6.542J, 24.966J, HST.712J LABORATORY ON THE PHYSIOLOGY, ACOUSTICS, AND PERCEPTION OF SPEECH
Fall 2001

Lab 13 *10/30/01*

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 via the computer room microphone (or you may use an utterance already recorded). 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 msec, 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*

lspecto file -f0 for fundamental frequency: *lpr file.ps_f0*

lspecto file -pf for formant frequencies *lpr file.ps_pf*

lspecto file -syn for printout: *print file.doc*)

If you are synthesizing a female voice, you should use a sampling rate (SR) of 11500 Hz, with the number of formants (NF) equal to 5. For a male voice, use the default SR of 10000 Hz. (Reasons for this will be explained in class.)

In order to get the vowel quality right, obtain a 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. 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.

6.542J, 24.966J, HST.712J LABORATORY ON THE PHYSIOLOGY, ACOUSTICS, AND PERCEPTION OF SPEECH
Fall 2001

Kenneth N. Stevens (36-517, x3209) Joseph S. Perkell (36-591, x3223)
stevens@speech.mit.edu *perkell@speech.mit.edu*

Stefanie Shattuck-Hufnagel (36-523, x3201) *stef@speech.mit.edu*

Course Secretary: Arlene Wint (36-511, x7309) wint@speech.mit.edu

Lab 13 Supplement *10/30/01*

Supplement to Speech Synthesis Using a Formant Synthesizer

Suggestions for copy synthesis

1. Start by setting the duration of the utterance to be synthesized (parameter *du*). For a male voice, use the default sampling rate of 10 kHz. For a female voice, use a sampling rate of 11.5 kHz for both the spoken speech sample and for the synthesizer (parameter *sr*).

2. Pick a point in the middle of a vowel. Make a spectrum of the vowel at this point. Then set the constant synthesizer parameters for F0, formant frequencies, OQ, TL, and formant bandwidths, to give a good match of the synthesized spectrum to the spectrum of the original utterance.

3. Draw the parameters for F0, F1, F2, and F3 as a function of time. Note that when F0 is set to zero, there is no voicing source. Formant parameters must be continuous and should **not** be set to zero or to weird values. Usually F4 and F5 should be set to constant values. After each synthesis iteration, give the *.doc* and *.wav* file a new name.

4. Once the formants and F0 are done, proceed with parameters for frication noise, aspiration noise, and nasalization. You will need to do some matching of spectra between the original utterance and the synthesized one.

Using the synthesizer

Log in on *labc*.

Type *xkl*

Click on default or, if you already have a *.doc* file, find that file.

There is a set of default settings for the synthesizer. To see these, type *p*.

To change a default (such as F0, F1, … AV…), type *c*.

To draw a time-varying parameter, type *d*. Enter the parameter to be changed. The time function of the parameter will be displayed. You are asked to enter numbers in pairs: a time value and a parameter value. Graph will show a piecewise linear function. The numbers must be entered sequentially.