

Chapter 1

1.1 Spectrograms

As we saw before, audio can be better represented as mix of different frequencies that define the sound we are hearing. The problem lies in the fact that, in speech, these frequencies change drastically during each conversation, sentence, word and syllable! This means that the perfect spectrogram would need to be a continuous representation of each moment of time and for every frequency.

This is impossible for various reasons, one being the limitations of computational representations, which forces us to make a discrete representation instead. But even if that wasn't the case, the fourier transform doesn't work on single points in time, instead it works on entire signals. That said, we can use the short time fourier transform to give us a spectrogram that is discrete in time. Combined with the discrete fourier transform, we will also have a discrete representation in the frequency axis.

Spectrograms have 3 dimensions, similar to how images are represented, but not exactly. In any given image there are two dimensions that represent space and one that represents intensity. Meanwhile Spectrograms have one dimension for time, one for frequency and the last represents the strength and sometimes the phase of the frequency signal at that time. Although some writers, the standard tends to place frequency as height, time as width and color as signal intensity in a given point.

Although this is a great step forward to giving us a better representation of audio, there are still too many parameters that play a role. Each one of these parameters affects how well the spectrogram can correctly contain the information in the audio. Although there are already well established defaults for most human tasks, it still hasn't been well studied in great part of automated and machine learning tasks.

The exception to this being Speech Recognition, but even so they haven't been heavily studied. There are few papers and datasets that can be used to prove when some parameters are better than others or if there are conservative settings that tend to work consistently in all problems. In the end, some authors question if spectrograms are even the representation that we are looking for for various reasons that we will see in more detail later in Chapter 7.

For now we'll hide our skepticism and focus on the advantages of using spectrograms. The spectrograms we have been referring to so far are what are called linear-spectrograms.

Although these are the simplest and easiest to implement, it is not the only one. We will mention the difference between Amplitude and Decibel Spectrograms as well as Frequency and Mel-Bin Spectrograms.

1.1.1 Amplitude vs Decibel

The values in amplitude are represented in the “color” dimension, giving us the intensities that we see in the spectrogram image. As we saw in the Second Chapter, the amplitude tells us the strength of the signal. In the spectrogram, however, there may also be a phase value. This is because the signals tend to be represented as a Complex number, which can be expressed in cartesian or polar coordinates.

Chapter4AmplitudeVsDecibel_1

Bibliography