Giaan Nguyen

September 27, 2019

Lab 01

ECE 3366: Intro to Digital Signal Processing

Introduction

In speech analysis, vowels generally represent low-frequency sounds whereas consonants have higher frequencies. In this lab, four different sounds are recorded: “sip”, “nip”, “rip”, and “i-a-i-a”. The first three sounds all have the same \i\ vowel sound and the same \p\ stop sound. Ideally, their spectra should be similar, except the addition of the starting consonant will add additional frequency content respective to that consonant. The fourth sound is a sequence of two different vowels. It is expected that its spectrum should be more concentrated in the low-frequency band. Since the \i\ vowel sound in the first three sounds are somewhat similar to the \i\ in “i-a-i-a”, spectral peaks for \i\ and \a\ should be distinguishable. In addition, four subjects are each recording all four sounds: two males and two females. Because females naturally have higher pitched voices than males, the frequency spectra of the females should be shifted to the right relative to the male spectra.

Procedures

Four subjects – aptly named Female 1, Female 2, Male 1, and Male 2 – were asked to record the four sounds in succession and email the recording. Because of the different kinds of recording devices used, .mp3 and .m4a files were obtained, to which audioread() and audioinfo() in Matlab were used to extract the audio signals and its sampling rate Fs. (It should be noted that while Male 1 unintentionally swapped the order of “nip” and “rip” when reading the sequence, he managed to sneak an “o” at the end of the sequence as a play on a nursery rhyme.) The duration d of the audio signal is given by the number of samples in the signal divided by the sampling rate, with the sampling times defined in increments of the sampling period $T = 1/Fs$ from 0 to $(d - T)$. For each of the four signals, the four sounds were extracted by clipping the appropriate time interval and their respective output.

The fast Fourier transform (FFT) were then applied to each of the extracted sound clips. Since Matlab’s fft() generates two-sided complex Fourier components, a single-sided amplitude spectrum was manually created from the fft() output. To compensate for both odd and even number of samples n, the fft() output was clipped from the first value to $(floor(n/2)+1)$, where the floor will always truncate to an even number, and the addition of 1 will compensate for the duration of discrete-time intervals. Since the spectrum is single-sided, all components except for the first component corresponding to zero frequency are doubled. To define the respective frequencies, the sampling theorem is used, in which the maximum frequency is bounded by twice the sampling rate; therefore, the frequencies are defined from 0 to $Fs/2$ with increments of Fs/nf where nf is the number of components in the adjusted FFT output. Lastly, the amplitude (obtained via absolute value) of the adjusted spectra is plotted, with peaks labeled on the plot. Note that while there are multiple peaks plotted, only the prominent peaks will be analyzed.

Results and Discussion

Figure 1 shows the entire signal of the four sounds for all four subjects. Again, note that for Male 1, the second and third waveforms should have been swapped. When examining the first three sounds, all four subjects have a high-amplitude signal for each signal, with most high amplitudes followed by a short burst. It is inferred that the short burst corresponds to the \p\

sound. Interestingly, Female 1 does not enunciate her \p\ as audibly, if at all, as the other subjects, and hence the lack of the short burst feature following each of the first three sounds. With the exception of Female 1, the sound clips were extracted from the onset of the high amplitude to the offset of the short burst.

