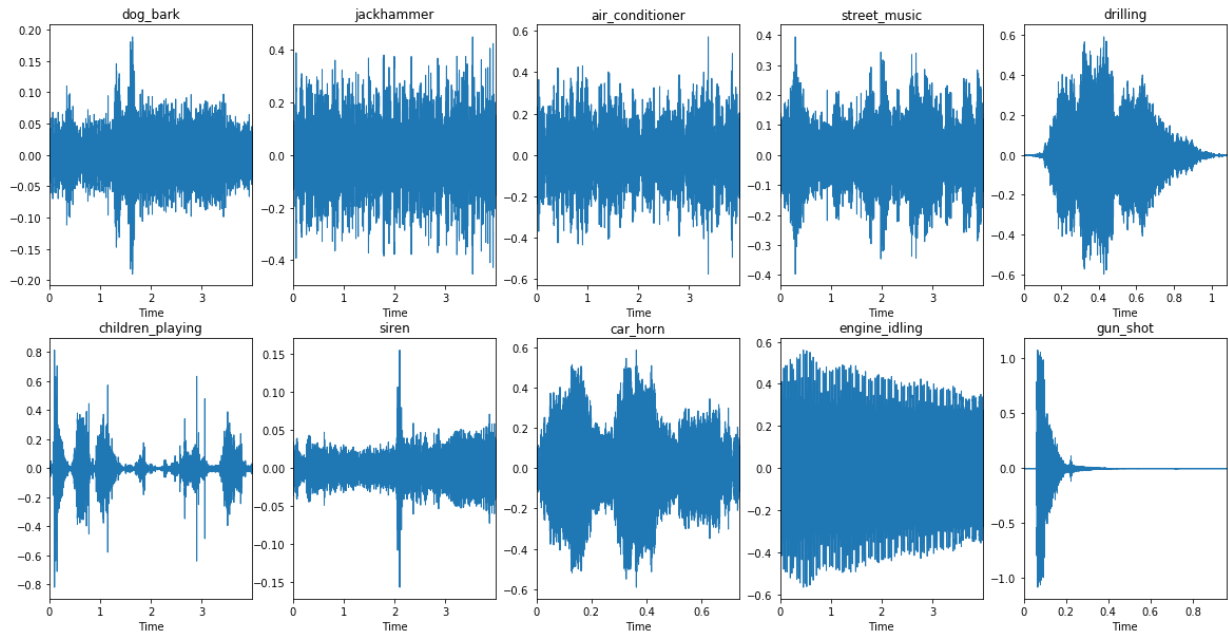


Figure 1: Wave plots of chosen samples

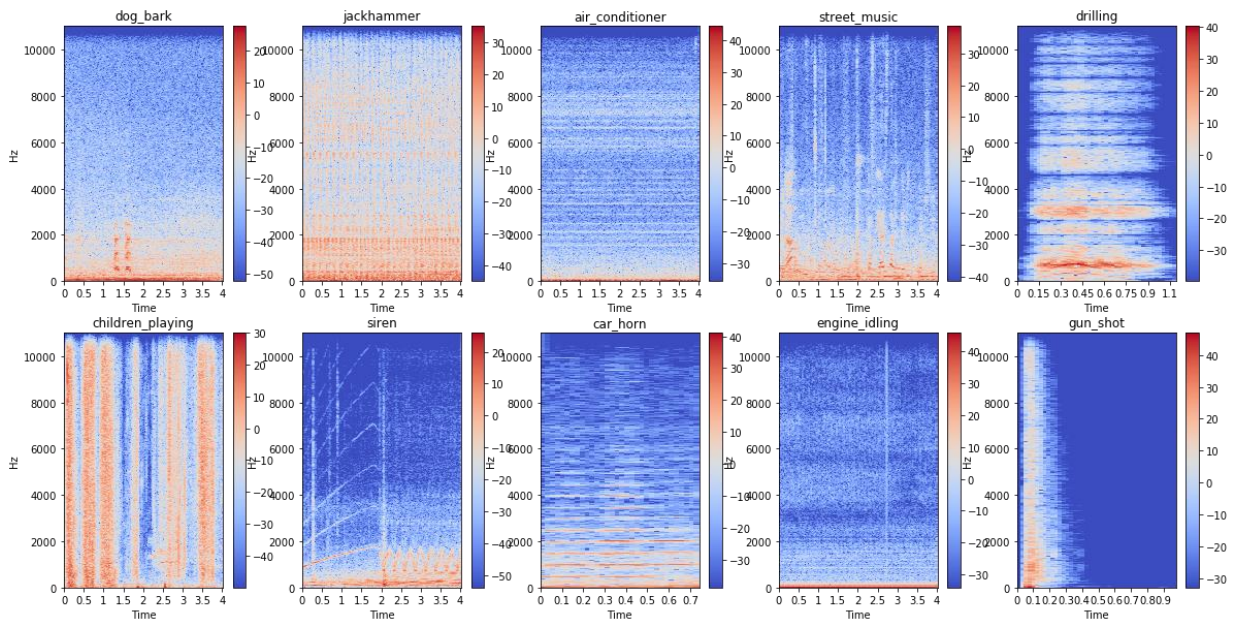


Figure 2: Spectrograms of chosen samples

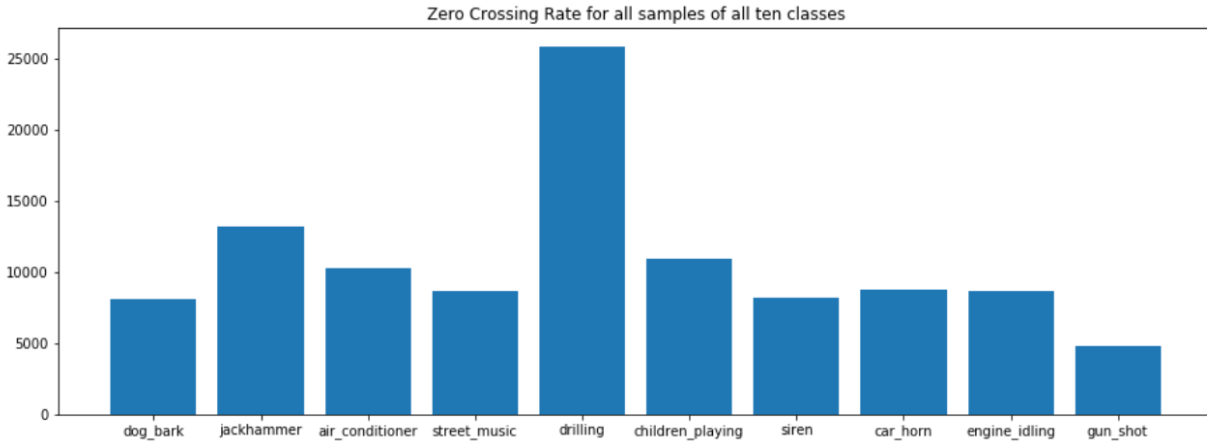


Figure 3: Mean Zero Crossing Rate

Zooming in on a signal's wave plot in figure 1, the number of times the signal changes from positive to negative and vice-versa can be quantized as the zero-crossing rate, which is a feature used heavily in sound and music classification. Figure 3 shows the computed mean zero crossing rates for all samples in the dataset. Looking at the wave plots and spectrograms in figures 1 and 2, the "children playing" spectrogram has the highest rate of change i.e. signal variability, while "engine idling" is the least variable. Just looking at the wave plots of "jackhammer", "air_conditioner" and "street_music", one cannot discern any differences. However, their spectrograms tell a different story. The spectrogram of "jackhammer" is almost high and vertical while that of "air_conditioner" is very low. "street_music" is not as steady as either of the former categories but is intermittently louder than "air_conditioner". "drilling" almost has a horizontal spectrogram, unlike "street_music" and "children_playing". The spectrogram of "drilling" is closest to that of "car_horn" and "engine_idling", albeit it is louder than both former. The spectrograms of "gun_shot" and "dog_bark" are very impulse-like as is that of "siren" that has a delayed impulse start but is steady after that, like "jackhammer". Here, impulse-like refers to the brief time duration the frequencies are present in the signal. "dog_bark" has a lot more random noise in its signal over time than "gun_shot".

