Giaan Nguyen

November 22, 2019

Lab 03

ECE 3366: Intro to Digital Signal Processing

Introduction

When sampling, the type of window applied to a signal – as well as its length – can affect the distribution of spectral energy within a spectra plot. In particular, leakage describes a phenomenon in which spectral energy at a mainlobe trickles down to its surrounding sidelobes rather than remaining concentrated solely in the mainlobe, resulting in an inaccurate frequency spectrum. In this lab, the effects of the rectangular and Hamming windows on the frequency resolution and spectrum are observed at four different window lengths for the four previously recorded audio signals, each with four sound artifacts ("rip", "sip", "nip", "i-a-i-a").

While the lab mentions a spectrogram, it is important to note that method of using overlapping fast Fourier transforms (FFTs) to obtain a time-series estimate of spectral power is actually Welch's method of periodograms. The spectrogram is a gradient map of time against frequency, with the gradients defined by the power spectral density (PSD). For future purposes, the term "periodogram" will be used.

Procedures

Overlapping segments of each audio signal with 50% overlap are created using windows of length 50, 100, 200, and 400 [ms], separately. (Again, recall that Male 1 flipped the order of two artifacts and added an "e-i-e-i-o" joke towards the end.) In each case, a rectangular window is used first, before using a Hamming window. Applying Welch's method, the FFT for one segment is computed, accounting for overlapping; each overlapped FFT is then averaged and plotted against time to obtain the time-series average periodogram. Welch's method is then repeated for all other window sizes for both types.

A separate piece of code should ideally be able to calculate the fundamental frequency of the audio recording. Keep in mind that the typical fundamental frequency for males tends to be around 85-180 [Hz] and for females, 165-255 [Hz].

After decimating each audio signal by a factor of 8, Welch's method is once again applied to observe the periodogram.

Results and Discussion

In Figures 1-4, the time-series periodograms for all four audio signals for both rectangular (top) and Hamming (bottom) windows at lengths 50, 100, 200, and 400 [ms] with 50% overlap are displayed. A common feature across all four figures, regardless of window function, is that as the window length increases, the PSD becomes less noisy. Obviously, since the sampling frequency f_s is pre-defined by the recording device, the temporal resolution Δt for each audio signal remains unchanged, regardless of windowing. However, by increasing the window size, the frequency resolution Δf decreases – that is, the frequency resolution improves. This is validated by the periodograms becoming less noisy as the window size N increases. This should make sense when considering the formula, $\Delta f = f_s/N$, for the parameters of the FFT.

