Friday 7
Name:_____

Speech synthesis and analysis

What parallels exist (what is the analogy) between the vocal tract and a music synthesizer?

Both can make a wide variety of sounds. Both have a power source. Both start with an adjustable oscillator (vocal chords) and noise generator ("fricatives" as in the letter f) and subsequently modify the sound by filters (formants). Both have ways of controlling the envelope of sounds (“plosive” sounds such as p and t depend on the envelope). What the synthesizer does with voltage controlled filters, is accomplished by flexing muscles to change the size and shape of air pockets for the human voice.

One other effect, demonstrated on the didjerido, is nonlinear generation of “combination tones” due to the nonlinear response of the lips and of the vocal chords when a sung sound is combined with a played sound. This effect is very similar to how a “ring modulator” combines sounds in an electronic synthesizer.

What is a formant and how does it affect speech and singing?

The vocal system has a variety of cavities (mouth, nasal cavities, larynx) and these form simple harmonic oscillators (Helmholtz oscillators) which filter vocal sounds. These filters are called formants. As filters go they are "band pass" filters. This means they tend to enhance frequencies in a central "pass band" relative to other frequencies on either side. They are usually fairly broad filters on a frequency axis but they do "shape" the sound by enhancing frequency components that are in their pass band.

How are vowel sounds recognized by a computer (also, what representation is used)?

First sound is represented using a spectrogram (time-frequency) representation. Vowels are generally tones with tone quality that is adjusted by the formants of the vocal cavities. The person speaking will change the frequencies of the formants in order to make different vowel sounds. Thus, by observing the frequencies of the formants you can identify vowel sounds. These formant frequencies for a given vowel differ a little bit from person to person (and between men and women) so “training” the computer with sound clips of a given voice is usually helpful.