# Speech Perception

Ben Weissman

COGS 4780

# Speech Perception phenomena

- Top-down effects
  - When ambiguous /d-t/ sounds appear within words, participants exhibit a bias towards the sound that makes a real word (Ganong, 1980)
    - e.g., __ash vs. __ask
  - **Phoneme restoration effect** – when a speech sound is artificially replaced with a non-speech sound (Warren 1970)…

  [s] in *legislature* replaced by a coughing sound

  … participants "hear" the /s/ sound as if it were really there, along with the cough

# Speech Perception phenomena

- Cross-modal cues:
  - McGurk effect
  - McGurk effect with nonwords (Dekle et al., 1992)
  - aero-tactile integration (Gick & Derrick, 2009)
  - Tactile McGurk effect (Fowler & Dekle, 1991) – *ba* vs. *ga*

any case, relevant parts of her face were covered by the subject's hand.) The subject stood with his or her right hand in a disposable glove placed over the lips of the model. After some piloting, we found that subjects were most successful if we placed the forefinger on the upper lip and the next finger on the lower lip of the model. This is not the procedure used by Tadoma perceivers (who generally place a thumb vertically against the lips); however, it allowed subjects to distinguish open from closed lips. After presentation of paired felt and heard syllables, subjects indicated to a second experimenter which syllable ("ba" or "ga") they had heard and then which syllable they had felt, in each case guessing if necessary. They made their responses by pointing to printed syllables on an 8½ × 11-in. (21.59 × 27.94 cm) sheet of paper held on a clipboard by a second experimenter (Dekle). On the sheet of paper, the printed syllables "BA" and "GA" appeared twice: on the left under the heading "heard" and on the right under the heading "felt." The second experimenter then marked an answer sheet analogous to those used in the orthographic condition (i.e., with response columns "heard" and "felt"). As in the ortho-

# Speech Perception cues

- Not all input in speech perception is treated equally
- The articulatory phonetics perspective simplifies things (voicing, manner, place)
  - [p] vs. [b] differ in voicing and aspiration
- From an acoustic perspective, there are lots of different cues
- *rapid* vs. *rabid*
  - VOT, preceding vowel length, lip closure duration, preceding pitch contour, following pitch contour, etc.
  - potentially up to 16 different cues

# Speech Perception cues

- Not all input in speech perception is treated equally
- Listeners engage in **cue weighting** and we learn to pay attention to some cues over others
  - Some cues are naturally more informative than others, and we learn which cues are the most productive via statistical learning
    - Can differ from language to language (VOT is more productive in English than in Korean, for example)
  - Some cues are more noticeable than others, and we learn to pay attention to those when useful
  - Some cues are more susceptible to noise than others, so we learn to focus on the less variable ones

# Speech Perception cues

- The way one listener picks up on speech cues can even vary from speaker to speaker
- Nygaard & Pisoni (1998) trained participants on sentences from 10 different speakers, labeling the speakers' recordings with names and playing these recordings over a 3 day period
- After 3 training sessions, some participants heard new sentences from those speakers, but the sentences were mixed with white noise
- Some heard new sentences (again noise-mixed) from other speakers that they hadn't yet heard

# Speech Perception cues

- The way one listener picks up on speech cues can even vary from speaker to speaker
- Nygaard & Pisoni (1998) trained participants on sentences from 10 different speakers, labeling the speakers' recordings with names and playing these recordings over a 3 day period
- After 3 training sessions, some participants heard new sentences from those speakers, but the sentences were mixed with white noise
- Some heard new sentences (again noise-mixed) from other speakers that they hadn't yet heard
- Participants better at deciphering speech from familiar voices