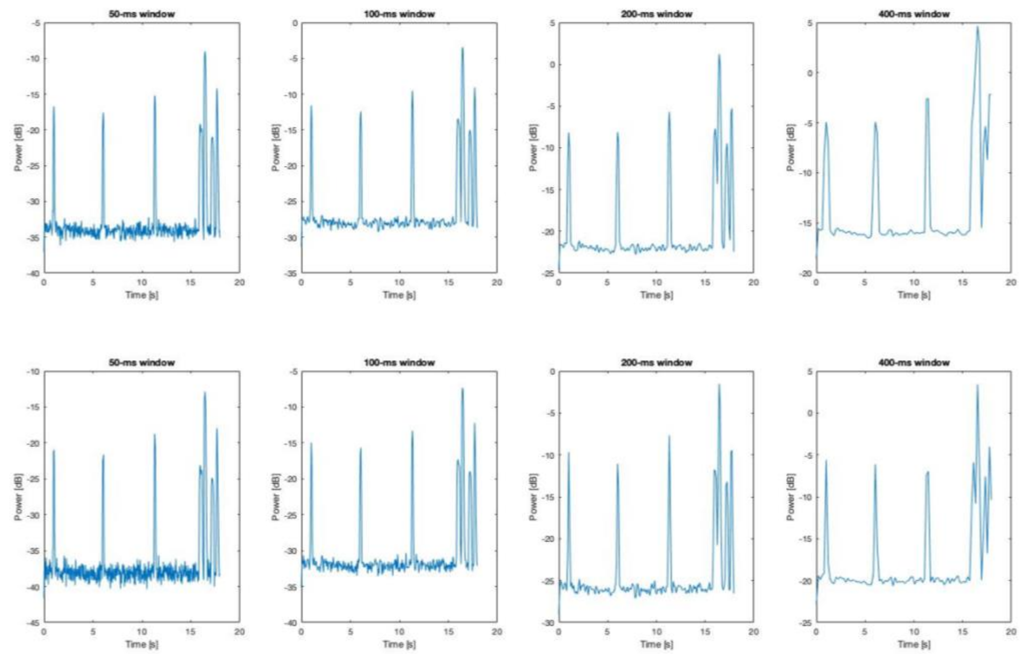

Figure 1. Female 1, windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

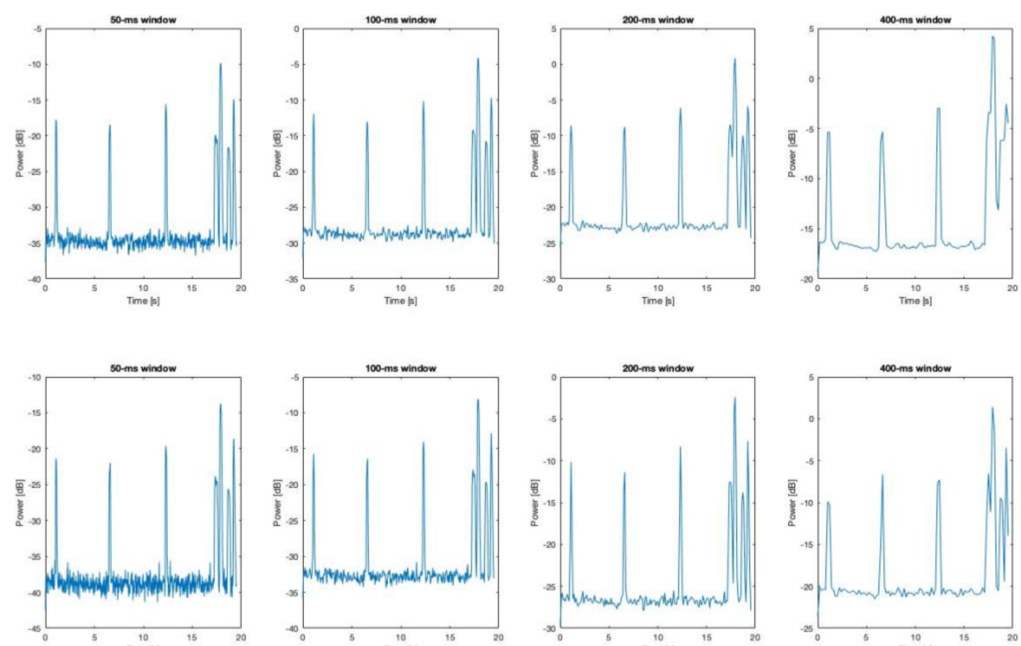

Figure 2. Female 2, windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

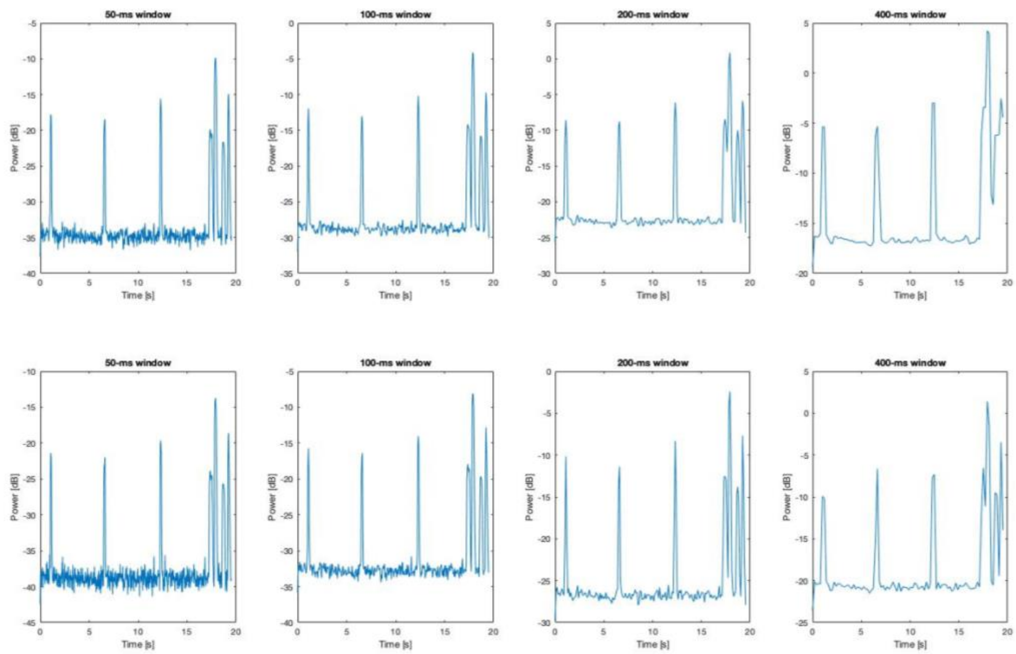

Figure 3. Male 1, windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

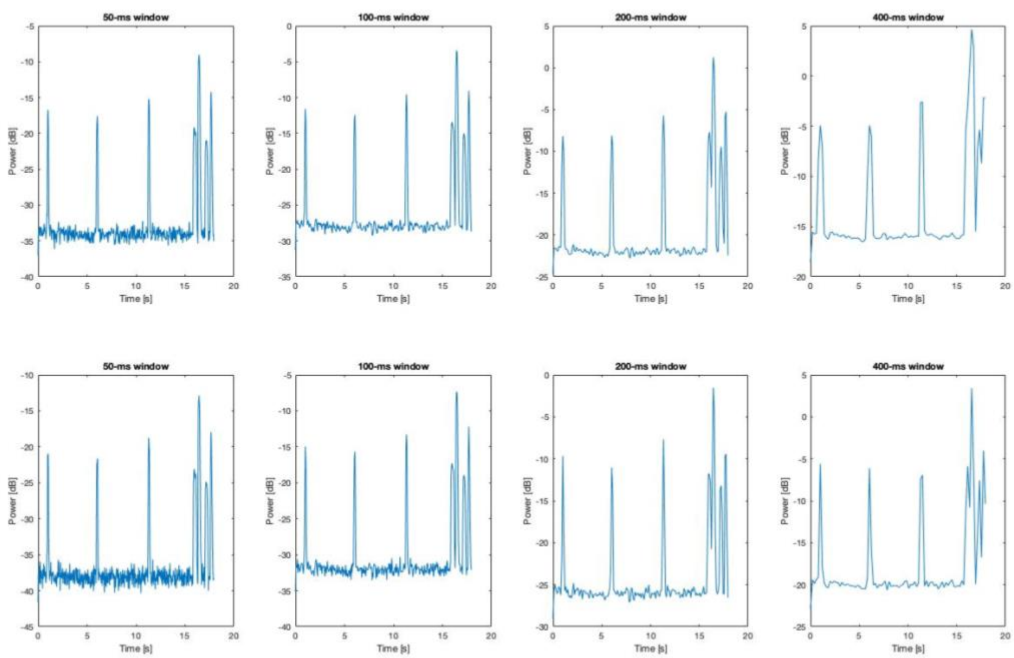

Figure 4. Male 2, windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

Comparing the periodograms due to the rectangular and Hamming windows from Figures 1-4, the dB-levels for the Hamming windowed periodograms are lower than that for the rectangular windowed. That is, while windowing will result in spectral leakage, the sidelobes for the Hamming windowed spectrum approaches zero more than that of the rectangular windowed, minimizing leakage into the sidelobes.

