Encoding Audio With DFT

EECS 16ML

Sampling

- Sampling a continuous time signal gives a discrete time signal that we can use for signal processing
- **Nyquist–Shannon sampling theorem**: a signal can be sampled and perfectly reconstructed from its samples if the waveform is sampled over twice as fast as it's highest frequency frequency
    - $f_s > 2 f_{max}$

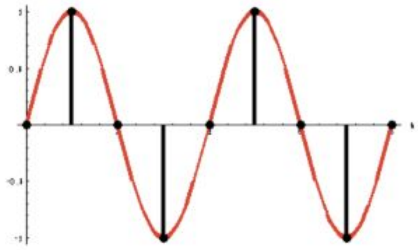


Figure 3: slower sampling rate (source: Berkeley microscopy)

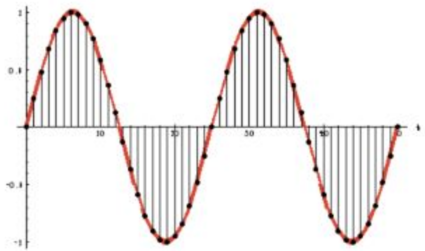


Figure 4: faster sampling rate (source: Berkeley microscopy)

Wave Fundamentals - Simple Waves

- **Wavelength** (λ): The distance between 2 similar points on a periodic wave
  - Here we measure from peak to peak
- **Period** (T): wavelengths per cycle of unit circle
- **Frequency** (f) := 1/T = $2\pi/\lambda$
  - Wavelengths per cycle of unit circle
- **Amplitude** (A): vertical distance from center of wave to peak of wave
  - Tells us the strength of the wave

$f(x) = \cos(x)$

$\lambda = 2\pi$,

$f = 2\pi/\lambda = 1$

$A = 1$

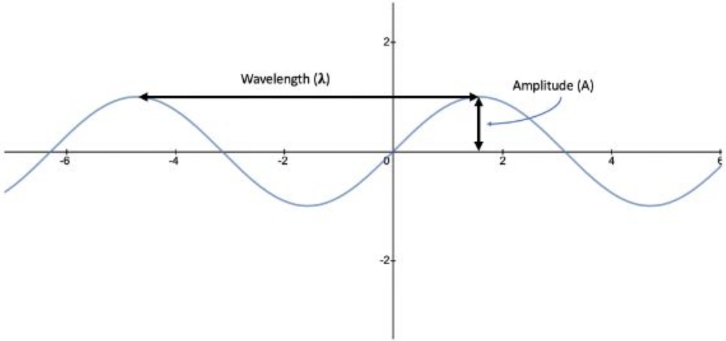


Figure 1: sine wave labeled with amplitude and wavelength

Wave Fundamentals - Slightly More Complex Waves

- Complex signal consists of cos(5x) and cos(2x)
- Each cosine contributes its own frequency to the signal
- Since the amplitudes of each cosine are the same, both cosines contribute an equal amount of their respective frequencies

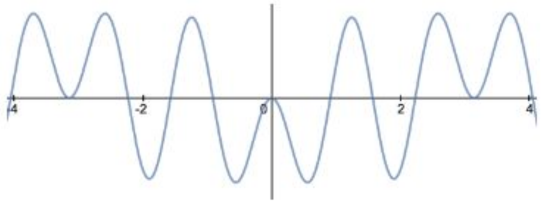


Figure 2: plot of $\cos(5x) + \cos(2x)$

As an exercise, calculate the wavelengths of cos(5x) and cos(2x)

Sound Waves

- A single tone is defined by its frequency
- All sound waves are a linear combination of tones with varying frequencies
- Human range of hearing from 20 Hz to 20,000 Hz

