Short Term Fourier Transform

After exploring a few different algorithms and techniques in the previous section, we have settled on using the Short-Term Fourier Transform (STFT). This algorithm takes a sound file and splits it up into smaller sound files, often into time steps of one to two milliseconds. Recall from our discussion of sampling that audio is typically sampled at around 44.1 kHz; this means that a one millisecond audio clip will have 44 samples. A Fast Fourier Transform (FFT) algorithm is then applied to these smaller sound files; each FFT is combined into a plot known as a spectrogram. A plot of a spectrogram can be seen in Figure 1.

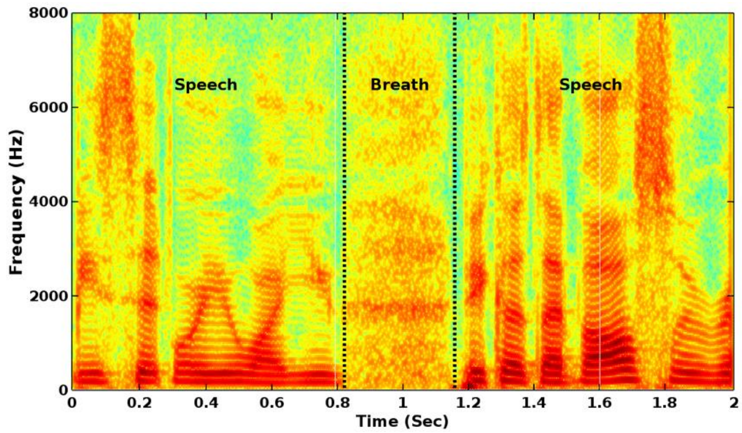

Figure 1: Spectrogram Plot

The color of the plot indicates the intensity of the frequency. For this plot, darker colors indicate a larger presence of the frequency, while lighter colors indicate a smaller presence. This plot can also be viewed as a three-dimensional construct, with peaks in the Z-axis representing intensity of sound. This can be seen in Figure 2.

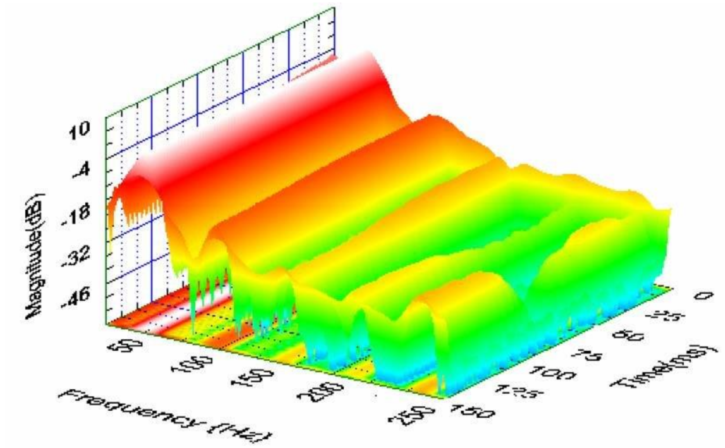

Figure 2: 3D Spectrogram Plot

These representations offer a good intuition of how the spectrogram works; specifically, it shows how much control we have over both frequency resolution and time resolution. Applying this to a sound file of a musical piece, we will know what frequencies occur with what intensity, and when they occur. Both two dimensional and three dimensional plot will be generated for different qualities of violins; from this, we will be able to discern patterns in the upper harmonics.