# Subtitles vs. Dubs

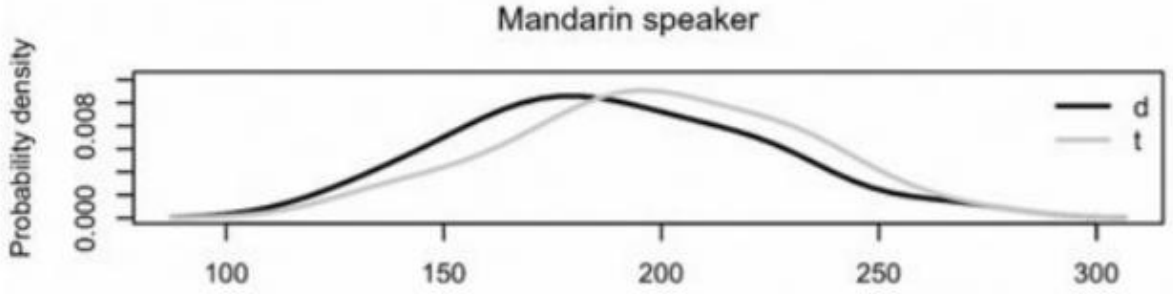
- Discuss from a speech perception and language processing perspective:
- What are the pros and cons of using subtitles when watching a foreign language show?
- What are the pros and cons of using dubbed voices when watching a foreign language show?

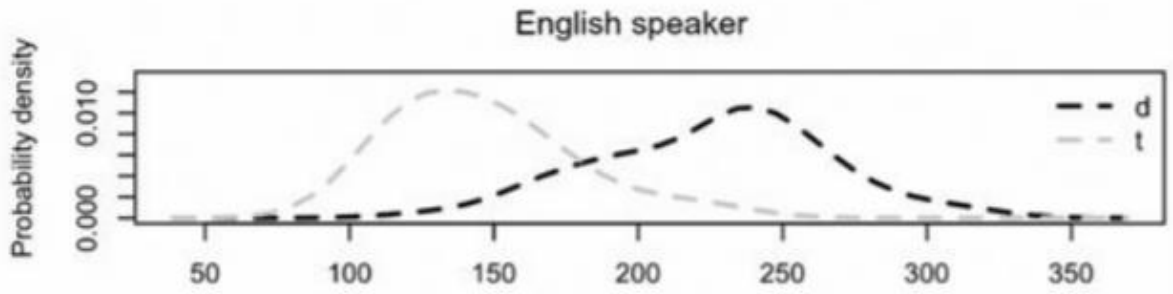
# Speech Perception cues

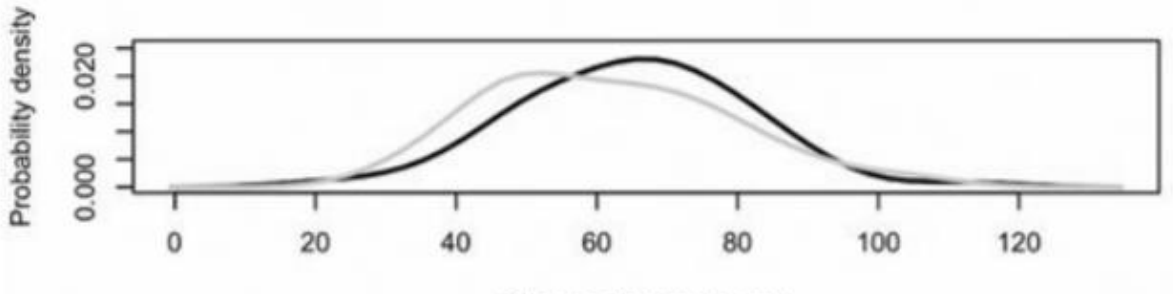
- Some cues are linked to social factors (e.g., speaker gender, regional accent)
  - First formant (F1) has, on average, different distributions for male and female speakers
  - VOT does not
  - So F1 is used as a cue for speaker gender and VOT is not
- Familiarity with certain regional accents makes you better at deciphering speech from those accents

