

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering,
Department of Speech and Hearing Sciences, and
Department of Linguistics

ECE 598 THE SPEECH CHAIN

Laboratory Exercises 6
Fall 2006

Labs for the week of: Monday, October 23, 2006

Write-up due in lab during the week of: Monday, October 30, 2006

Problem 6.1

Record yourself (or your lab partner) saying the following three phrases: “the bug,” “the dud,” “the guppy.” Look at wideband spectrograms of these three utterances.

- Identify the formant frequencies of the first two vowels in each utterance (these vowels have SAMPA transcriptions AX and AH). Are these two vowels different primarily because they have different formant frequencies, or because of duration?
- Find the formant loci of the three consonants (for F_1 , F_2 , and F_3). Remember that the formant “locus” is the resonance of the vocal tract right at the moment of consonant release (at the instant of the stop burst), even if the formant frequencies are not visible at that instant in time.
- How long does the formant transition take, from stop consonant to following vowel? Do the F_1 , F_2 , and F_3 transitions take the same amount of time? Do the /b/, /d/, and /g/ transitions take the same amount of time? Is the formant transition a straight line (in frequency), or an asymptotic approach, or some other type of curve?
- Take the DFT of a relatively long window of samples from the middle of one of the vowels (25-30ms). Plot the log magnitude spectrum, in decibels (this should be the default). Notice that the log magnitude spectrum looks like a sampled transfer function: there are a series of peaks (one at every multiple of the pitch frequency), and the tops of the peaks sketch out a transfer function with peaks at F_1 and F_2 . The spectrum is given by

$$\begin{aligned} S(\omega) &= T(\omega)E(\omega) \\ 20 \log_{10} |S(\omega)| &= 20 \log_{10} |T(\omega)| + 20 \log_{10} |E(\omega)| \end{aligned}$$

where the differentiated source spectrum $E(\omega)$ is

$$E(\omega) = R(\omega)U_G(\omega) \propto \omega U_G(\omega)$$

Recall that $20 \log_{10} |T(0)| = 0$. Find L_1 and L_2 , the levels of the transfer function at the first and second formant frequencies. Find B_1 and B_2 , the bandwidths of the first two formants in Hertz — you may need to interpolate between pitch harmonics in order to find the 3dB-down points. Verify that $L_1 \approx 20 \log_{10}(F_1/B_1)$, and $L_2 \approx 20 \log_{10}(F_2/B_2) - 20 \log_{10}(F_2/F_1)$.

Problem 6.2

Record yourself or your lab partner producing a nasalized vowel—for example, say the word “can” slowly, without closing your tongue tip at the end. Look at a spectrogram of the utterance; can you see B_1 broadening as the vowel becomes more and more nasalized? Look at a DFT spectral slice from the nasalized part of the vowel. Can you identify the two F_1 peaks, and a dip in between them at the frequency of the zero? The effect may be subtle, because of sampling by the pitch harmonics. For example, you may find that the F_1 peak has one tall harmonic, then a short one (near the zero), then another tall one; or you may not see any shortened harmonics, if the zero falls between two pitch harmonics.