Reading: Johnson Ch.2.3.3-2.3.6, Ch.5.5 (today); Liljencrants & Lindblom; Stevens (Tues)
reminder: no class on Thursday

Spectral Analysis

1. There are two major spectral analysis techniques: Fourier analysis and Linear Predictive
   Coding. Fourier analysis is used to calculate the spectrum of an interval of a sound wave.
   LPC attempts to calculate the properties of the vocal tract filter that produced a given interval
   of speech sound.

| waveform and spectrogram | FFT | LPC |
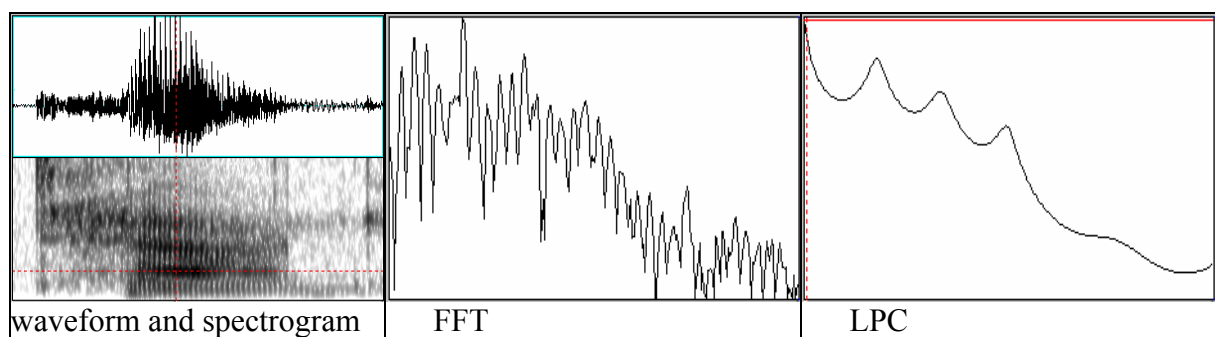
Fourier Analysis (DFT or FFT)

2. Recall that a complex wave can be described as the sum of sinusoidal components. A
   Fourier analysis determines what those components are for a given wave.
   - The analysis technique that we will use is called Discrete Fourier Transform (DFT).

3. The basic idea is to compare the speech wave with sinusoidal waves of different frequencies
   to determine the presence/amplitude of the component with that frequency in the speech
   waveform.
   - Ideally, we would compare a single period of the speech wave with a one period of the
     sinusoidal wave. But generally we don't know the location of a period, so we select an
     arbitrary <u>window</u> (usually about 20-45 ms) and treat it like one period.
   - The analysis then calculates how well sine and cosine waves of various frequencies
     correlate with the speech wave.

        Revealing part of what's in the 'black box': The amplitude of each point in the
        speech wave is multiplied by the amplitude of the corresponding point in the
        sinusoidal wave, and the results are summed. This is called the dot product of the
        waves. If the waves are both going in essentially the same direction at the same
        time, the multiplications will give positive numbers; if they are going in opposite
        directions, the multiplications will give negative numbers. So a high dot product
        means a good correlation. And the degree of correlation indicates the relative
   amplitude of the frequency component in the complex wave.
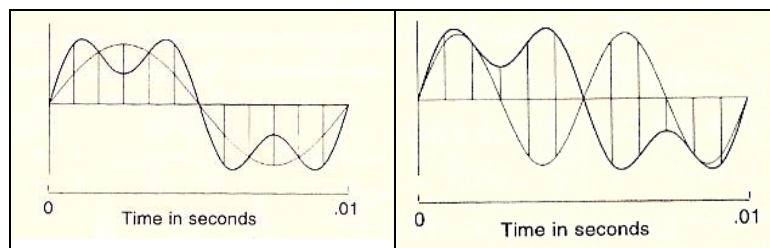
(from Ladefoged, 1996)

