Human speech perception

Ling 205
What goes on in your mind? Standard conception

- Converting continuous acoustic signal ...
- to a sequence of discrete units, perhaps like IPA symbols (perception)
- which can then be matched up with entries in the mental lexicon (recognition) from which meaning can be computed.
What are the units of perception?

- syllables
- segments (i.e. consonants and vowels) as allophones or phonemes
- phonetic features, e.g. voiced, nasal, fricative, etc.
- regardless, there is an assumption of linearity: the units must occur in sequence
Linearity problem

• But acoustic properties are not always linear, localizable

• For example, in the sentence 'Benny saw her', the lowering of F3 due to the 'r' in 'her' may extend backward in the phrase.
  
  – i.e. the F3 of [ɛ] in 'Benny' is typically slightly lower in 'Benny saw her' than in 'Benny saw him'.
General research program of perceptual phonetics

- Focused on experimental studies
- Presenting audio stimuli to human subjects (often synthetic speech, allows for controlled manipulation of particular aspects of the stimuli)
- Measuring subjects' responses to various tasks
Vowel perception

• Basically has to do with frequency of F1, F2 and (to some extent) F3, but
  – Speaker variation: Formant frequencies vary depending on length of the speaker's vocal tract, particularly for men vs. women.
    • There appears to be a gender effect on top of this anatomical difference: men learn to have 'manly vowels', women learn 'womanly vowels'.
  – Phonetic context variation: Formant frequencies are affected by nearby sounds, e.g. [u] in 'toot' has higher F2 than in [kook]
  – Speech rate variation: in faster speech vowels not only shorten in duration, but also move closer to [ɨ], raise and centralize (undershoot).
Vowel perception: normalization

- Some sort of transposition of speaker's vowel space to a 'normal' vowel space in the mind of the listener.
- Extrinsic: listeners first get a sense of the size of the speaker's vowel space and use this to calibrate the categorization of particular vowels.
- Intrinsic: using information in the vowel itself, e.g. F0, and auditory scaling to better differentiate vowels.
Alternative: no normalization

- Exemplar approach: store every token of every vowel, including 'indexical' information (speaker characteristics), to build up vowel categories.
- An incoming male vowel can be compared to all the stored tokens of male vowels, and matched to the 'winning' category.
- Perhaps give added weight to tokens from similar phonetic context, or similar speech rate.
- Can handle a rich range of sociophonetic effects, e.g. dialect variation (consider New Zealand vowels).
- Can be computationally very expensive.
  - But human brain computation is massively parallel, unlike today's electronic computers which are predominantly serial. Parallelization could make exemplar approach very feasible.
Consonant perception

- More complex
Approximants

- [j] vs. [w] categorization principal cue is onset F2 frequency.
- [ɹ] vs. [l] cue is F3 freq.
- [b] vs. [w] is duration of F2 transition.
- Notion of 'category boundary'
  - boundary can shift depending on various factors, e.g. speech rate, and actual word status, e.g. 'yellow' vs. 'wellow' (non-word), j/w boundary shifts towards [w].
- We rarely categorize based on a single cue
  - clusters of cues are 'integrated', giving more weight to some, less weight to others, but taking all of them into account.
  - the sets of cues, and their relative weightings, have to be learned on a language-specific basis.
Fricatives

- Sibilants vs. non-sibilants distinguished by overall noise intensity
- Also frequency of spectral peak.
Nasals

- skip
Stops

- Place of articulation: basically depends on F2 frequency, but this is relative to the following vowel.

- Best predictor: locus equations
  - $F_{2c} = k(F_{2v} - L) + L$
  - where $k$ and $L$ depend on the consonant place of articulation
  - categorize the incoming $F_{2c}$ according to the place of articulation that gives the closest match.
Theories of speech perception

- Motor theory
- Auditory theories
- “Double-weak” theory
Motor theory

- Assumes perception involves a mental transformation of acoustic signal into articulatory 'space'
- in hopes that this will eliminate or reduce the variation in the signal
- i.e. the units of perception are articulatorily defined
- but if there is acoustic variation, there must be articulatory variation that's causing it.
- vague as to how this acoustic -> articulatory transduction is done.
- what are we doing when we perceive the speech of parrots?
Categorical perception

- Perceptual studies frequently show category boundary behaviour
- Within category variation doesn't affect response as much as variation near and across the boundary
Is speech special?

- Motor theory says yes
- Exemplar theory says no