EE114D - Computer Assignment #3

Abstract

In this assignment we studied the effect that window size has on the spectral representation of sound. Specifically we tried to determine optimal window lengths for two different speakers. To determine the optimal window length, we took several samples. We looked at the spectral representation using several window sizes determined the lowest frequency where the individual harmonics were still recoverable. We did this for both a male and female speaker. In a broader case we examined window size effects on simple tones and ramping tones. We saw that our spectral resolution increased as the length of the window decreased. Finally we examined how filtering can affect the quality and intelligibility of speech signals. We studied how using a low pass filter can make two sounds that differ only in the high frequency range appear the same.

Questions

I) Window Size

1) For each (male a, female a) approximately what window size is necessary so that harmonics are no longer resolved?

Female spectrum – window length = 320 samples
Female spectrum – window length = 60, 70, 80 samples (blue, red, green respectively)

It is clear that the harmonics are not visible with a window of 60 samples, while they are completely visible with a window of 80 sample. They start becoming visible with a window of 70 samples. So for the female, at a window of 60 samples, the harmonics are no longer resolvable.

Male Spectrum – window length = 320 samples

Male spectrum – window length = 110, 120, 130, 140 samples (blue, red, green, yellow respectively)

I had to zoom in to see these harmonics. But you can see that the harmonics are clearly visible with a window of 140 samples. They are completely unresolved with a window of 110 samples.
What is the DFT-bin spacing (in Hz) for these window lengths?

The zpfft.m file contains the lines
length=max(size(data));
fft_length=length*zpfactor;

we call the zpfft.m function with the line
zpfft(male_a(1:110),8000,8);
which means the zpfactor = 8.

So for a window length of 60 samples, the DFT size is 8*60 = 420.
For the 110 sample window, the DFT size is 8*110 = 880

The DFT-bin is 8000Hz/420samples = 19 Hz/sample for the 60 sample window (female).

The DFT-bin is 8000Hz/880samples = 9 Hz/sample for the 110 sample window (male).

The zpfft function uses a raised-cosine window which has an effective duration that is one half the length of the input. What is the (effective) frequency resolution implied by these windows (in Hz)?

The effective frequency resolution is dependent on the window size. So for a 60 sample window (female), the effective window is really just half that, so it is 30 samples. This means the frequency resolution is 8000 (samples/sec) / 30 samples = 266 Hz

For the male sample, with the 110 sample window, this means the effective window length is really 55 samples. This means the frequency resolution is 8000 (samples/sec) / 55 samples = 145 Hz
II) Spectrograms

2) Describe the spectrogram of a tone, and of the sweep function using the default (initial) settings.

The spectrogram of the tone shows that the tone is constant in frequency and lasts for 2 seconds.
The spectrogram of the sweep shows the tone sweeping across the frequency range. You can see the effect of windowing and overlap, because the image appears jagged as the tone is CONTINUOUSLY changing in frequency, but the FFT is not continuously computed, it is only computed at discrete intervals, this image is showing the default 50% overlap.

By changing to a 240 point overlap (93.75%), you can see that the spectrogram is much smoother, because FFT is computed much more frequently.
3) *How many samples is a 256-point window incremented when the overlap is 240 points?*

If the window is 256 points, and the overlap is 240 points, then the window is incremented 256-240 = 16 samples each time.

*What's the length of the increment in seconds?*

Sample rate is 8000Hz. Sample period = 1/8000 = 0.000125 seconds

16 samples per increment means the length of the increment is 0.000125*16 = 0.002 seconds = 2 ms.

*At what rate, in Hz, are we computing successive FFTs?*

1/0.002 = 500 Hz

4) *When the window length is long, what do the horizontal stripes during voiced sounds correspond to?*

The horizontal stripes correspond to the harmonics during voiced sounds. Using a 980 point window
specgram(female_sentence,1024,8000,980,850)
The spectrogram looks like this:
5) For the male sentence, when the analysis window length is 60, you will see (very fine) vertical striping during voiced sounds. Why do these stripes appear at short window lengths?

As you can see in the spectrogram for a 60 point window, there are vertical stripes in the spectrogram.

![Spectrogram](image)

The vertical stripes are caused by the limited frequency resolution of the short window. In this case with a hamming window of length 60, the resolution is $8 \times 4000 \text{Hz}/60 \text{ samples} = 533 \text{ Hz/sample}.$

6) Of the 3 window lengths (480, 120, 60), which length is best to analyze the resonances of the female vocal tract, and not the pitch period of the driving function?

I think the smaller windows are a bit better because the harmonics don’t even show up, which means I’d choose 60.

Which is best for the male vocal tract?

I felt that the 120 window showed the formants better for the male sample.
7) What happens to the spacing of the harmonics (in frequency) of the note that drops in frequency at the beginning of the music sequence (does the spacing increase or decrease)?

The spacing between harmonics decreases as the note drops in frequency.

What happens to the harmonics of the note that rises at the end of the sequence?

The separation between harmonics increases as the note rises.
Filtering

8) Did you find a cut-off frequency that made /f/ and /s/ nearly indistinguishable?

The original spectrogram is like this:

By changing the filter cutoff to 1500Hz, you can see that the high frequency energy is reduced, the filter also removed the energy at around 1800 Hz from the sibilance in /s/.

If so, what is that frequency?

The best we can do is to change the cutoff frequency to 1500Hz.