Adaptation During Speech Perception

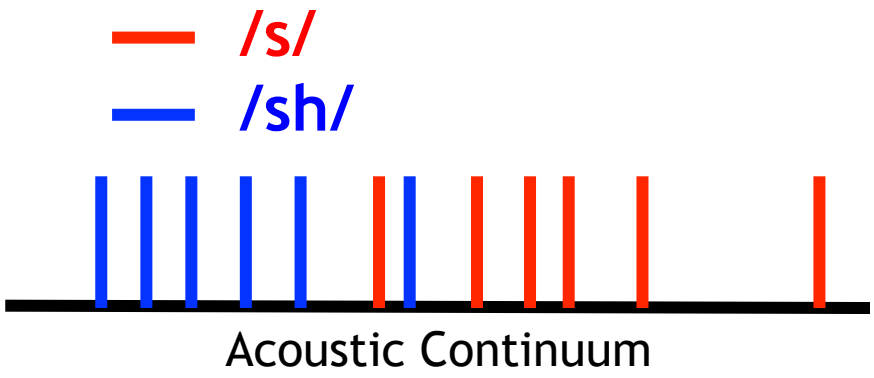
BCS 152

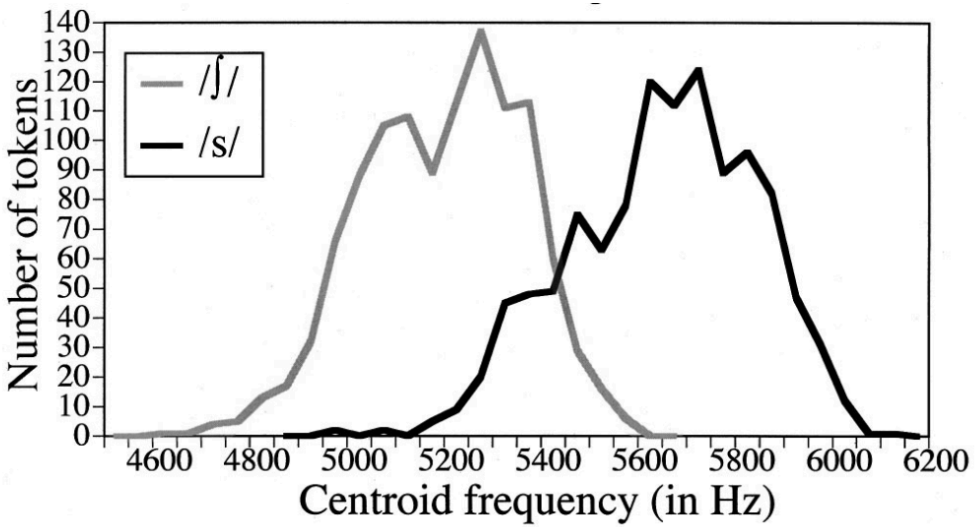
3 October 2018

Wednesday Bushong

Lack of Invariance Problem

- Different instances (***tokens***) of the same phoneme are slightly different





- Given this variability, how can listeners figure out what phonemes they're hearing?

figure from Newman et al. (2001)

But it gets worse...

- Different people are different too!