4. Window length, also called <u>analysis size</u>, is often measured in points (1 point = 1 sample). Duration of the window also depends on the sampling rate. E.g., 256 points at a sampling rate of 10kHz = .0256 sec (25.6 ms).
   - Most speech analysis software uses a Fast Fourier Transform (FFT) algorithm to calculate DFTs. For computational ease, the number of samples in the FFT window must be a power of 2 (e.g., 64, 128, 256 points).
   - Larger analysis size gives better frequency resolution, but where spectral properties may be changing, the window needs to be short enough to represent the time accurately.
   - Praat allows you to manipulate window size in *seconds*, rather than in points. To know the number of points, you have to calculate based on the sampling rate. E.g., 30 ms at a sampling rate of 10kHz = 300 points.

5. Recall that a spectrogram consists of a sequence of spectra, and the band width of the spectrogram depends on the window-length used to calculate the spectra.
   - This is why in Praat, you use the same window length parameter to adjust both the bandwidth of spectrograms and the window length of FFTs.
     - standard window length for a spectrogram: .005s = 5ms
     - standard window length for a spectrum: .025 or .03s = 25 or 30ms

6. Windows have not only a length, but also a <u>window shape</u>.
   - If we simply take an arbitrary slice out of a waveform, it may begin and end abruptly. As a result, the spectrum of such a wave segment might include spurious high frequency components.

rectangular window (no window)
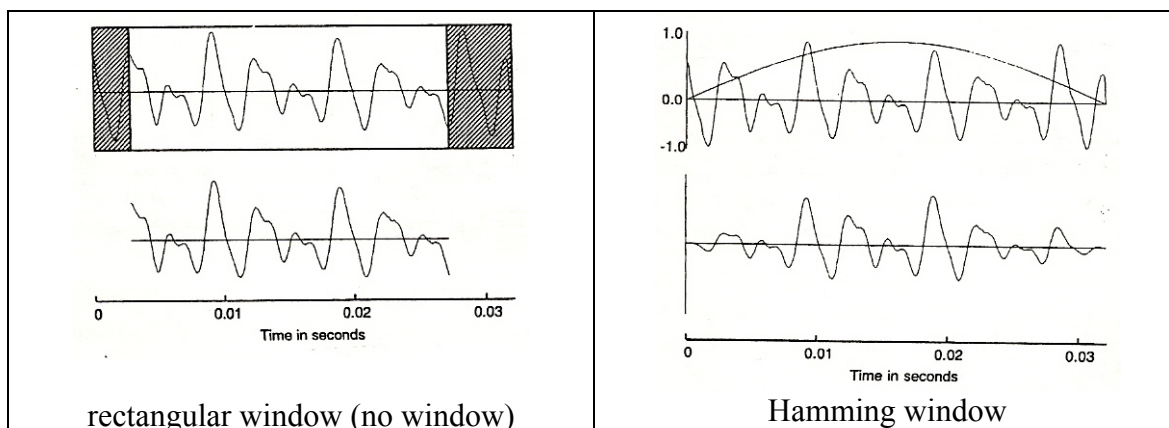
Hamming window

- To avoid this problem, we use shaped windows that go gradually to/from zero at each edge. Speech analysis software usually offers a choice of window shapes. There is relatively little difference between them, but Hamming is probably most common. Gaussian is the default in Praat, but you can choose Hamming or one of the others.

7. There are at least two potential sources of error in this type of FFT analysis.
   - The FFT assumes that the spectrum is stationary through the window of analysis.
   - The window does not correspond to exactly one period of the waveform, so frequencies may be shifted very slightly, but systematically.

Linear Predictive Coding (LPC)

8. Linear Predictive Coding (LPC) analysis attempts to determine the properties of the *vocal tract filter*. In particular, it tries to determine the formant frequencies, or peaks in the filter.

