

**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 17

11/10/05

Use of Landmarks and Features for Speech Recognition

References

Stevens, K.N. (2002), *Toward a model for lexical access based on acoustic landmarks and distinctive features*. **J. Acoust. Soc. Am.**, **111**, 1872-1891.

Part 1: Estimating features from the speech signal (bottom up)

In this part of the lab, you will be analyzing and annotating acoustic events in two sentences. The waveforms of these two sentences are stored in files labeled **senta** and **sentb**, under lab17. You can choose various ways of analyzing the sentences by making spectrograms, doing spectrum analysis using various time windows, measuring fundamental frequency, etc. You must **not** listen to the sound generated by these waveforms. You are asked to carry out the following tasks for each sentence. Your results should be recorded in the attached tables labeled **Sentence A: Vowels**, **Sentence B: Vowels**, **Sentence A: Consonant landmarks**, and **Sentence B: Consonant landmarks**.

- (a) In each vowel, find the point in time at which the amplitude of the first formant prominence in the spectrum is a maximum. Record these times in the table for Sentence A and B. Also record the relative amplitude A1 (in dB) of the first formant peak.
- (b) At the landmarks identified in (a), find the frequencies of the first three formants F1, F2, and F3, and the fundamental frequency F0. Enter the values in the table.
- (c) Based on the measurements in (b), and possibly on other observations in time intervals preceding and following the landmarks, estimate the distinctive features of each vowel. The features are given in the table, and you should enter + or – for each. If you are uncertain, you may enter up to three estimates for each feature. There is room in the table for you to enter the confidence with which you are estimating each feature, and also to enter up to 3 estimates of the confidence for each estimate. The confidence should be specified by a number from 1 (little confidence) to 5 (very confident). If you make more than one estimate for a feature, the sum of the confidence numbers should be 5. Note there is a

feature [reduced]. If a vowel is judged to be reduced (as the first vowel in about), then most of the other features do not need to be given. However, try to estimate the value of the feature [back] for the reduced vowels.

- (d) In the appropriate column of the table, give the phonetic symbol for the vowel (or more than one symbol if you are in doubt about one or more features). Write down the sum of the confidences for the various (up to 3) estimates of the phonetic symbols.
- (e) Find the places in the sentence where one edge of a vowel is adjacent to a consonant, creating an abrupt change in the spectrum amplitude in some frequency ranges. Tabulate the times at which these landmarks occur and indicate whether the vowel is to the left (L) or the right (R) of the landmark. Use the tables Sentence A consonants and Sentence B consonants.
- (f) For each landmark in (e) indicate whether the consonant is [+sonorant] or [-sonorant], i.e., sonorant or obstruent. If it is [-sonorant], indicate whether it is [+continuant] or [-continuant], and, if [+continuant], whether [+strident] or [-strident]. Enter your estimates in the tables. As in (c) above, you may make more than 1 estimate (and up to 3) of each feature, again giving confidences.
- (g) Identify two of the words in Sentence A and two in Sentence B. If there is a discrepancy between the features of the segments in your estimated words and the features entered in the tables, explain or discuss.

Part 2: Verifying hypotheses about the features and segments (top down)

In this part of the lab, you are given two brief utterances in which two of the words are unknown. These utterances are **sentc** and **sentd** under lab17. Each utterance is of the form “Say --- --- *please*.” For each utterance, an error-prone automatic speech recognizer has come up with a list of six possible word sequences, as given below. Your task is to determine, for each utterance, which of the word sequences is the correct sequence, **or** to state that **none** of the sequences is correct. Knowing what you have learned about the acoustic patterns for the features that make up each hypothesized word sequence, you should check whether this pattern is consistent with what you see in the signal. Give a brief discussion of the measurements that support your answer.

Again you may use any kind of signal analysis you wish (e.g., spectrograms, F0, spectra, durations) but you must **not** listen to the utterances.

Sentence C

the button
the soccer
beware of
his ribbon
her waking
his wiggle

Sentence D

that fan
his can
dish pan
this pan
big bomb
this pen