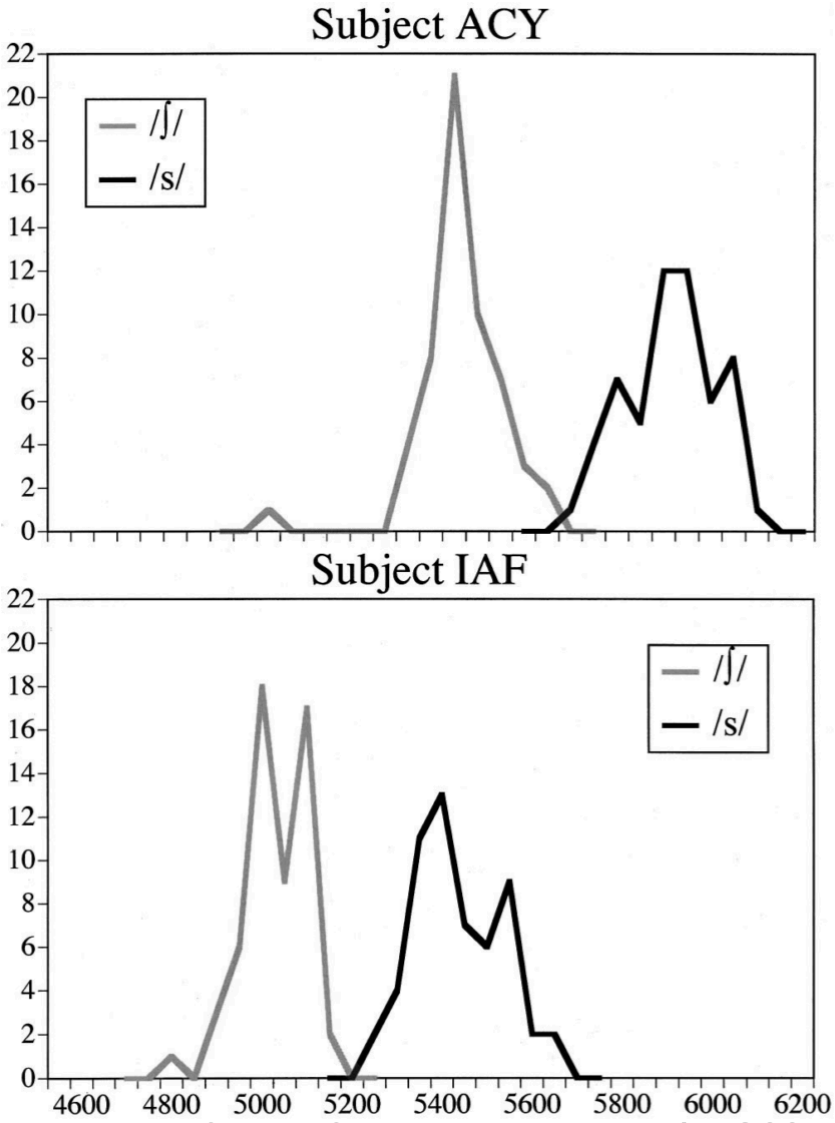


figure from Newman et al. (2001)

Inter-speaker variability

- Pronunciation
- Lexical choices
- Syntactic choices
- Social choices (e.g., politeness)
- Assumptions about common ground
- Etc.

Inter-speaker variability

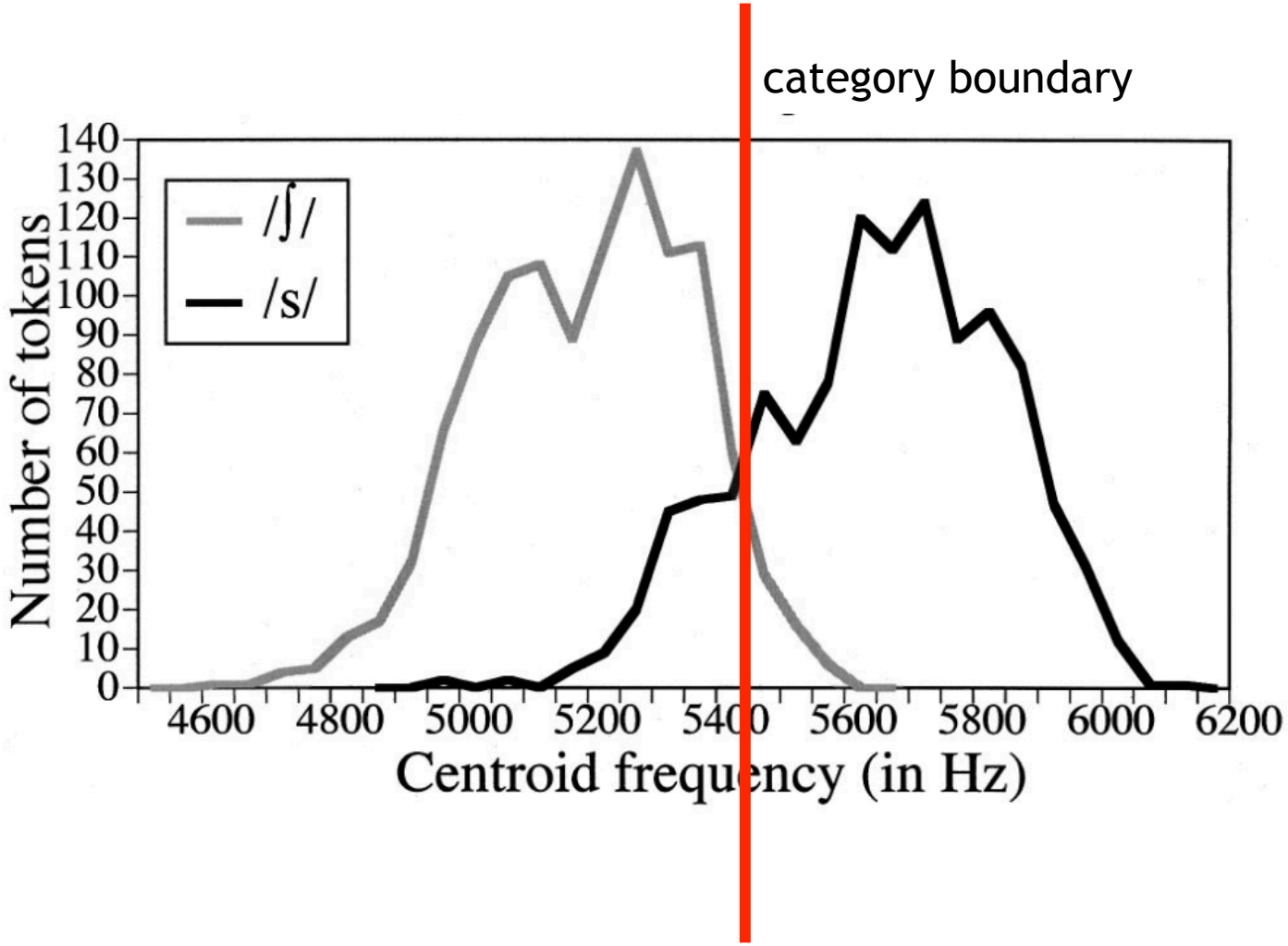
- **Pronunciation**
- Lexical choices
- Syntactic choices
- Social choices (e.g., politeness)
- Assumptions about common ground
- Etc.