9. The basic principle is 'analysis by synthesis'. If we knew the form of the source, and we know the output waveform, we could calculate the properties of the filter that transformed that source into that output.
   - Because we don't know the exact properties of the source, we make the simplified assumption that the source (for voiced sounds) is a flat spectrum. So the filter calculated by LPC analysis includes the effects of shaping the source (making it slope downwards), as well as the effects of the vocal tract.
   - The analysis seeks to minimize the difference between the predicted (synthesized) signal and the actual signal (i.e., the error).
        > Revealing part of what's in the 'black box': An LPC filter is expressed as a function with a set of <u>coefficients</u>. The number of coefficients is called the <u>order</u> of the filter. Each pair of coefficients defines a resonance of the filter. The order of the filter is specified prior to the analysis (i.e., the phonetician tells the analysis how many resonances to expect). The object of the analysis is to find the coefficients that minimize error.

10. How do you pick the order for the filter? Since it takes a pair of coefficients to specify a resonance of the filter, the number of coefficients should be twice the number of formants you expect to find.
    - The number of formants you expect depends on the range of frequencies contained in the digitized speech signal. Remember, only frequencies up to the Nyquist frequency (which is half the sampling rate) are represented in a digital speech signal.
    - As a rule of thumb, we expect to find about one formant per 1000Hz for a male, less for a female.
    - So if your sampling rate is 22kHz, the signal contains frequencies up to 11kHz, which is the Nyquist frequency. Therefore, you would expect approximately 11 peaks, so you would choose an order of approximately 22.
    - Praat asks you to specify the number of *peaks*, rather than the filter order. So you would simply enter 11 for the case above.

- You can try a range of filter orders (or numbers of peaks) and see what works best. If there are too many coefficients (or predicted peaks), there may be spurious peaks in the LPC spectrum. If there are too few, some formants may not appear in the LPC.

11. FFTs show harmonics, not resonances. LPCs show resonances, but not individual harmonics. An LPC smoothes the FFT using a speech-appropriate vocal tract-like function (based on a simple source filter model of speech), so it is generally well-suited to the analysis of speech and facilitates finding formants.

Spectra in Praat

12. FFTs
- In the Sound window, go to "Spectrogram settings…" in the "Spectrum" menu. Set "window length" to 0.025s (or whatever FFT window length you need). Note that this will also change your spectrogram to be narrow-band rather than wide-band.
- Go to "Advanced spectrogram settings…" in the "Spectrum" menu. Set "window shape" to "Hamming" (or whatever FFT window shape you need—not "square (rectangular)").
- Select a point in the waveform/spectrogram at which you would like to take the FFT. Select "View spectral slice" in the "Spectrum" menu, or press Ctrl-S.
- You will see a spectrum in a new window. The frequencies will go from 0 to the Nyquist frequency. If your sampling rate is high, you may want to zoom in and look at just the first 5000Hz or so.

13. LPCs
- After you have made an FFT, highlight the "spectrum slice" in the "Praat Objects" window. Click "LPC smoothing" to the right.
- In the "Spectrum: LPC smoothing" dialog box, set "number of peaks" to about one per 1000Hz up to the Nyquist frequency of your sound. E.g., if your sampling rate is 22kHz, the Nyquist frequency is 11kHz, so you would choose approximately 11 peaks. (Each peak you specify is the same as *2* coefficients.)
- A new "spectrum slice" object will appear in the "Praat Objects" list. Highlight the new spectrum slice and click "Edit". This will display your LPC. You may want to zoom in to look at just the first 5000Hz or so.

14. A practical note about Praat spectral settings: Since there is just one set of window parameters that is used for all analyses, you will want to remember to reset your window length and window type for making normal spectrograms.

15. Another important note for both of these analyses: For most purposes, do not use the Spectrum button/options in the "Praat Objects" window. This will give you an FFT averaged over your entire sound file (rather than at the point of your cursor). Likewise, do not use the Formants & LPC button with a Sound object highlighted. Only use the LPC Smoothing button when you have selected a Spectrum slice.