The fundamental frequency is calculated by extracting the maximum PSD peak from the time-series periodogram where there is a large density concentration, using a Hamming window size of 400 [ms] with 50% overlapping. A minimum and maximum bound from the peak time with deviation 0.5 seconds is defined, and the original audio is clipped according to such bounds. The FFT is then applied to the clipped signal. Limiting the scope of interest to 80-300 [Hz] when considering typical fundamental frequencies of the human voice, the location of the maximum peak within the magnitude squared of the FFT is set as the estimated fundamental frequency.

Using "find_fund.m" on the four audio signals, the fundamental frequencies are as follows: Female 1, 90.8203 [Hz]; Female 2, 228.1174 [Hz]; Male 1, 155.4428 [Hz]; and Male 2, 121.582 [Hz]. "find_fun.m" also displays the periodogram and the magnitude squared spectrum of the signal of interest. For Female 2, Male 1, and Male 2, we observe that the maximum PSD is observed during the "i-a-i-a"; in contrast, Female 1's maximum PSD occurs during the second artifact. We can safely say that Male 1 and 2 falls within 85-180 [Hz] and can be classified as male; similarly, Female 2 falls within 165-255 [Hz]. While Female 1 would erroneously be classified as male, it is likely that the short duration of the /i phoneme within "nip" attributes to inaccurate classification. In addition, the selection of "nip" over "i-a-i-a" may be attributed to the mechanics of the recording device, the noisy environment, and the emphasis of the speaker's enunciation on particular words. In general, it can be inferred that to obtain an accurate fundamental frequency, it is best to extract vowels from a speech signal.

Lastly, the original audio signal is decimated by a factor of 8, with Figures 5-8 showing the results of applying a rectangular or Hamming window of lengths 50, 100, 200, and 400 [ms] with 50% overlap, similar to before.

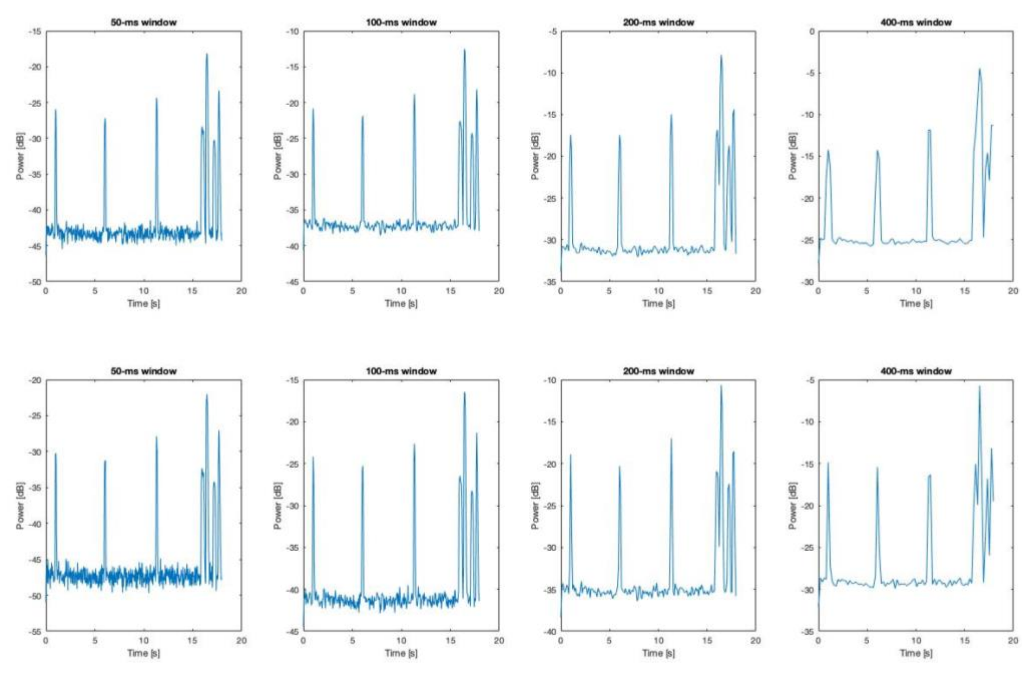

Figure 5. Female 1, decimated and windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