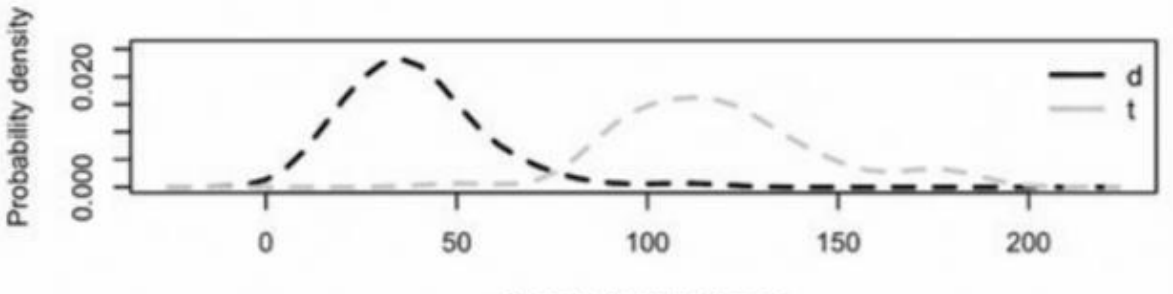
# Speech Perception cues

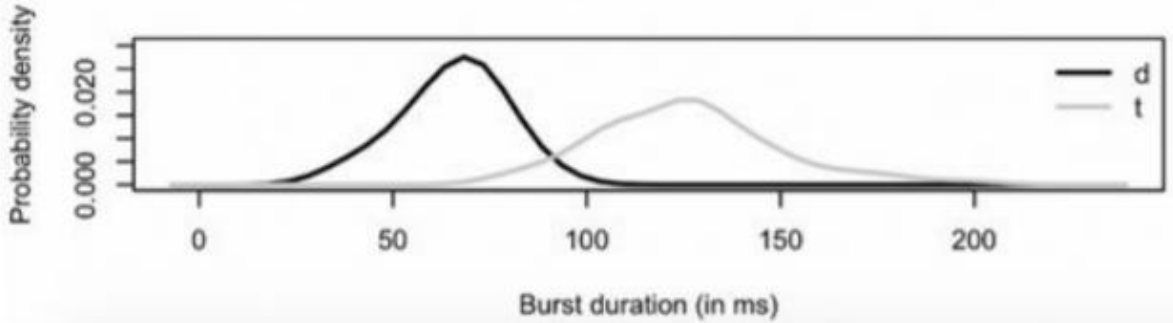
- Familiarity with idiosyncratic foreign accents makes you better at deciphering speech from that speaker
  - Xie et al. (2017) trained English-speaking participants on a Mandarin-accented speaker
  - Speaker did not use the typical English cues to distinguish /d/ and /t/ (VOT, preceding vowel length, closure duration) but used a different cue (burst duration)
  - Participants trained with just 60 words of this speech learned to rely on burst duration

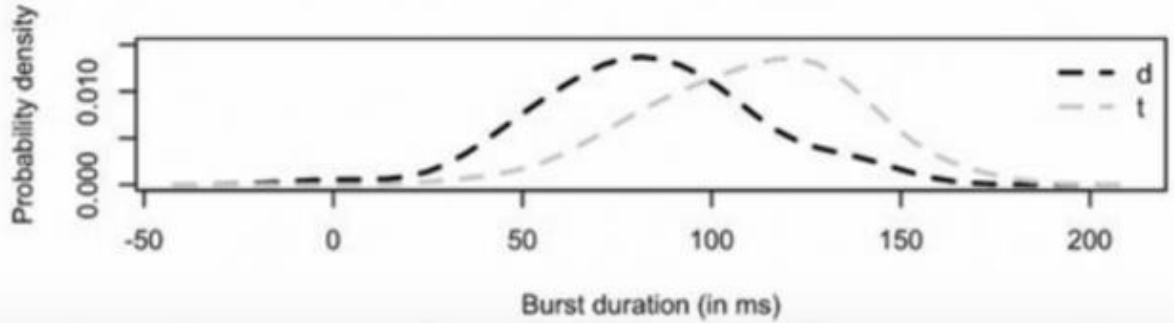

Mandarin speaker
Probability density
0.000
0.008
100
150
200
250
300
Vowel duration (in ms)
d
t


English speaker
Probability density
0.000
0.010
50
100
150
200
250
300
350
Vowel duration (in ms)
d
t


