Acoustic phonetics (speech perception) summary

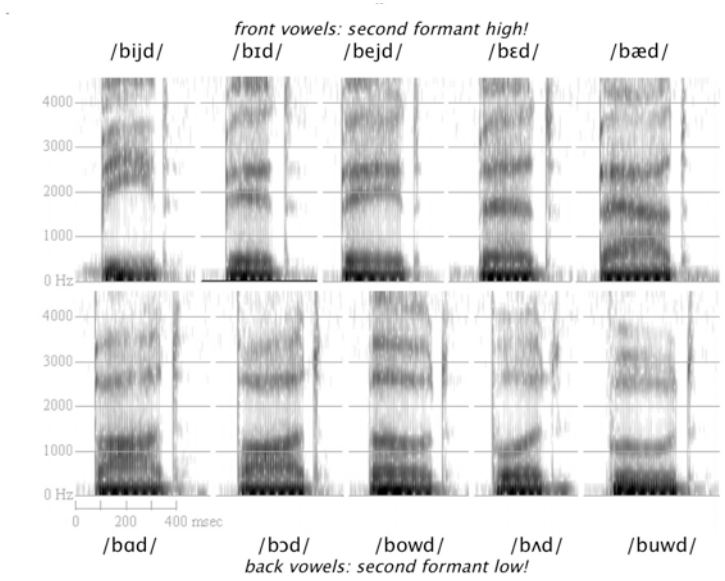
1. Speech Perception

- We hear speech sounds in such a way that we can generally tell how they were pronounced. That is what the "b-ness" of "b" consists in, and the "g-ness" of "g".

- What we pay attention to are the frequency bands that are reinforced, not suppressed, by the resonating chambers of the vocal tract. These are called **formants**.

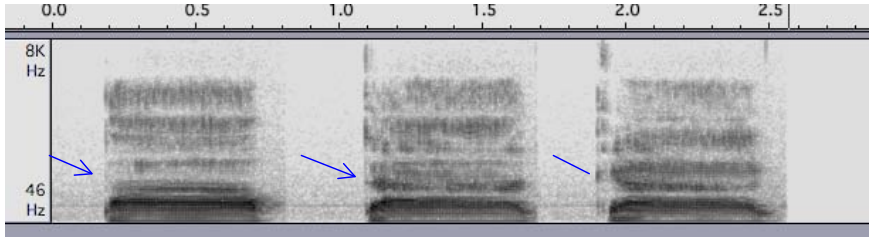
- Vowel formants (except for diphthongs) are steady-state formants.
  The higher the vowel, the lower the first formant.
  The fronter the vowel, the higher the second formant.

(1)



- Consonant sounds, especially stops, involve rapid changes in formant pitch called **transitions** — as you expect given how the vocal tract produces a consonant. The transition is visible in F2 & F3.

(2) **Spectrograms (x=time *y* = frequency)** of ***ba - da - ga*** **(spoken by me!)**
Note the rise (not that clear) for *b*, the shallow fall for *d* and the sharp fall for *g*.



- **Fact**: If a signal *sounds like* a set of formants that could be created by human speech, your mind perceives the signal as speech, even if other properties of the signal are odd. The duck call web site shows this for vowels. Go to
  http://www.exploratorium.edu/exhibits/vocal_vowels/vocal_vowels.html

- In fact, an extremely reduced formant structure can still be perceived as speech. That was the point of the presentation of **sinewave speech.** Go to **http://www.haskins.yale.edu/research/sws.html** for more.

A relevant slogan: ***Speech perception is in your mind, not in your ears.***

2. Categorical perception

- Many phonetically important distinctions lie along a *continuum* of acoustic possibilities.

  For example, one can artifically produce (say, using computer software) the appropriate formant stucture for the syllable *ba*, then gradually change the initial transition through *da* to *ga*. In principle, one might expect to be able to produce a series of signals that gradually morphs from *ba* to *da* then *ga*,
- That is not what we perceive, though. Instead of a gradual shift from *ba* to *da*, and from *da* to *ga*, we hear a bunch of *ba*s, then a bunch of *da*s, then a bunch of *ga*s. That is, each stimulus is assigned to one of the three categories, and we hear nothing as in between two categories. This phenomenon is called **categorical perception**.