Measuring Formants

16. Formant Frequency definitions
    - Technically, a formant is a resonance frequency of a vocal tract (of a given size and shape).
        - a property of the *filter*
    - Practically, a formant is a strong harmonic or harmonics in the speech signal since we can really only see or predict formants based on their effect on the source.
        - property of the output

17. You can measure formants from the spectrogram itself, from a formant track, or from an LPC. (You could also estimate formants from an FFT, but these other choices are usually better.)

18. Measuring formants from a spectrogram
    1. find ⎧ steady ⎫ part (the most characteristic part)
            ⎨ extreme ⎬
            ⎩ middle ⎭
    2. find center of broad band
        - measure frequency
    3. expect ~1 formant per 1000 Hz
        (F1 usually occurs between 200 & 1000 Hz)

    - Typically, you want to measure in a steady state portion of the vowel, but if there's no steady state, choose the point where F1 is at a maximum value. In a diphthong, make sure you're in the right part of the vowel.
    - Unless you're specifically interested in transitions, try to avoid a part of the vowel that is a transition to a following consonant. You can tell you're in a transition if one or more of the formants points up or down right at the edge of the vowel. Transitions are often more visible in higher formants, so make sure to look at F2 and higher.

19. Measuring formants from a formant track
    - Most speech analysis software offers a formant tracking feature, which provides a trace of the formants overlaid on a spectrogram.
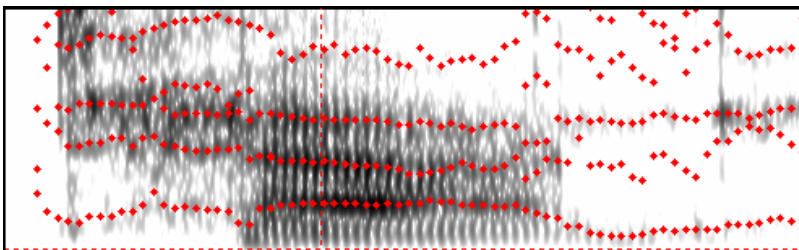
    - Generally, formant tracking is the generated from automatic repeated LPC analyses.
    - Once you have chosen the point where you want to take the measurements, you can query them directly from the formant track.

20. Measuring formants from an LPC
   - Once you have chosen the point where you want to take the measurements, place your cursor there and perform the LPC analysis. (In Praat, this requires you to make a spectral slice and then to apply LPC smoothing. See points 12 and 13 above.)
   - Make sure there are at least 25ms of steady state vowel to the right of the point you are measuring from (assuming you used a spectrum window length of 25ms). The LPC actually looks at a window of data that starts at the cursor.

21. Regardless of the method you use, you should verify that your formant measurements are reasonable. If the formants seem off a little (or even a lot), try moving the cursor on the spectrogram a bit and trying again at a slightly different point. Often one point will show the formants more clearly than another.

22. Fairly small frequency differences are audible. For F1, a 14 Hz change can be heard. For F2, a change of 1.5% can be heard. Repeated measurements, then, should be in this range. However, formant measurements tend to be a bit noisy and they are rarely this accurate.

23. Using Praat to measure formants
   - To turn on formant tracking, select "Show formants" in the "Formant" menu in the Sound window. All of the default settings should be fine.
   - To start a log for your measurements, go to "Log settings…" in the "Query" menu on the "Sound" window.
        - Choose a location and file name for "Log file 2:" I recommend something like:
             C:\Documents and Settings\All Users\Desktop\Formant Log.txt
        - In "Log 2 format:", type the following:
             't1:4' 'tab$' 'f1:0' 'tab$' 'f2:0"tab$' 'f3:0'   (yes, type the single quotes)
          This will give you the start time (t1) and first three formant frequencies (f1, f2, f3) at the cursor point. Click "ok".
   - Now you can record the formant frequencies by simply selecting the relevant point in the waveform (which is linked to the spectrogram, so you can see your cursors in both displays simultaneously) and hitting Shift-F12 (for log 2).
        - Measurements will display in a "Praat: Info" window as well as write to the file you set up.