Probability density
0.000
0.020
0
20
40
60
80
100
120
Closure duration (in ms)
d
t

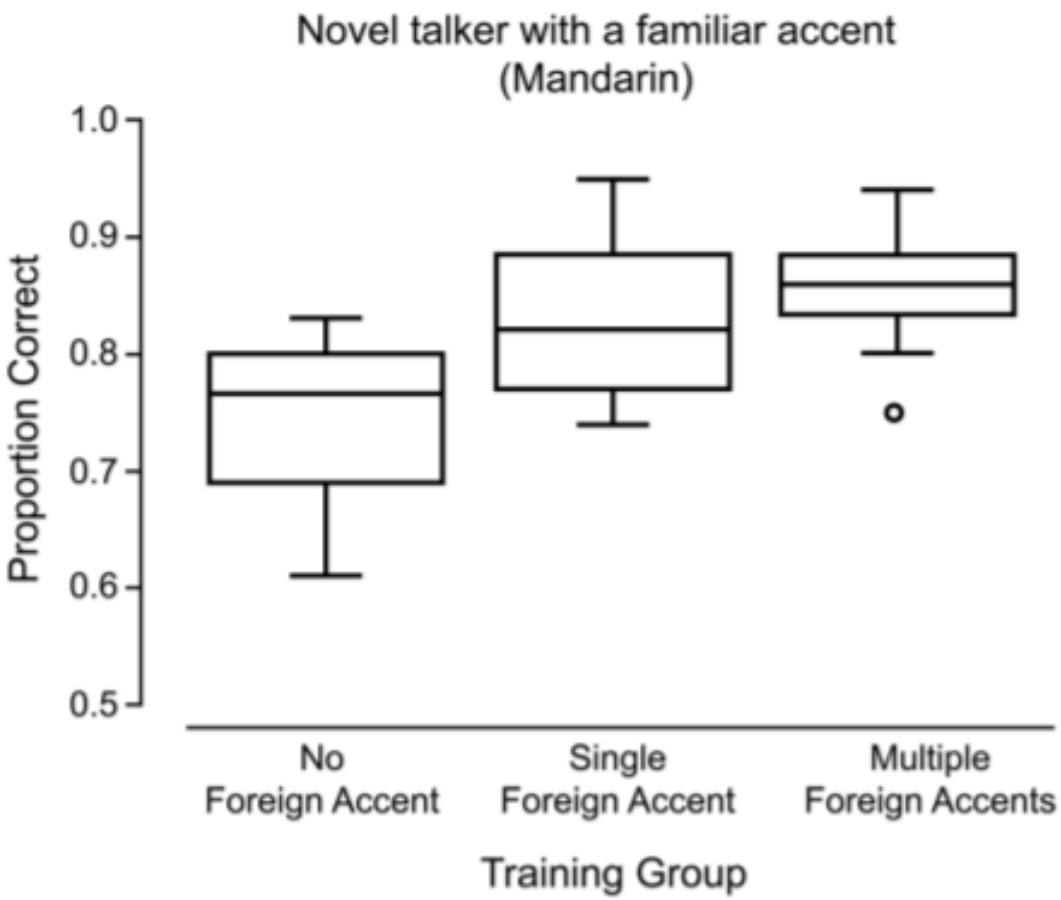

Probability density
0.000
0.020
0
50
100
150
200
Closure duration (in ms)
d
t

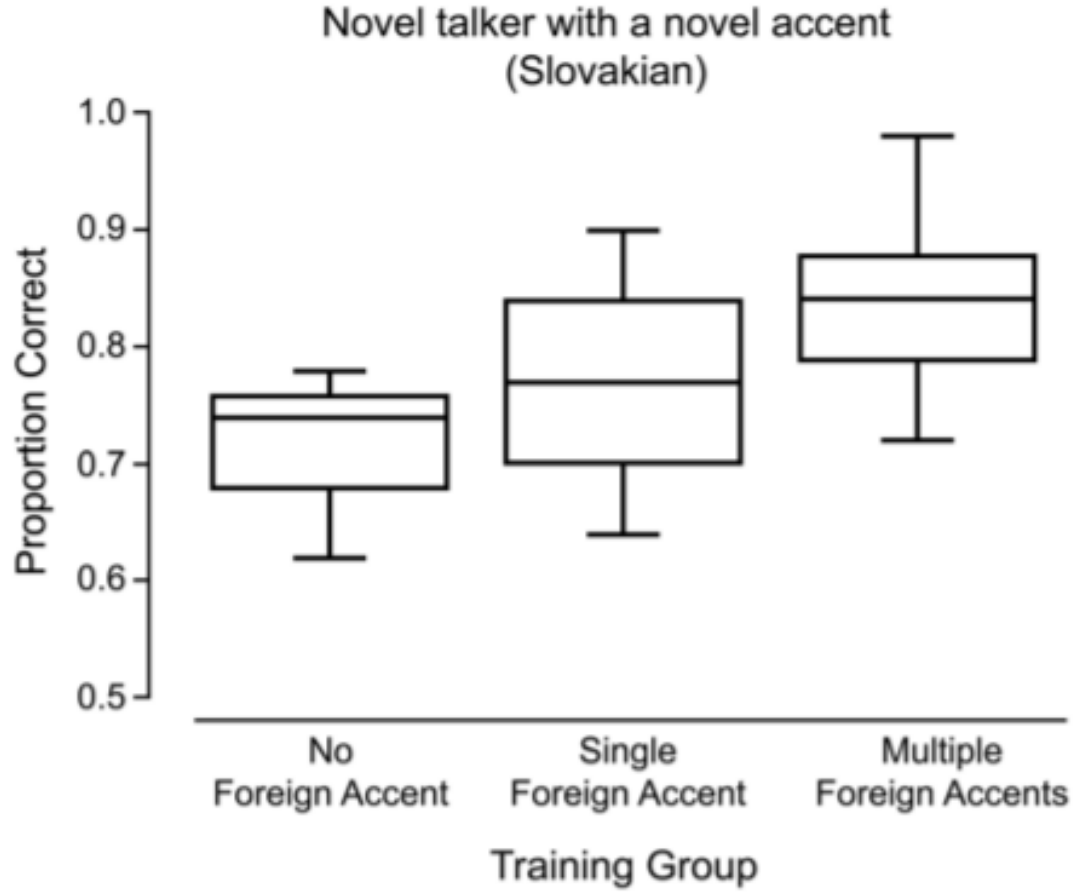

Probability density
0.000
0.020
0
50
100
150
200
Burst duration (in ms)
d
t


Probability density
0.000
0.010
-50
0
50
100
150
200
Burst duration (in ms)
d
t

# Speech Perception cues

- Familiarity with idiosyncratic foreign accents makes you better at deciphering speech from other speakers…..?
  - Baese-Berk et al. (2013) exposed participants to English speech in 5 different foreign accents (Thai, Korean, Hindi, Romanian, Mandarin)
  - Exposure to those accents improved performance on speech from a different Mandarin-accented speaker
  - And even a Slovakian-accented speaker


Novel talker with a familiar accent
(Mandarin)
Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
No
Foreign Accent
Single
Foreign Accent
Multiple
Foreign Accents
Training Group


Novel talker with a novel accent
(Slovakian)
Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
No
Foreign Accent
Single
Foreign Accent
Multiple
Foreign Accents
Training Group

# Speech Perception cues

- Familiarity with idiosyncratic foreign accents makes you better at deciphering speech from other speakers…..?
  - Baese-Berk et al. (2013) exposed participants to English speech in 5 different foreign accents (Thai, Korean, Hindi, Romanian, Mandarin)
  - Exposure to those accents improved performance on speech from a different Mandarin-accented speaker
  - And even a Slovakian-accented speaker

  - **What are two possible explanations for this?**

# Speech Perception cues

- Familiarity with idiosyncratic foreign accents makes you better at deciphering speech from other speakers…..?
  - Baese-Berk et al. (2013) exposed participants to English speech in 5 different foreign accents (Thai, Korean, Hindi, Romanian, Mandarin)
  - Exposure to those accents improved performance on speech from a different Mandarin-accented speaker
  - And even a Slovakian-accented speaker

