

**6.542J, 24.966J, HST.712J LABORATORY ON THE PHYSIOLOGY,
ACOUSTICS, AND PERCEPTION OF SPEECH
Fall 2005**

Kenneth N. Stevens

Joseph S. Perkell

Stefanie Shattuck-Hufnagel

Lab 2

09/13/05

Spectrographic and Spectrum Analysis of the English Vowel System

The main purpose of this laboratory is to get you familiar with some of the speech analysis and editing tools that are available in the Speech Group. You will examine some more of the utterances that you recorded in Lab 1, and you will use various analysis methods to examine the sounds. Then you will do more detailed analysis of some of the vowels from the utterances in Lab 1.

Part 1. Spectral analysis and editing with the Klattools.

In this part we will do several kinds of spectral analysis on the vowel [ʌ] and consonant [ʃ] in **shutter**. Log in to a Linux terminal as *labc*. Use the *cd* command to change to the directory for your group. You can choose to create a subdirectory for this lab, i.e. *mkdir lab2*.

To get into the spectral analysis program, type *xkl shutter*. You will see displays of (1) the entire waveform, (2) the waveform with an expanded time scale, i.e. “zoomed in,” and (3) a window that will display the spectrum. Type *i* to include a fourth window with a spectrogram.

(a) Isolate one portion of the waveform (say the vowel [ʌ]) and play it back. This is done by placing the cursor (left mouse button) at the beginning of the portion, typing *w*, and then placing it at the end of the portion, and typing *e*. To play, type *p* or press middle button. Try other portions of the utterance.

(b) Default spectrum. Locate the middle of the vowel and place the cursor there. Look at the spectrum (with a smoothed version overlaid over the discrete Fourier transform). What is the fundamental frequency? What are the frequencies of the first three formants? (The duration of the time window for this spectrum is about 26 ms.) Print the spectrum.

(c) Spectrum with a short time window. Change the window size to 6.4 ms. To do so, either click on the Spectrum menu and select “change spectrum parameters” or type the letter “c.” Both options

will bring up a dialog box. Enter 6.4 in the space next to “# ms in Hamming window for DFT spectrum.” To look at the spectrum, type *d* and give the time. With the mouse, adjust the location of the window to different times within a pitch period. Note the variation in the spectrum. The best location is centered on the first part of a pitch period. Measure the frequencies of the first three formants. If you type */*, the program will automatically find the frequency of the peak nearest the cursor when you click the mouse. Print the spectrum.

(d) Use of averaging. Calculate the average spectrum over a time interval of 15 ms, using the same window size of 6.4 ms. (Type *a* and follow instructions.) Again measure the formant frequencies. Print the spectrum.

(e) Look at (and print) the spectrum in the middle of [ʃ] with the same window sizes of 26 and 6.4 ms. To obtain the average spectrum, (as in (d)), use a longer averaging time, such as 50 ms. Print some relevant spectra.

(f) Comment on the advantages and disadvantages of the different analysis methods for examining vowels and fricative consonants.

Part 2. Measurement of American English Vowels.

References

Peterson, G.E., and Barney, H.L. (1952) *Control Methods Used in a Study of the Vowels*, **J. Acoust. Soc. Am.** **24**, 175-185. Also in R. Kent et al. (1991) **Papers in Speech Communication: Speech Production**, Acoustical Society of America, Woodbury, NY., 585-594, and in J. Hillenbrand et al. (1995) *Acoustic Characteristics of American English Vowels*, **J. Acoust. Soc. Am.** **97**, 3099-3111. See also pp. 257-322 of Chapter 6 of **Acoustic Phonetics**, K.N. Stevens.

Procedure

Spectrograms of stored waveforms can be calculated and printed out using the computer in the lab. Broadband spectrograms are to be made of each of the word pairs containing examples of English vowels (use *lspecto*). Select the *best* speaker among the lab partners (other speakers may be analyzed, if desired).

Determine the midpoint in time for each simple vowel. Select two or three measurement points for diphthongized vowels, but try to exclude formant motions associated with consonantal transitions. Measure and tabulate the frequencies of the lowest three resonances of the vocal tract (formants). Plot each vowel as a point or as a trajectory (with two or three points) on the graph paper in this handout. From this plot, formulate the simplest possible acoustic relationships between the first two formant frequencies and the distinctive features that were described in lecture:

1. Raise tongue body to a HIGH position
2. Lower tongue body to a LOW position

3. BACK tongue body
4. Make TENSE (non-LAX) articulation
5. Round lips

For two of the vowels, use the xkl program to obtain spectra sampled at the midpoint. It is best to use pre-emphasis to calculate the spectrum (First difference should be 100). In each case, display three spectra: (1) using the S command, a dft and a smoothed spectrum superimposed on the dft; (2) using the l command, a dft and an lpc spectrum superimposed; (3) setting the dft window to 6.4 ms, placing the window on the first part of a pitch period, and clicking on the peaks to get the frequencies. Compare the estimated values of F1, F2, and F3 obtained from the spectra and from the spectrograms. You can create a printout with up to four spectra per page, using the “g” command to open a .ps file and the “#” command to save the spectrum window to the file.

The lab report should include (1) spectrograms with the measurement points marked in pencil, (2) the plot of F1 vs. F2, (3) your acoustic definitions of distinctive features, and (4) spectra and observations concerning formant frequencies obtained from the spectra.

Part 3. Influence of 2nd subglottal resonance.

For the diphthong /aɪ/ (in the word buy) the F1 trajectory moves from a high to a low frequency, while F2 moves from a low to a high frequency. Make a spectrogram of the word buy. The second resonance of the subglottal system for most adult speakers is in the frequency range 1400 to 1500 Hz. For the diphthong /aɪ/, F2 passes through the frequency of this second subglottal resonance. Often there is a brief decrease in the amplitude of the vowel waveform as F2 passes through this point. Look at the spectrogram, the spectrum in the vicinity of this point, and the waveform, to see whether there is a “glitch” at about 1400 to 1500 Hz. Evaluate the size of the discontinuity in amplitude and frequency.

The theory that describes how a vocal-tract resonance can be influenced by a nearby subglottal resonance is discussed briefly in *Acoustic Phonetics*. The significance of this interaction (and others) will be discussed further in a laboratory on the “quantal theory” (Lab 8).

Graph paper