Flashback to speech perception lectures

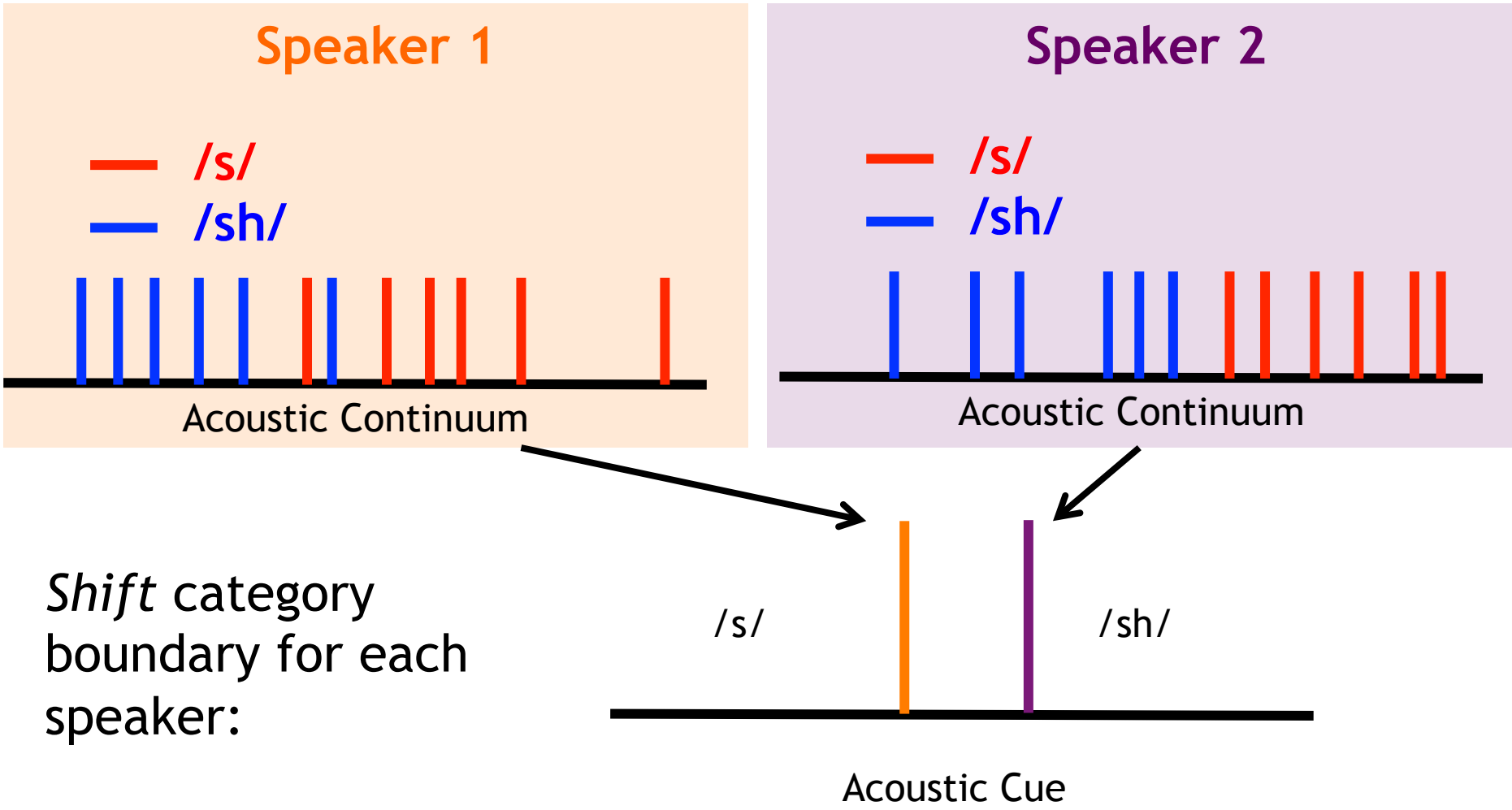
- How do people deal with the lack of invariance?

