

Audio files for CNN input.

1. I downloaded dog wav files from:
<https://www.kaggle.com/datasets/mmoreaux/audio-cats-and-dogs?resource=download>
2. I also download dog wav and mp3 files from <https://freesound.org>.
3. Susan Smith categorized the sounds into 4 groups: Red, blue, green and uncertain. Red means that the dog was angry; blue means the dog was not angry, but not sad; green means the dog was happy; and uncertain means that the sounds could not be classified into any of the groups.
4. One can visualize sound by using four approaches: waveform, spectrogram, MFCCs, and pitch tracking. See Jupyter notebooks: 'dog_sounds_overview.pdf' Below is the analysis of the four approaches for the same dog sound file.
 - a. The waveform (Figure 1) visually represents the changes in air pressure over time, providing a graphical representation of the sound signal and is useful for offering insights into the characteristics of the sound, including its intensity, frequency, and duration.
 - b. Spectrograms (Figure 2) are created using a mathematical technique called the Short-Time Fourier Transform (STFT). The audio signal is divided into short, overlapping segments, and the Fourier Transform is applied to each segment to analyze its frequency content. By plotting the results over time, you get a detailed representation of how the frequency content of the signal changes.
 - c. MFCCs (Figure 3) provide a compact representation of the spectral features of an audio signal, capturing both the frequency and amplitude information in a way that is more relevant for human auditory perception and less sensitive to noise. They are commonly used as input features for machine learning models in audio processing tasks.
 - d. Pitch tracking (Figure 4) is a process in audio analysis that involves determining the pitch or fundamental frequency of a sound signal over time. The pitch represents the perceived frequency of a periodic sound and is closely related to the concept of musical notes.
5. Any of these approaches can be used as inputs to a CNN. I created a C++ program called "crop_images.cpp" that removes the axes and labels for the spectrograms and MFCC files for input to a CNN. The purpose of the CNN is to determine if sound files could be used to distinguish different dog emotions.
6. Two Pytorch files were created for processing the wave files: "dog_sounds_spectrogram.ipynb" and "dog_sounds_mfcc.ipynb". Both take input files from the downloadfolder and produce image files that are saved in the Desktop. The outputted files are then used as inputs to the CNN.

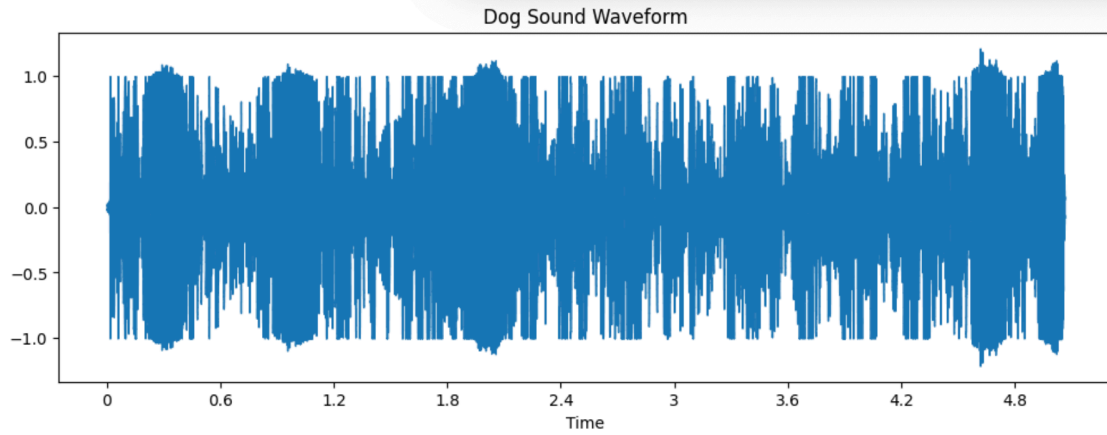


Figure 1. An example of a waveform.