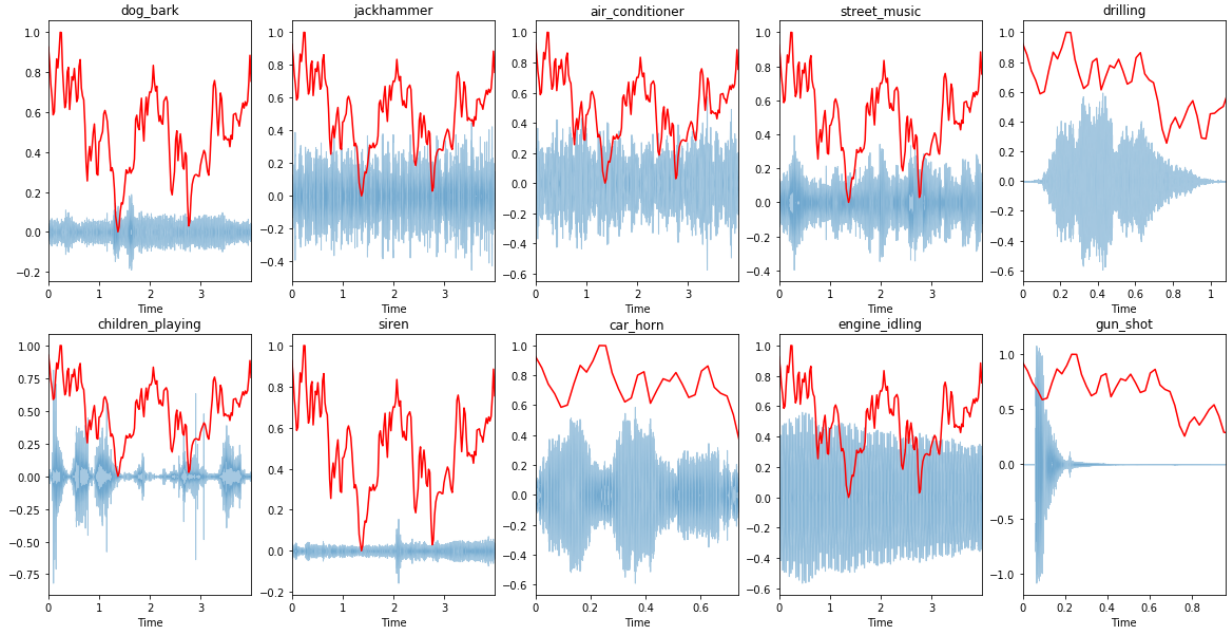


Figure 4: Spectral Centroid of samples

A statistical feature used in sound signal analysis is the spectral centroid. It is defined as the weighted mean of frequencies present in a signal frame. In other words, it could be explained as the center of mass of the sound signal in each time frame. Just looking the plots in figure 4, we can cluster them into two groups, one group includes “car_horn”, “gun_shot” and “drilling” and the other group has the rest.

3.2. Power or Energy Spectrum

A spectral roll-off is defined as the frequency of the signal at which 85% of the spectral energy is bounded.

