Big issues in speech perception

- Categorical Perception (already covered)
- Lack of invariance
  - Coarticulation
  - Context effects
- Constancy (dealing with variation in both time and talker)
- Segmentation
- Speech "specialness"

Sources of variability

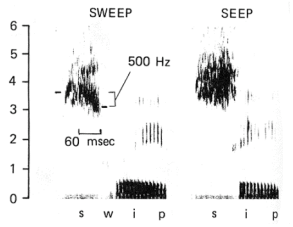
- Coarticulation
- Undershoot
- Cue trading
- Talker differences
- Speaking rate

Coarticulation

Examples:

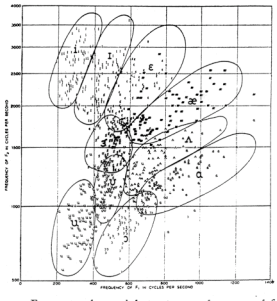
SHOE vs. SHIP

SWEEP vs SEEP



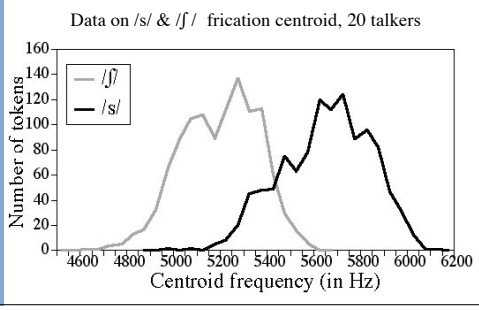
Hockett's analogy

Talker differences

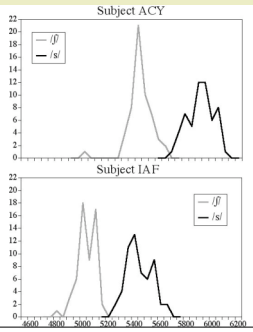


Frequency of second formant *versus* frequency of first formant for ten vowels by 76 speakers.

Talker variability - /s/ & /ʃ/



Talker variability - /s/ & /ʃ/



Adjusting for variability

- Mullennix, Pisoni, and Martin (1989)
  - Naming task, with trial-to-trial variation in talker
  - Identification was more accurate and naming was faster for a single-talker condition
- Magnuson, Yamada & Nusbaum (1994)
  - Same results when voices are spouse & children
- Sommers, Nygaard, and Pisoni (1994)
  - Similar decrements for rate variability, but not amplitude variability

Effects of variability on memory

- Palmeri (1993)
  - Continuous recognition memory task (had to decide if each word was old or new)
  - Better performance if word was spoken in same voice (so voice was stored)
  - Effect of voice was irregardless of gender match
  - Increasing # of voices (2-20) had little effect
- Sheffert & Fowler (1995)
  - Did A/V version of this; voice mattered, but only a trend for visual face

Working memory/processing effects

- Martin, Mullennix, Pisoni & Sommers (1989)
  - Serial recall test
  - Early list items were recalled better for lists spoken by a single talker (weakened primacy for multiple talkers)
  - Suggested that processing lists by multiple talkers required more resources
- Goldinger (1991)
  - But if presentation is slowed (so you have those resources), recall was better in multiple-talker lists
- Nygaard, Sommers, & Pisoni (1995)
  - At short ISIs, both talker and rate, but not amplitude, variability led to processing costs
  - At long ISIs, only talker led to an advantage

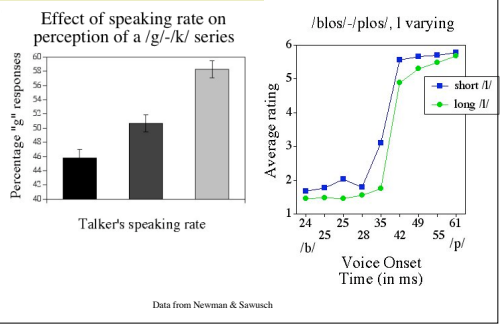
Clinical populations

- Sommers, Kirk & Pisoni (1997)
  - Word recognition tasks with NH, CI, and normal-hearing-noise-masked (NHNM)
  - Talker effects were only in open-set tasks (not closed-set tasks), so they're more sensitive
  - Similar talker effects for all groups, so talker adjustment is an issue for CI, despite impoverished stimuli

Clinical populations

- Sommers (1997)
  - NHY, NHE, & HIE (hearing-impaired elderly)
  - Age effects: greater effects of variability for NHE than NHY (talker & amplitude)
  - Hearing impairment effects: greater effects of variability for HIE than NHE (rate & talker)
- Adjusting for variation requires cognitive resources, which may be why it is particularly problematic for older individuals & those with hearing impairments

Speaking rate



Data from Newman & Sawusch

Speaking rate adjustment

- Miller & Liberman (1979)
  - A longer following vowel altered the location of a b-w boundary
- Kidd (1989); Summerfield (1981)
  - Precursor sentence duration similarly alters perception
  - There seems to be an effect of the immediately preceding phonemes (a local effect) and a more distal effect based on rhythmic effects

Obligatory processing

- Miller & Dexter (1988)
  - In a /bi/-/pi/ series, listeners who responded slowly had a rate effect; those who responded quickly did not
  - But, fast responders acted as if the vowel was physically short (as if it ended when they responded) - their boundary was the same as for a short vowel, not the same as for a "mid-length" vowel)
  - Even though they didn't know what the vowel-length would be, they couldn't help processing on the basis of what they had so far

Following rate effects

- Newman & Sawusch 1996
  - Phoneme identity is irrelevant
    - effects aren't larger for more similar phonemes - chaes-shaes series)
    - Vowel isn't particularly privileged (blos-plos series)
    - Legality/Phonotactics doesn't matter
  - It's all about distance: a temporal window of about 250 ms

Speech specificity

- Rate normalization occurs for tone analogs (Pisoni, Carrell & Gans (1983) to Miller & Liberman's study
  - Others argue that parallel findings do not imply similar mechanisms; Fowler 1990
- Similar effects occur for nonspeech stimuli that can't be heard as speech (Diehl & Walsh, 1989)
- Suggests an early auditory process, not a language-specific process