

**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 15

11/01/05

**Speech Synthesis Using a Formant Synthesizer**

**References**

Klatt, D.H. (1980) *Software for a Cascade/Parallel Formant Synthesizer*, **J. Acoust. Soc. Am.** **67**, 971-995.

Klatt, D.H. and L.C. Klatt (1990) *Analysis, synthesis, and perception of voice quality variations among female and male talkers*, **J. Acoust. Soc. Am.** **87**, 820-857 (optional).

Klatt, D.H. *Description of the cascade/parallel formant synthesizer*. **Chapter 3 of a book in preparation.**

**Procedure**

The computer facility will be used to synthesize a monosyllabic English word or (better still) a phrase, using the formant synthesizer described in Klatt (1980) and in Klatt and Klatt (1990). It will take two lab sessions to finish this synthesis.

Select a short word or phrase containing one or two “interesting” consonants and have one lab partner record it (or you may use an utterance already recorded). If you are synthesizing a female voice, you should use a sampling rate (SR) of 11500 Hz. For a male voice, use a SR of 10000 Hz. (Reasons for this will be explained in class.) Use the *lspecto* program to produce a spectrogram, a fundamental frequency plot, and a plot of formant frequencies, in order to measure and tabulate the motions of the formant frequencies every 25 to 50 ms, and to measure fundamental frequency. You may also wish to obtain a printout of fundamental frequency and formant frequencies. (The commands are:

*lspecto* <file> for spectrogram; *lpr file.ps* to print

*lspecto -f0* <file> for fundamental frequency; *lpr file.ps\_f0* to print

*lspecto -pf <file>* for formant frequencies; *lpr file.ps\_pf* to print

*lspecto -syn <file>* for printout; *enscript -G file.doc* to print)

In order to get the vowel quality right, obtain an xkl printout of the spectrum in the middle of a vowel in the recording. Using the synthesizer with a short duration (DU = 50) try to adjust the various formant and glottal parameters (OQ, TL, AH, F0, AV, and the formant bandwidths, particularly B1) to synthesize a vowel segment with a spectrum that is a good match to the spoken vowel. Set the sampling rate (SR) to the same rate you used for your recording (11500 for females, 10000 for males). Use these glottal parameters as a basis for synthesis of the vowel or vowels in the utterance.

Next, use the *d* (draw) command of the synthesizer to enter synthesizer parameter values over time. If a synthesizer control parameter cannot be measured accurately from the analysis printout (e.g. formant bandwidths, relative source amplitudes), use theory to estimate an initial value, and trial-and-error spectral comparisons of synthetic and natural waveforms to improve input parameter estimates. After 2 or 3 trial-and-error iterations, the synthetic waveform should be both intelligible and similar to the speaker after whom it is modeled. Make a spectrogram of your final synthesis.

The lab report should contain a brief discussion of the problems you encountered, and should show the two spectrograms, together with examples of spectra sampled at crucial locations within the word. We will make a recording of some of the original and synthetic utterances to play in class.