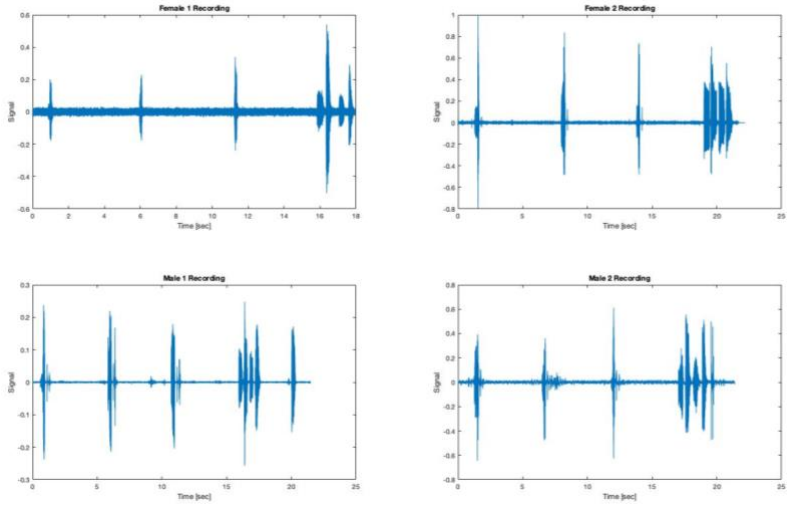

Figure 1. Audio signals of four different voices. Note that Male 1 has the second and third sound swapped and that he added an extra "o" at the end.

Figure 2 shows the spectra of all four voices for the sound "sip". Similarly, Figure 3 shows the spectra for the sound "nip", Figure 4 for "rip", and lastly Figure 5 for "i-a-i-a". For a better view of the figures, use the code provided.

Cross-examining the four figures, we first analyze the similarities between the different sounds per subject. For each subject, there is a general shape similar to each sound, in which there is a low-frequency band (with frequencies up to 500 [Hz]) with high peaks and conversely a high-frequency band (around 2-3 [kHz] for the first three sounds and around 1 [kHz] for the fourth sound) with low amplitude. As suggested in the introduction, the presence of high peaks across all sixteen low-frequency bands suggest that the peaks are features of vowels. Interestingly, the presence of lower peaks in all sixteen high-frequency bands, including those for "i-a-i-a", suggest that the vowels not only have fundamental frequencies within the lower-frequency region, but the vowels also have resonant frequencies within the high-frequency region.

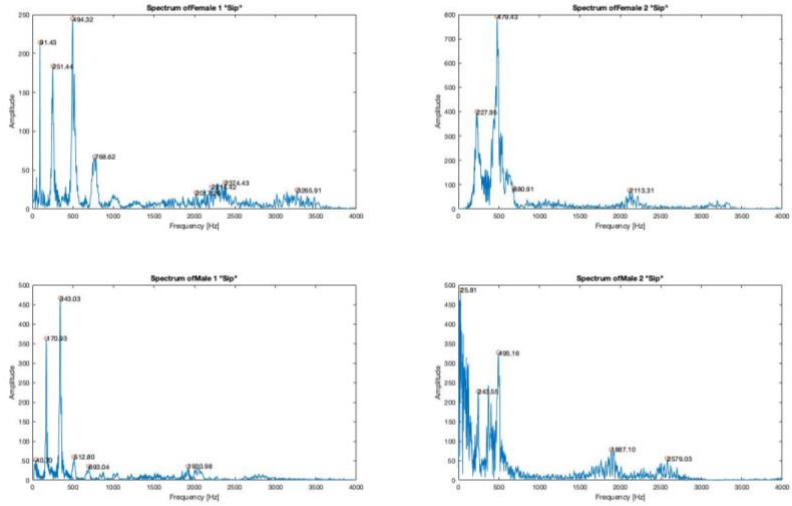

Figure 2. Spectra of "Sip". Female 1 has major peaks in the low-frequency (LF) region at 91.43, 251.44, and 494.32 [Hz], with minor peaks in the high-frequency (HF) region at 2374.43 and 3265.91 [Hz]. Female 2 has major LF peaks at 227.95 and 479.93 [Hz] and minor HF peaks at 2113.31 [Hz]. Male 1 has major LF peaks at 170.93 and 343.03 [Hz] and

minor HF peaks at 1920.98 [Hz]. Male 2 has major LF peaks at 25.81, 243.55, and 495.16 [Hz] and minor HF peaks at 1887.10 and 2579.03 [Hz].