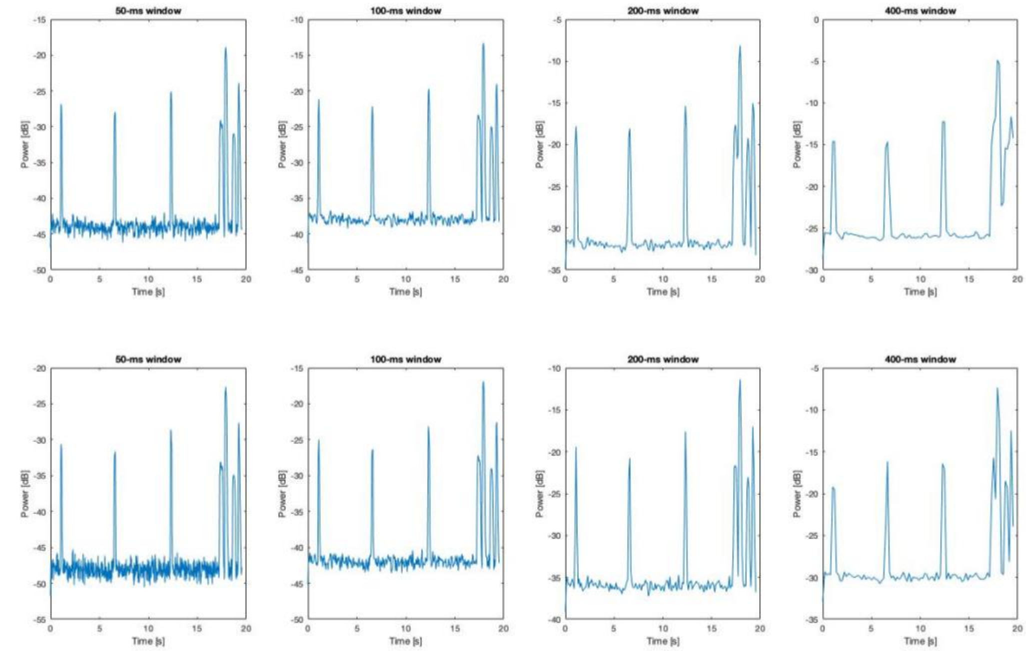

Figure 6. Female 2, decimated and windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

