**Theoretical Background**

Basically, there is two kinds of features that can be analyzed when we are dealing with sounds. These are amplitude and frequency. Sound is one particular type of a waveform which means the frequency or the amplitude value is depending on time. There are two kind of wave forms digital and analog. Sound output in real world can be considered as analog because there is some sort of output for every single moment while the voice recorded devices that we use store the recording in a digital way. Almost every time conversion of analog signal to digital signal should be reversable thus the recorder must have an acceptable sampling rate. This can be determined by sampling theorem.

For an equally spaced repeated sample with a period T and largest frequency component

In terms of sampling period

Given these samples, we can reconstruct x(t) by generating a periodic impulse train in which successive impulses have amplitudes that are successive sample values. This impulse train is then processed through an ideal lowpass filter with gain T and cutoff frequency greater than and less than . the resulting frequency will exactly equal x(t) (Oppenheim).

When we are analyzing real world sounds we don’t usually get smooth sinusoidal waves instead we get complicated and messy sound waves. Thanks to Fourier transform we can decompose the complex periodic sounds into superposition of sine waves which have different frequencies.

Figure 1: Visualize decomposition of complex signal.


Figure 1: Visualize decomposition of complex signal. (2)

There is one disadvantage of this method which is losing the time parameter. This restriction can be avoided by using method called short-time Fourier transform (STFT).

The STFT segments a time-domain input signal into several separated or overlapped frames by multiplying the signal with a window function and then applies the fast Fourier transform (FFT) to each frame. Because Fourier transforms are performed while moving the window, this technique can measure the frequency content changes of a signal over time (Jeon).

In our case what we must consider while evaluating the audio file is, determining the length of the window or fixed frame size. An interesting research done by Carlos Mateo introduces an algorithm for adaptive frame size for specific frequencies. According to researcher the algorithm shares some commonality with STFT multi-resolution technique but it does not require band-pass filter to compute (Carlos Mateo).

Open source python library librosa allows the user to easily get the solution of Fourier transform and short-time Fourier transform with simple function (Librosa). Then result can be visualized by another open source libraries such as matplotlib and seaborn(Matplotlib, Seaborn) .

**Resources:**

* 1] Oppenheim, Alan. (1996). *Signals & Systems* (2nd Edition). Prentice-Hall Series in Signal Processing.
* 2] Mateo, C., & Talavera, J. A. (2018). Short-Time Fourier Transform with the Window Size Fixed in the Frequency Domain (STFT-FD): Implementation. *SoftwareX*, *8*, 5–8. <https://doi.org/10.1016/j.softx.2017.11.005>
* 3] L. (2022, April 22). *GitHub - librosa/librosa: Python library for audio and music analysis*. GitHub. <https://github.com/librosa/librosa>
* 4] M. (2022b, April 25). *GitHub - matplotlib/matplotlib: matplotlib: plotting with Python*. GitHub. https://github.com/matplotlib/matplotlib
* 5] M. (2022b, April 24). *GitHub - mwaskom/seaborn: Statistical data visualization in Python*. GitHub. <https://github.com/mwaskom/seaborn>