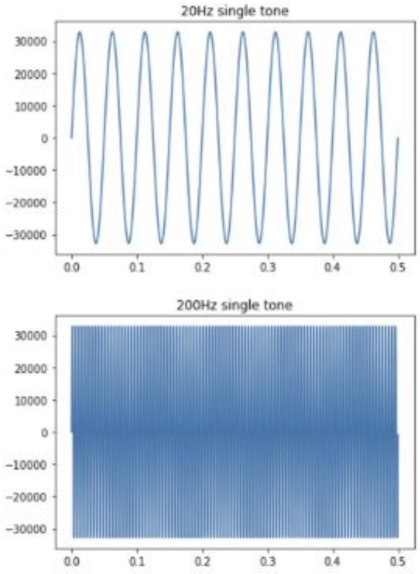


Figure 6: 20 Hz tone (top) and 200 Hz tone (bottom)

DFT of sound waves

- DFT tells us the intensity of different frequencies in a signal
- Since a tone only has a single frequency, the DFT of a tone is concentrated at a certain frequency
- Since sounds are a linear combination of frequencies, the DFT will show the spread of these tones across the frequency spectrum

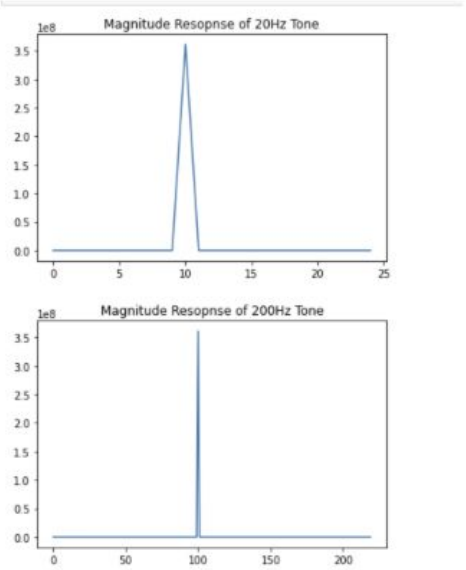


Figure 7: magnitude response of 20 Hz tone (top) and magnitude
200 Hz tone (bottom)

Frequency Spectrum

DFT:

- Transformation from time domain to frequency domain
- $$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-i\omega n}$$

STFT:

- Maps 1D signal to 2D spectrogram
- Gives us temporal information
-

$$X[n,\omega) = \sum_{m=-\infty}^{\infty} x[n+m]w[m]e^{-i\omega m}$$

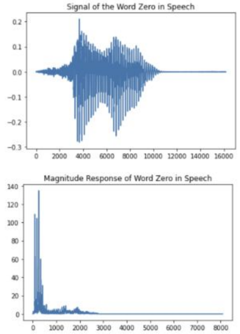


Figure 5: raw signal(top) and magnitude response (bottom) of the word "zero" in speech

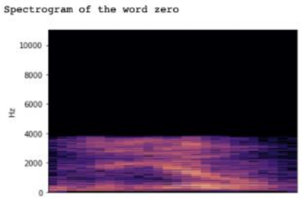
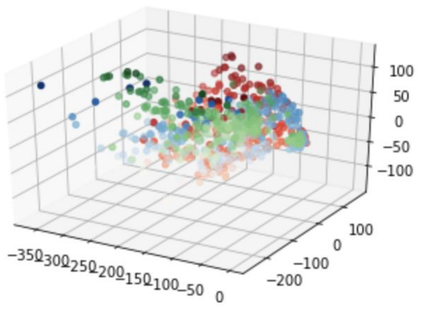


Figure 8: spectrogram of the word "zero"

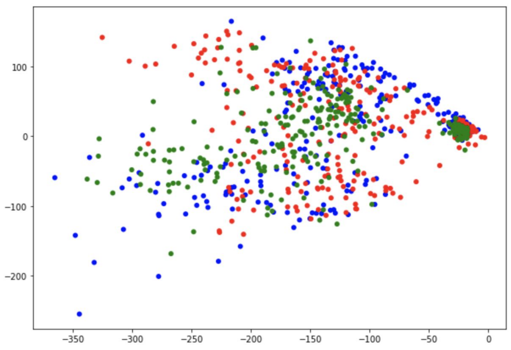
Raw Signal vs DFT vs STFT

Raw Time Varying Audio Signal

- Pretty bad clustering
- 3 words have don’t appear distinguishable in the scatter plots