- Category boundary
- Where does the category boundary come from?
  - Typical distribution of acoustic features along a continuum

All English Speakers



Phonetic Adaptation: adjust your categories to fit a new speaker's productions



Roadmap for next two lectures

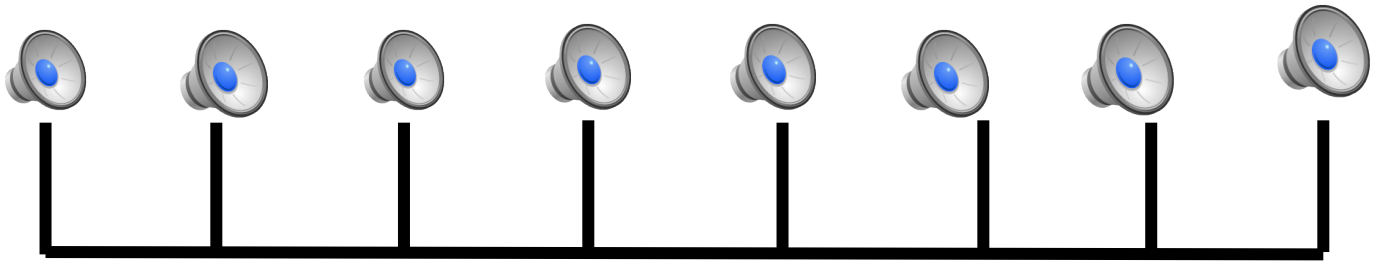
- Basics of speech perception studies
- Adaptation to a new speaker's production of a single sound
- Adaptation to dialects and accents
- Use of prior language & world knowledge during adaptation
- Adaptation as a broader theory of speech perception
- Speech perception & adaptation in the brain

How do we study speech perception?

- Categorization (also known as labeling or identification)
- Discrimination
- Transcription