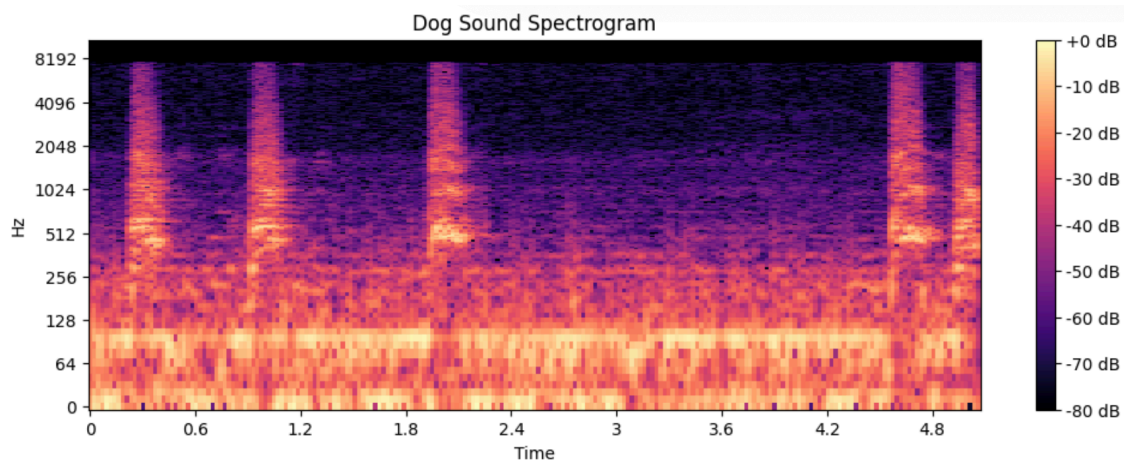


Figure 2. An example of a Spectrogram

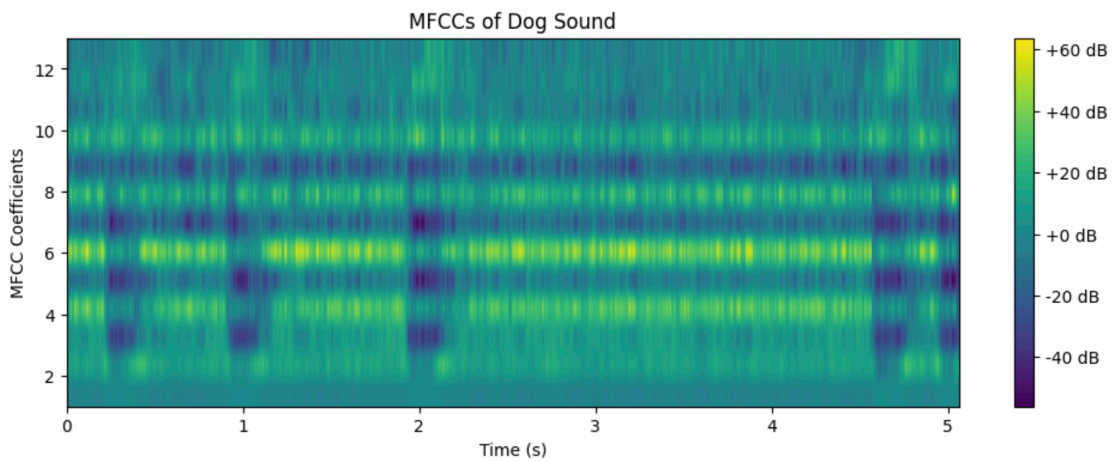


Figure 3. An example of MFCC sound format.

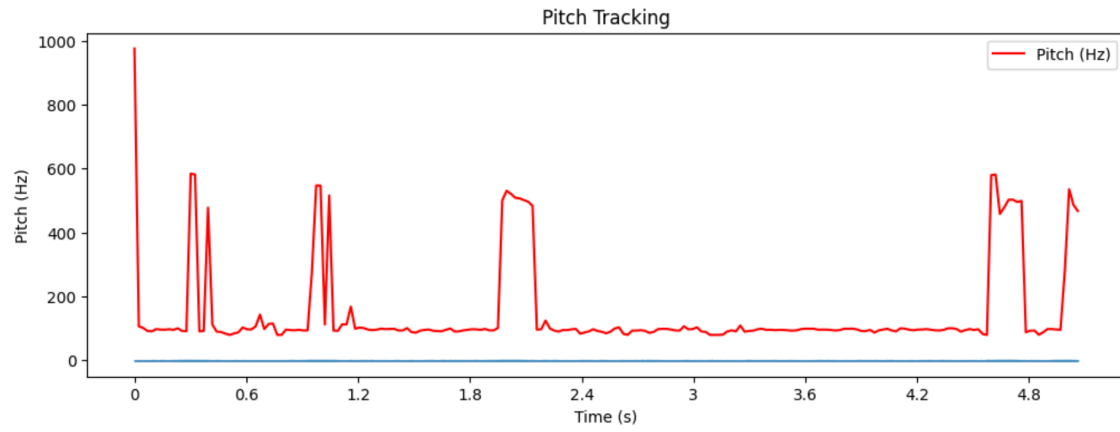


Figure 4. An example of pitch tracking.