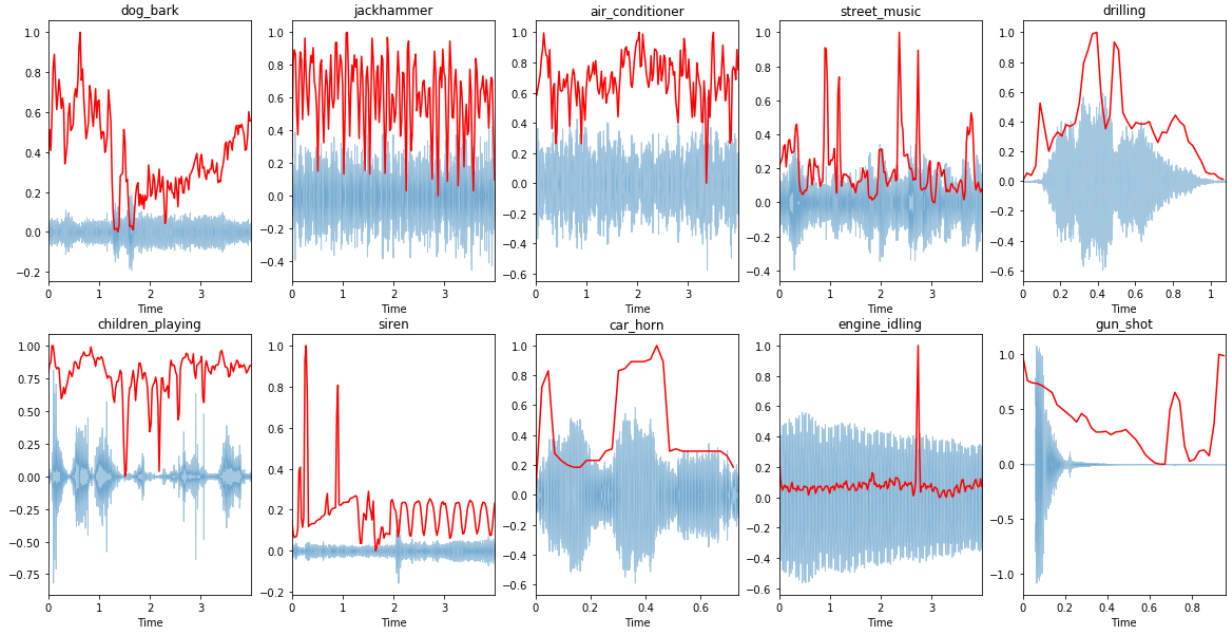


Figure 5: Spectral Roll-off of chosen samples

In figure 5, the red curves represent the computed spectral roll-offs for each time frame for each sample.

The power spectral density of a signal could be represented as a set of discrete coefficients known as the Mel-Frequency Cepstral Coefficients. These coefficients could be used as a feature for sound identification that are mapped to the nonlinear mel-scale that approximates the human auditory system.

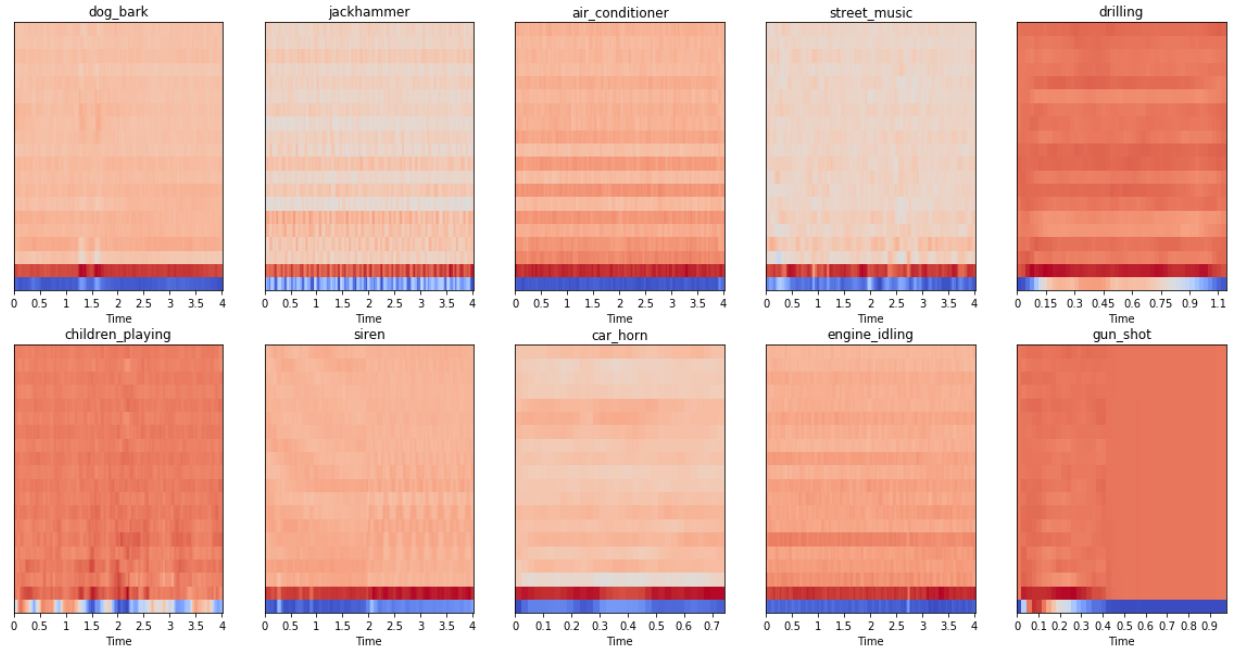


Figure 6: MFCCs for each chosen sample

Figure 6 shows a representation of 20 MFCC features computed for frames of each sample.

3.3. Conclusion:

In this chapter, we have explained and visualized some features most commonly used in sound signal analysis. We computed these features on a sample of signals from our training dataset. By visualizing the sound data, we could discern distinguishable features between sound classes. In the next chapter, we will express these findings in a more statistical framework.