Categorization Experiments

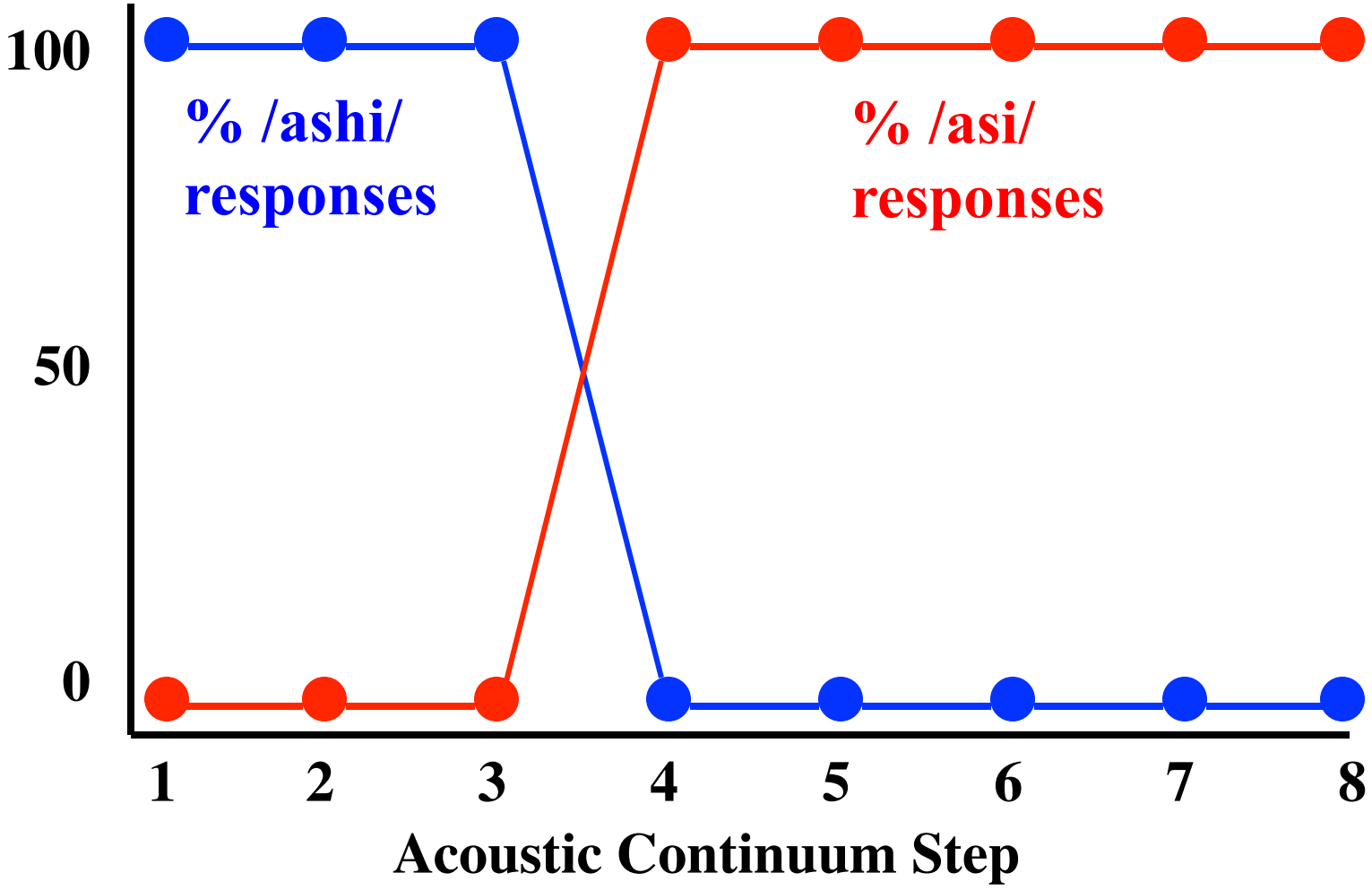
- Give subject a non-word syllable like “a?i”
- Manipulate “?” to vary between two sounds on an acoustic continuum



Acoustic Continuum

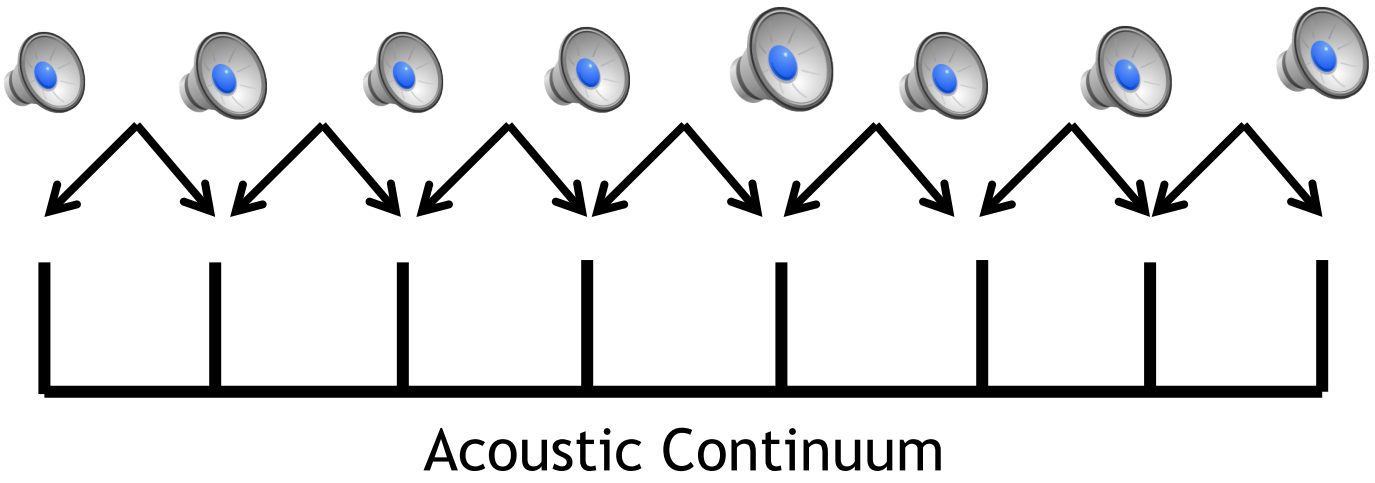
- Task: did you hear “ashi” or “asi”?
- **Dependent variable: % “ashi” or “asi” responses**
- From responses, we can estimate where their category boundary is (50% response probability)