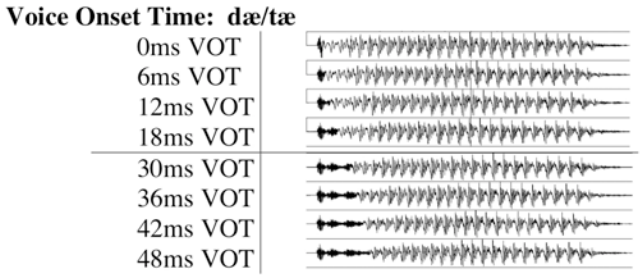
  In class, we saw this by looking at and playing the stimuli at

http://www.haskins.yale.edu/featured/bdg.php?audio=AIFF#

- **To refresh your memory, and because it is very neat — and because you should learn this! (hint) — I recommend you go there too.** Clicking on each stimulus will play it, and there is also a link to hear the entire series.

Other phonetic distinctions that show categorical perception include: voicing distinctions due to changes in Voice Onset Time (VOT).

(3)



- **Innateness:** The phenomenon of categorical perception for speech sounds appears to be **innate in humans**. (I'm not sure I remembered to tell you about this, but it will come up when we discuss the phoneme, and also when we discuss language acquisition later in the semester.) In a famous experiment, Eimas et al. (1971) showed that infants 1 month old have categorical perception for VOT [http://www.sciencemag.org/cgi/content/abstract/171/3968/303?ck=nck].

  Method: Infant's sucking on a nipple increases in amplitude and frequency in response to novel stimuli, then tapers off as the stimulus becomes "old". This allows one to see what sorts of distinctions are perceived as new and which are not, e.g. when varying VOT in an otherwise constant phonetic stimulus.

- **Language acquisition as *unlearning*:** Other work has shown similar categorical effect in infants for the **l/r distinction**. Strikingly, all infants appear to have categorical perception of this distinction, but adult speakers of languages in which l vs. r is not phonemic have lost categoricity! In this respect, language acquisition looks more like unlearning than like learning.

- **Unique to humans?** Kuhl and Miller (1975): Chinchillas trained to respond to a phonetic stimulus spontaneously exhibit a categorical *pa/ba* distinction.

  Method: Choose sounds that are equidistant from each other in VOT, e,g. at 15 ms, 45 ms, and 75 ms VOT. Pick two sounds in the *pa* category and one in the *ba* category. Train a chinchilla to run to the other side of the cage when it hears the central sound (45 ms), heard as /pa/ by English speaking humans. After training, researchers played the two sounds at the extreme ends (15 and 75 ms), which humans hear one sound as *ba* and *pa*, respectively.

  Outcome: Chinchilla ran when they heard the other English *pa*, but not when they heard English *ba*. Apparently chinchillas (who know no English) categorize these human speech sounds in a manner suggestive of human abilities.

- *Possible conclusion:* Categorical perception for speech sounds relies on innate properties of the mammalian auditory system not unique to humans. For example, there may be limits to our ability to resolve timing differences that play a role. Languages can capitalize on these properties and use them to distinguish speech sounds

---

3. McGurk effect

- **Methodology:** Show a video image of a person producing *ga*, with a soundtrack *ba* — so the visual image and the sound contradict each other with respect to the consonant, but are otherwise synchronized.

  Outcome: Most speakers will perceive an acoustic compromise between the video and the audio in the form of *da*! (I screwed this up when I stated it in class, and was corrected by Gillian.)

- Watch it at: http://www.youtube.com/watch?v=aFPtc8BVdJk. If that link gives you trouble, you can find the same movie all over the web. Google for "McGurk effect" and you will find it.

- The effect was discovered and first presented in:

  McGurk, Harry; and MacDonald, John (1976); "Hearing lips and seeing voices," *Nature*, Vol 264(5588), pp. 746–748

  — and is most often known as the ***McGurk Effect.***

- Why is this interesting to us? Well, it is our third demonstration of the very same slogan:

Speech perception is in your mind, not in your ears.

- And of course this slogan is a good characterization of almost all of language — from word boundaries to syntactic structures. It *is* all in your head.

24.900 Introduction to Linguistics

Fall 2012