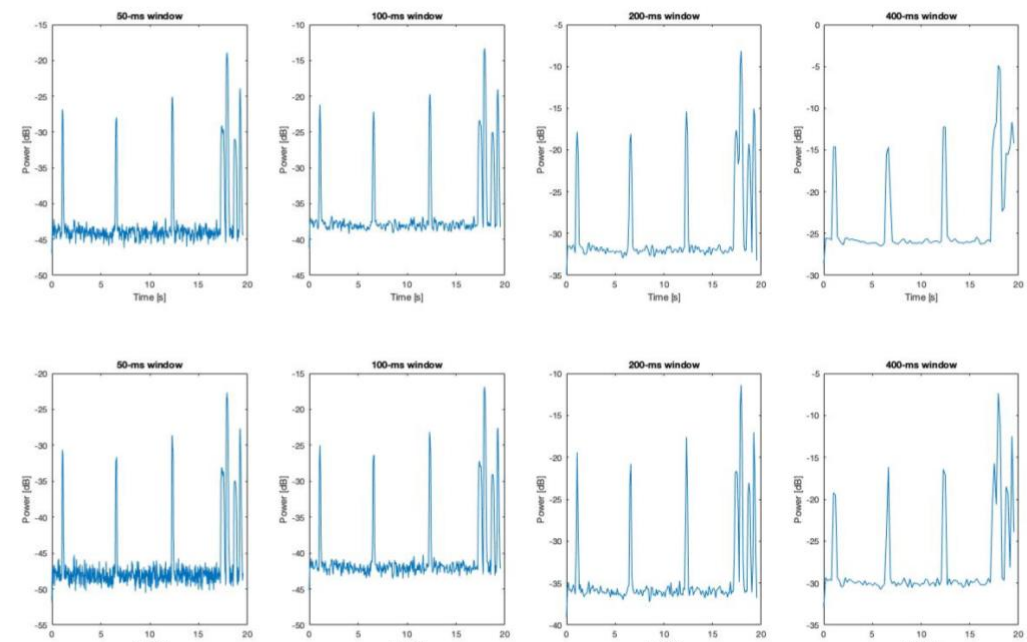

Figure 7. Male 1, decimated and windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

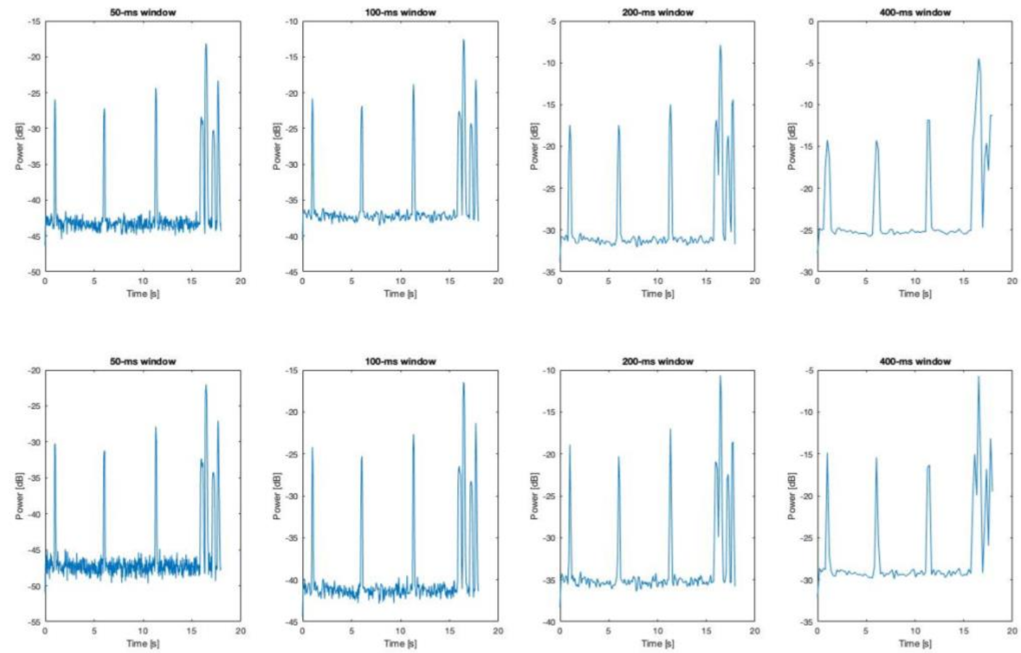

Figure 8. Male 2, decimated and windowed. The top row depicts the time-series periodogram after a rectangular window of the respective length is applied. The bottom row similarly depicts the periodogram after a Hamming window is applied.

Ideally, when decimating the signal, the temporal resolution Δt increases due to the change in the sampling rate $f_s = 1/\Delta t$ – that is, $f_s' = f_s/8$. Therefore, if the sampling rate decreases, the frequency resolution Δf also decreases. When looking at the dB-levels between the original and decimated periodograms, we see that the decimated periodograms undergo an overall decrease in PSD. This can be attributed to the stretching of the frequency axis due to decimation, in which the frequency spectrum is expected to become broader.