Typical Categorization Data

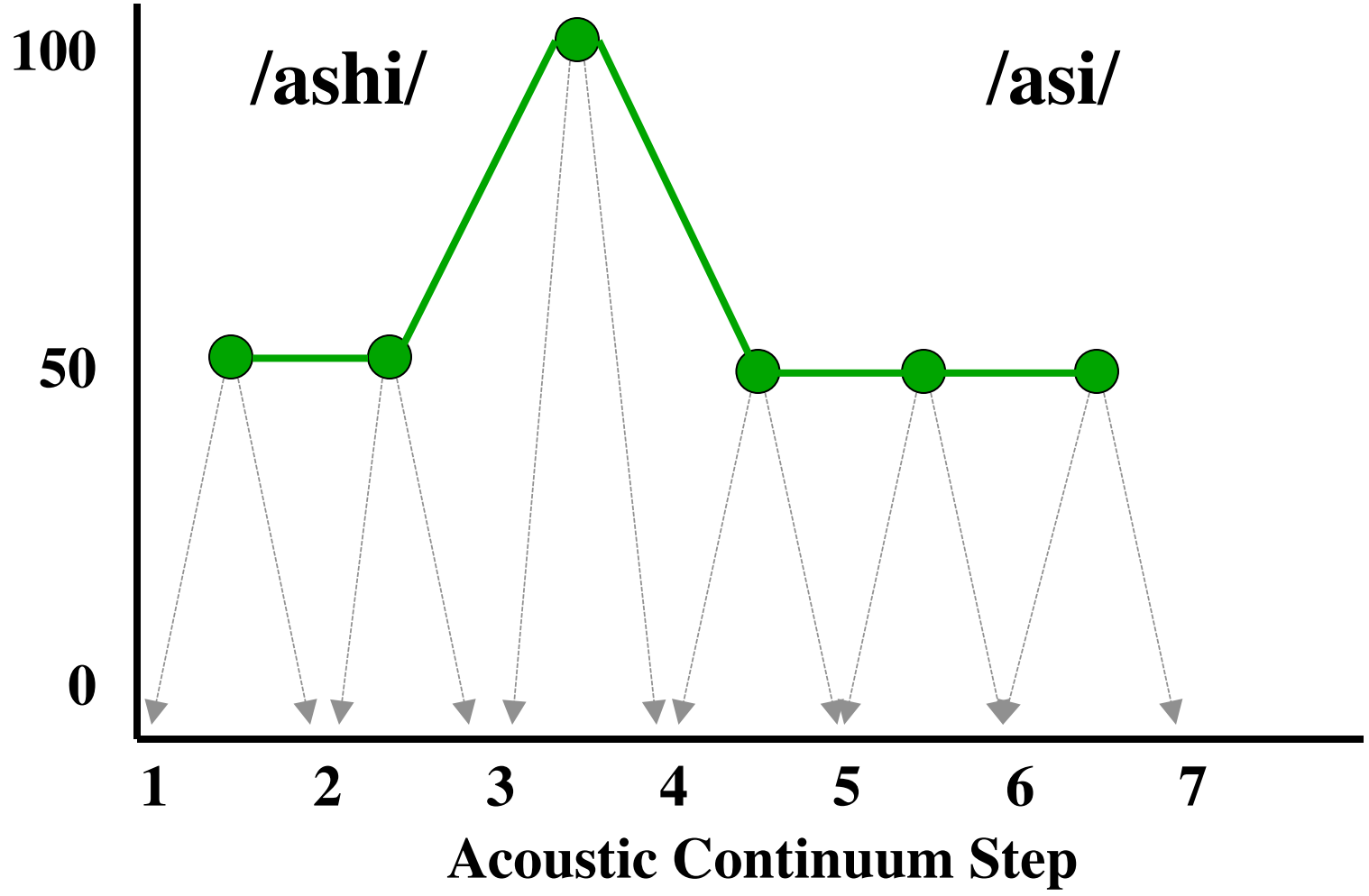


Discrimination Experiments

- Give subject a *pair* of sounds and ask whether they're the same or different
- **Dependent measure: accuracy**



Typical Discrimination Data



Transcription

- Give subject natural speech (single words or full sentences) and ask them to write down what they heard
- **Dependent measures: accuracy, reaction time (RT)**
- Con: not as fine-grained as categorization or discrimination experiments