While the low-frequency (LF) peaks are similar between all four sounds per subject, the "i-a-i-a" sound has high-frequency (HF) peaks at approximately half the frequency of the other three sounds' HF peaks. Since the \a\ sound is present, it is most likely that \a\ is more dominant and contributes more content to the spectrum, hence the left shift of the HF band. For "rip", the middle frequencies between the LF and HF bands also contribute content, indicating that \r\ is indicated by some activity in the 1-2 [kHz] range.

Similarly, "sip" has middle frequency (MF) content at around 1-2 [kHz]. However, compared to the MF peaks of "rip", "sip" has minimal MF activity. Interestingly, "nip" has negligible MF activity, possibly due to the \n\ sound starting from a stop, whereas the \r\ is significantly more loose, and the \s\ is somewhat more loose than \n\ but still starts from a closed sound.

When comparing the genders for each sound, it may seem at first that the females' spectra generally have peaks at frequencies greater than the males' spectra by about 100 [Hz]. That is, the females' spectra are shifted to the right by about 100 [Hz] or so. However, while Male 1 does show cases where LFs and HFs are approximately half of that for females for all four sounds, Male 2 shows the opposite. In fact, Male 2 has a larger LF band that starts below that of the minimum Male 1 LF peak and to extends to an LF comparable to the females. Evidently, the shape of the spectra for Male 1 and 2 are different, just as the shape of the spectra for Female 1 and 2 are different; this can be seen in the number of LF peaks. This suggests that while males generally do have lower pitched voices and thus low fundamental frequencies, the resonance frequencies, seen by the number of LF peaks and the ratio relative to the first peak (i.e., the fundamental frequency),

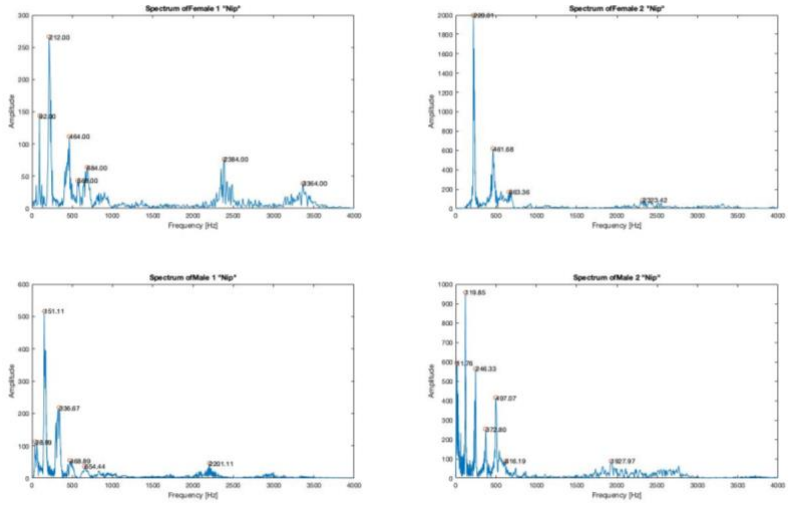

Figure 3. Spectra of "Nip". Female 1 has major LF peaks at 92, 212, and 464 [Hz], with minor peaks in the high-frequency (HF) region at 2384 and 3364 [Hz]. Female 2 has major LF peaks at 220.01 and 461.68 [Hz] and minor HF peaks at 2323.42 [Hz]. Male 1 has major LF peaks at 151.11 and 336.67 [Hz] and minor HF peaks at 2201.11 [Hz]. Male 2 has major LF peaks at 11.76, 119.85, 246.33, and 497.07 [Hz] and minor HF peaks at 1927.97 [Hz].

