24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

Waveform Editing and Speech Perception
due Tues 3/13

This exercise involves editing waveforms with reference to the waveform and spectrogram, listening to the results, and then trying to explain what you hear. Write up your observations and explanations.
The idea is for you to get more familiar with speech as an acoustic signal and to learn a bit about the nature and distribution of perceptual cues to phonological contrasts. The Wright (2004) reading should be useful – it contains an overview of the perceptual cues to various contrasts between speech sounds.

You can do the editing using PRAAT.

General principles for waveform editing:
When deleting or inserting segments of speech you want to avoid creating discontinuities – that is, you want the edges of an inserted portion to join smoothly with the adjacent waveform. If there are any abrupt jumps they will sound like clicks. Tips for achieving smooth edits are included below.

The basic method of editing waveforms in PRAAT is by cutting and pasting: select a portion of the waveform, cut it out using the 'cut' command from the 'edit' menu, then place the cursor where you want to insert that waveform segment, and select 'paste' from the edit menu. (You can also copy a selection).

Exercises

1. **gun.wav**
        This file contains the words 'gun' and 'gum'. The basic idea of this part is to replace the [n] of 'gun' with the [m] from 'gum', and vice versa – i.e. exchange the final nasals.
        Make sure that you make the cut between vowel and nasal as close to the point of closure for the nasal stop as possible.
        To get a smooth edit with voiced speech, make all cuts at the beginning of a period, where the waveform crosses the horizontal 'zero' line (known as a 'zero crossing'), then when you insert a segment, insert it at the beginning of a period, where the waveform crosses zero, and the two sections of wave should join smoothly. Under the 'select' menu in the edit window, there is a command to move the cursor to the nearest zero crossing, and similar commands to move the beginning or end of a selection to the nearest zero crossing. (NB not all zero crossings are at the beginning of a period, so you still need to check the location of the cursor carefully even if you use these commands).

Listen to the edited words. What do they sound like? In particular, did the editing change 'gun' into 'gum' and vice versa? It may help to play your edited file to a naïve listener, and ask what they hear. Hypothesize an explanation for your observations.

2. **lose.wav**
This file contains the word 'lose'. Delete about 0.16s from the vowel [u], reducing the vowel duration from about 0.31s to 0.15s.

For this to sound natural, you shouldn't delete a single block of 0.16s, because the result will have an abrupt drop in amplitude – it's better to cut out several shorter portions, being careful to make the cuts at zero crossings between periods, as above. Also, don't delete the very beginning and end of the vowel.

Listen to the result. What does it sound like? Why? Again, it may be helpful to have a naïve listener tell you what they hear.

3. **say.wav**
This file contains the word 'say' (it ends abruptly because it is excised from a sentence). Place your cursor at the end of the [s] of 'say'. Play the file from that point on. What word does it sound like? Try playing from different points near the transition from [s] to [eɪ].
Can you get the word to sound like 'day'? How? Why?
Can you get it to sound like (non-word) 'tay'? If not, why not.
What do you have to do to make it sound like there is no initial consonant (or perhaps an initial glottal stop)?

4. **stops.wav**
This file contains the CV syllables [bi, di, ba, da]. These are to be used in a cross-splicing exercise: separate the CV syllables into stop release (burst+frication) and vowel. Paste the burst from [bi] onto the vowel from [di] and vice versa. Play the results. What do they sound like? Why?
Do the same with [ba] and [da]. What happens in this case? How does it differ from the pattern of percepts you obtained by cross-splicing [bi] and [di]? Suggest an explanation.