- Pro: good for getting a general idea of how easily perceived (intelligible) a speaker is
  - used often in experiments on adaptation to foreign speech

Match-Mismatch

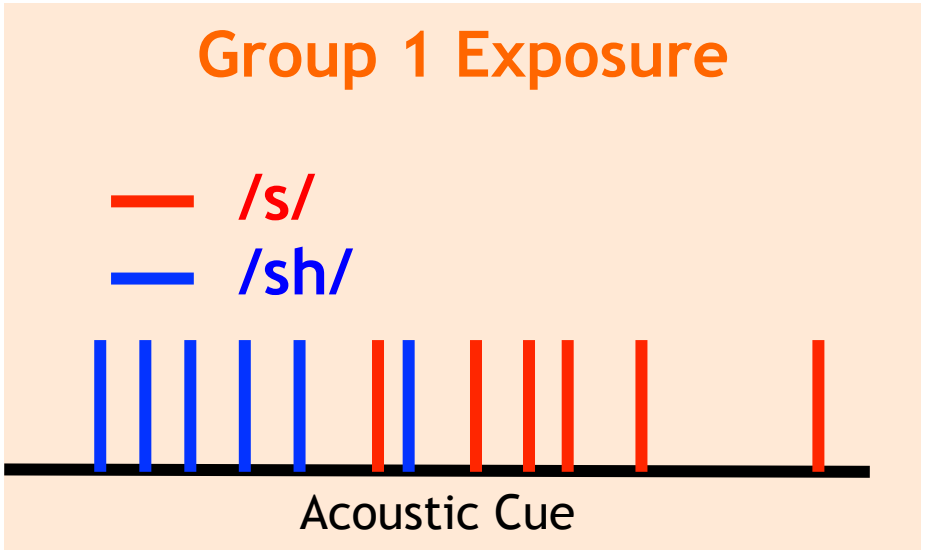
- Like transcription, listen to a fluent sentence
- After sentence, word appears on screen
- Participant judges whether the word matches the last word of the sentence they heard
- **Dependent measures: accuracy, RT**

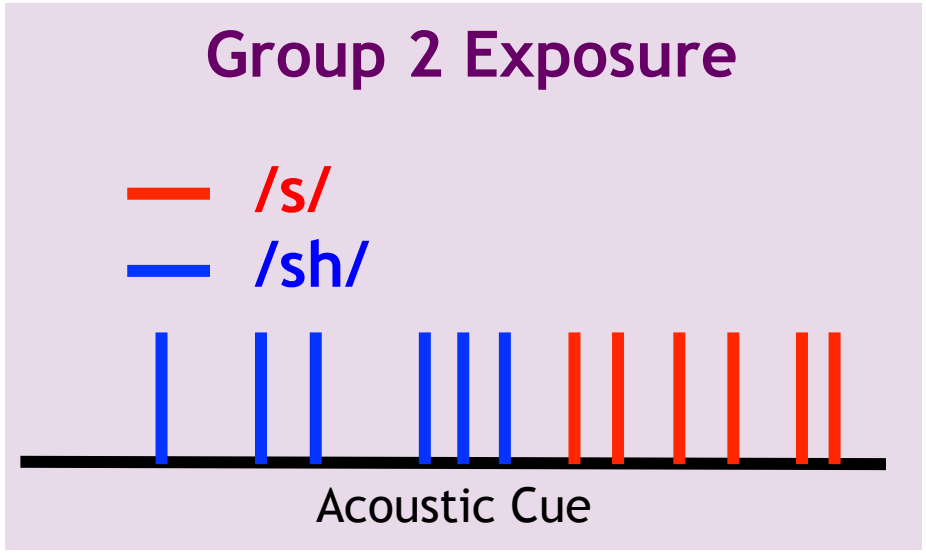
Adaptation to Single Sounds

How do we study this?