  - **What are two possible explanations for this?**
    - There's some overlap between the different non-native accents, so exposure to some accents prepares the listener for identification in others
    - Hearing more accents makes participants more generally flexible and perceptive of varying cues

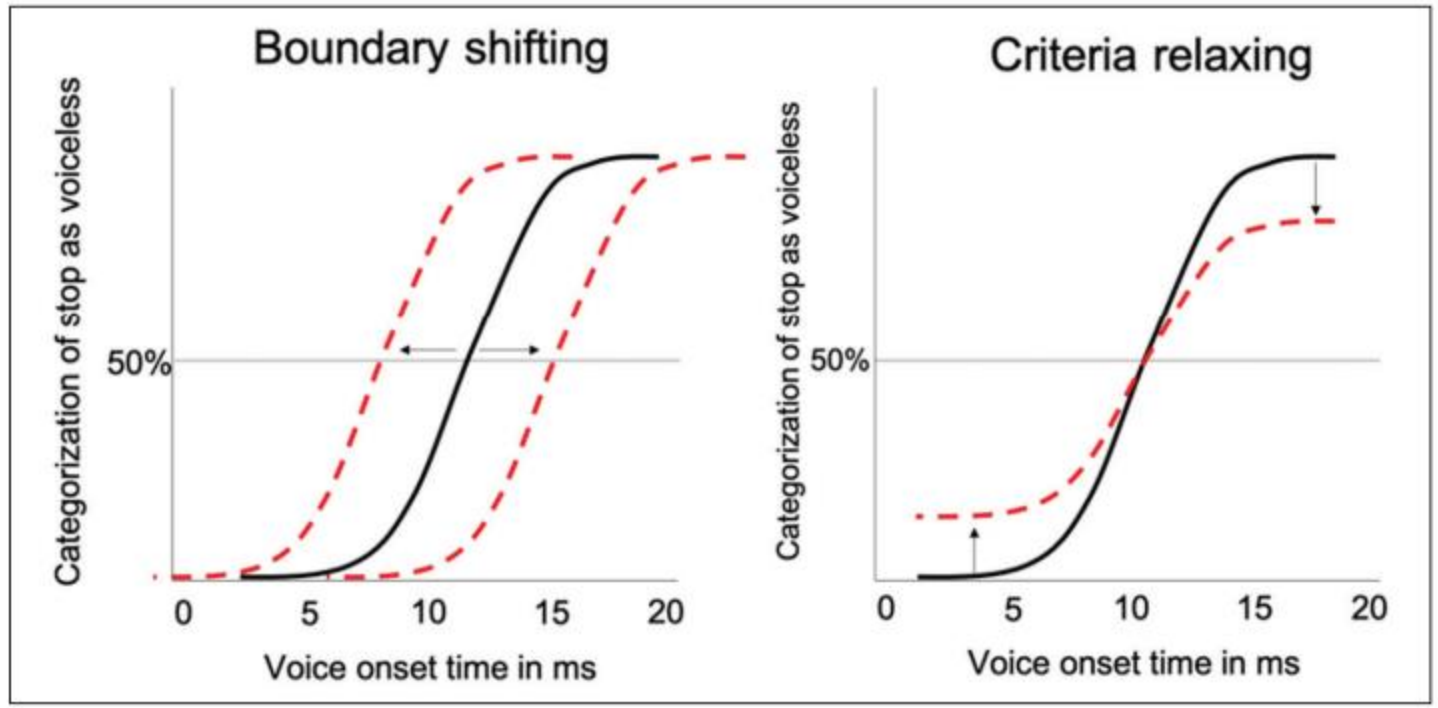



**Figure 1.** Demonstration of boundary shifting (left) and criteria relaxing (right). The solid curve represents pre-retuning categorization of a stop continuum based on voice onset time (VOT). The dotted curves represent how categorization may change after retuning. The point where the curves cross the 50% vertical line represents the category boundary.