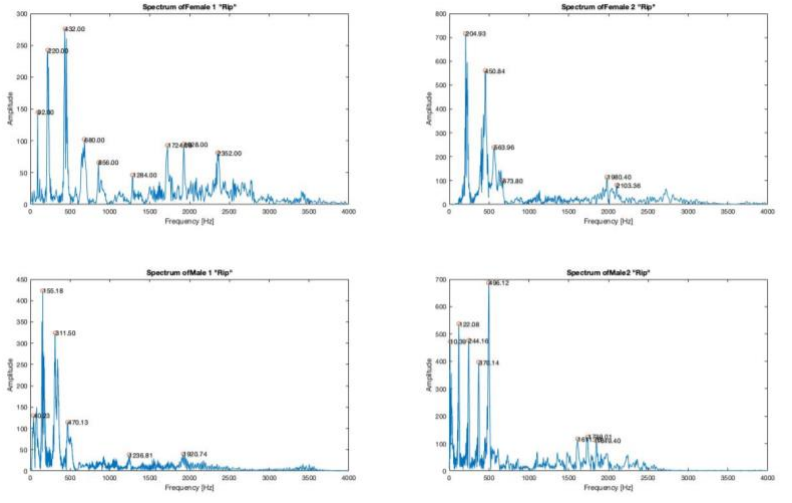

Figure 4. Spectra of "Rip". Female 1 has major LF peaks at 92, 220, and 432 [Hz], with minor HF peaks at 1724, 1928 and 2352 [Hz]. Female 2 has major LF peaks at 204.93 and 450.83 [Hz] and minor HF peaks at 1980.40 [Hz]. Male 1 has major LF peaks at 155.18 and 311.50 [Hz] and minor HF peaks at 1920.74 [Hz]. Male 2 has major LF peaks at 10.39, 122.08, 244.16, 370.14, and 496.12 [Hz] and minor HF peaks at 1611.70, 1729.01, and 1849.40 [Hz].

are what makes each voice – the timbre
– unique.

However, when looking at Male 2 spectra in comparison to the other three subjects, the spectra does appear more distorted than the three counterparts. It is possible that the environment in which each person was recording has varying degrees of noise. In addition, the devices they were using has varying thermal noise. Therefore, for future purposes, the procedures should be done on a single device at the same environment rather than have the subjects record themselves wherever they want. Applying a smoothing filter may also help find major peaks more easily. Lastly, the observations of \s\, \n\, and \r\ is best suited if the individual phoneme is isolated rather than isolating the word.

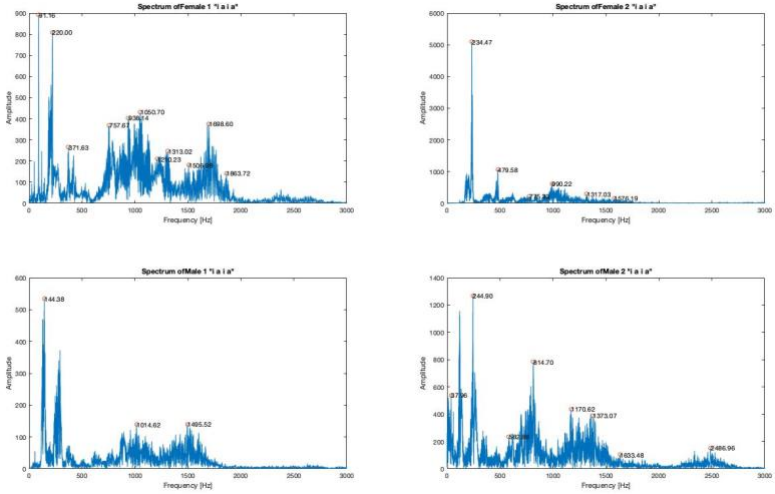

Figure 5. Spectra of "i-a-i-a". Female 1 has major LF peaks at 91.16 and 220 [Hz], with minor HF peaks at 1050.70 and 1698.60 [Hz]. Female 2 has major LF peaks at 234.47 and 479.58 [Hz] and minor HF peaks at 990.22 [Hz]. Male 1 has major LF peaks at 144.38 and approx. 300 [Hz] and minor HF peaks at 1014.02 and 1495.52 [Hz]. Male 2 has major LF peaks at 37.96 and 244.90 [Hz] and minor HF peaks at 814.70 and 1170.62 [Hz].