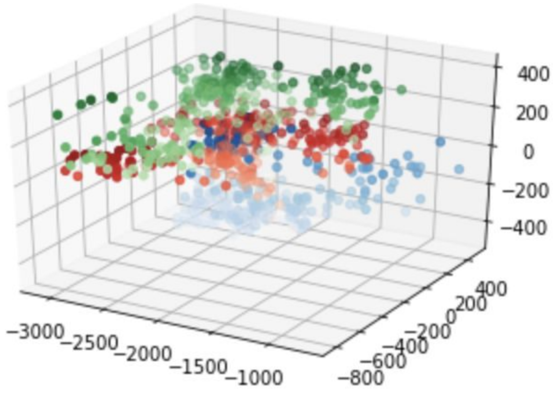


Plotting clusters of time domain signals for 3 different words

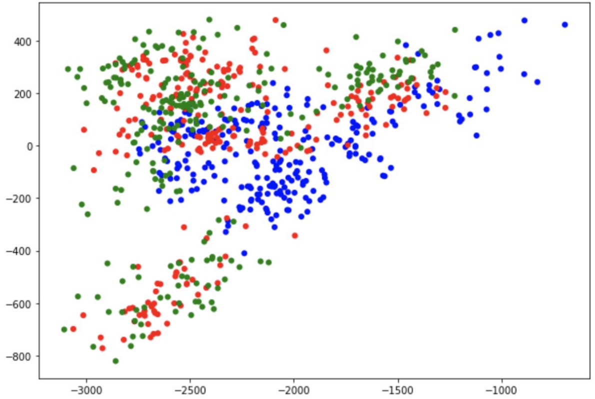


DFT of Audio Signal

- Much better than the raw signal
- 3 words seem fairly distinguishable, especially in 3d scatter plot

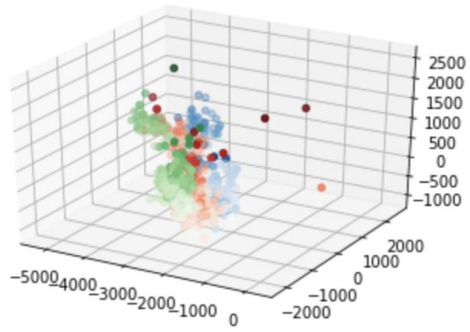


Plotting clusters of the DFT of signals for 3 different words. Here we use decibels to measure the magnitude spectrum to help us better differentiate between very small magnitudes

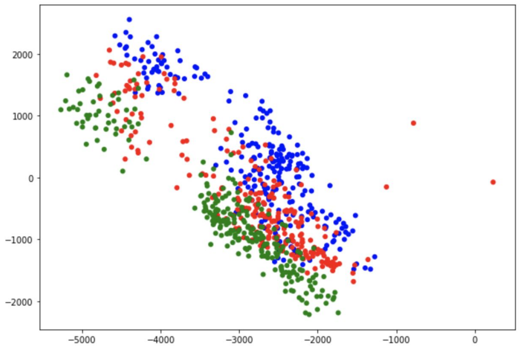


STFT of Audio Signal

- Very similar to DFT of the audio signal, but still performs a little better
- 3 words are even more distinguishable than when using t



Plotting clusters of the STFT of signals for 3 different words



Using CNNs with STFT Spectrograms

- STFT spectrograms are very similar to 2D images
- Can use image classifying techniques such as CNNs
- The spectrogram’s intensities at each time step and frequency can be transformed into a matrix input for the CNN, which can be trained to determine the presence or identity of target signal spectra

References

[1] Allen V. Oppenheim, Signals and Systems, Second Edition, 1997

[2] Berkeley Microscopy, Capturing images, http://microscopy.berkeley.edu/courses/dib/sections/02Images/sampling.html

[3] Dima Shulga, Speech Classification Using Neural Networks: The Basics, https://towardsdatascience.com/speech-classification-using-neural-networks-the-basics-e5b08d6928b7

[4] Jarno Seppänen, Audio Signal Processing basics, 1999, https://www.cs.tut.fi/sgn/arg/intro/basics.html

[5]M. Lustig, EE123 Digital Signal Processing Lecture 5B Time-Frequency Tiling, EECS UC Berkeley