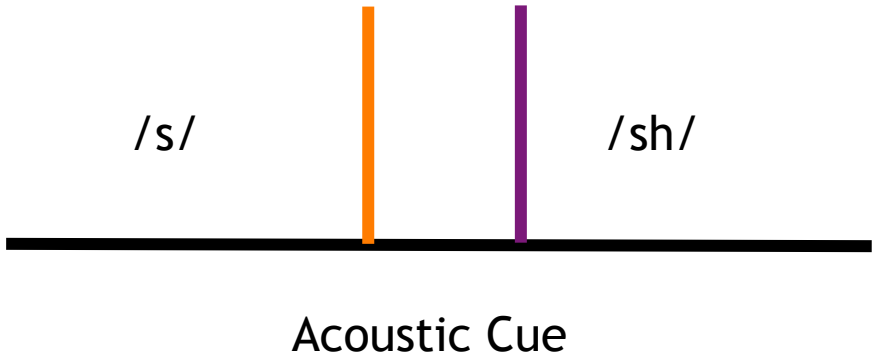
- **Exposure Phase:** expose research participants to sounds on a specific part of an ***acoustic continuum*** that are ***labeled*** as belonging to one category or another
- **Test Phase:** categorization or discrimination task along the entire acoustic continuum

- How do we label?
  - Use a word that doesn't have a minimal pair
  - e.g., /s/ in 'medicine'





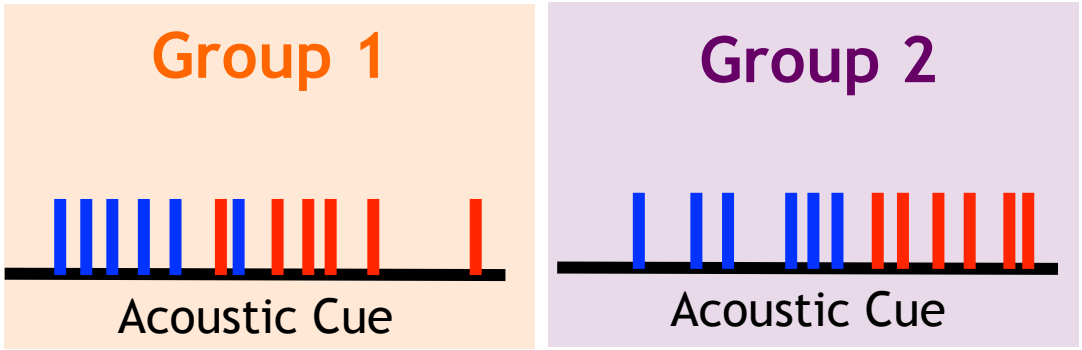
Shift category boundary given exposure:



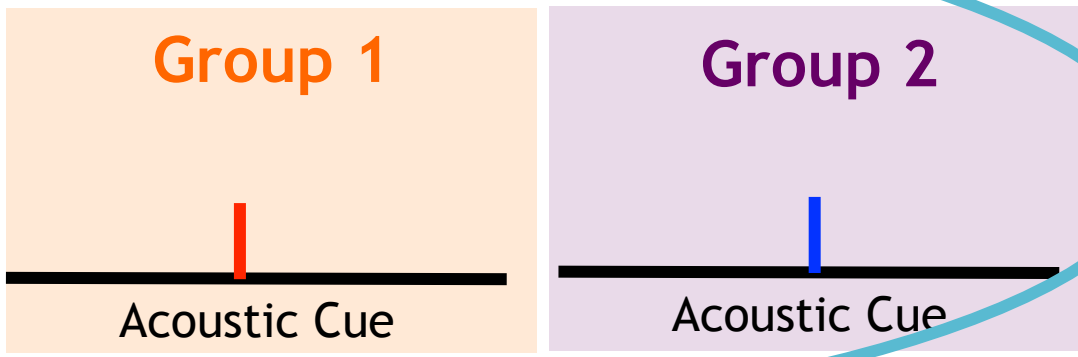
2 Methods of Exposure

/s/
/sh/

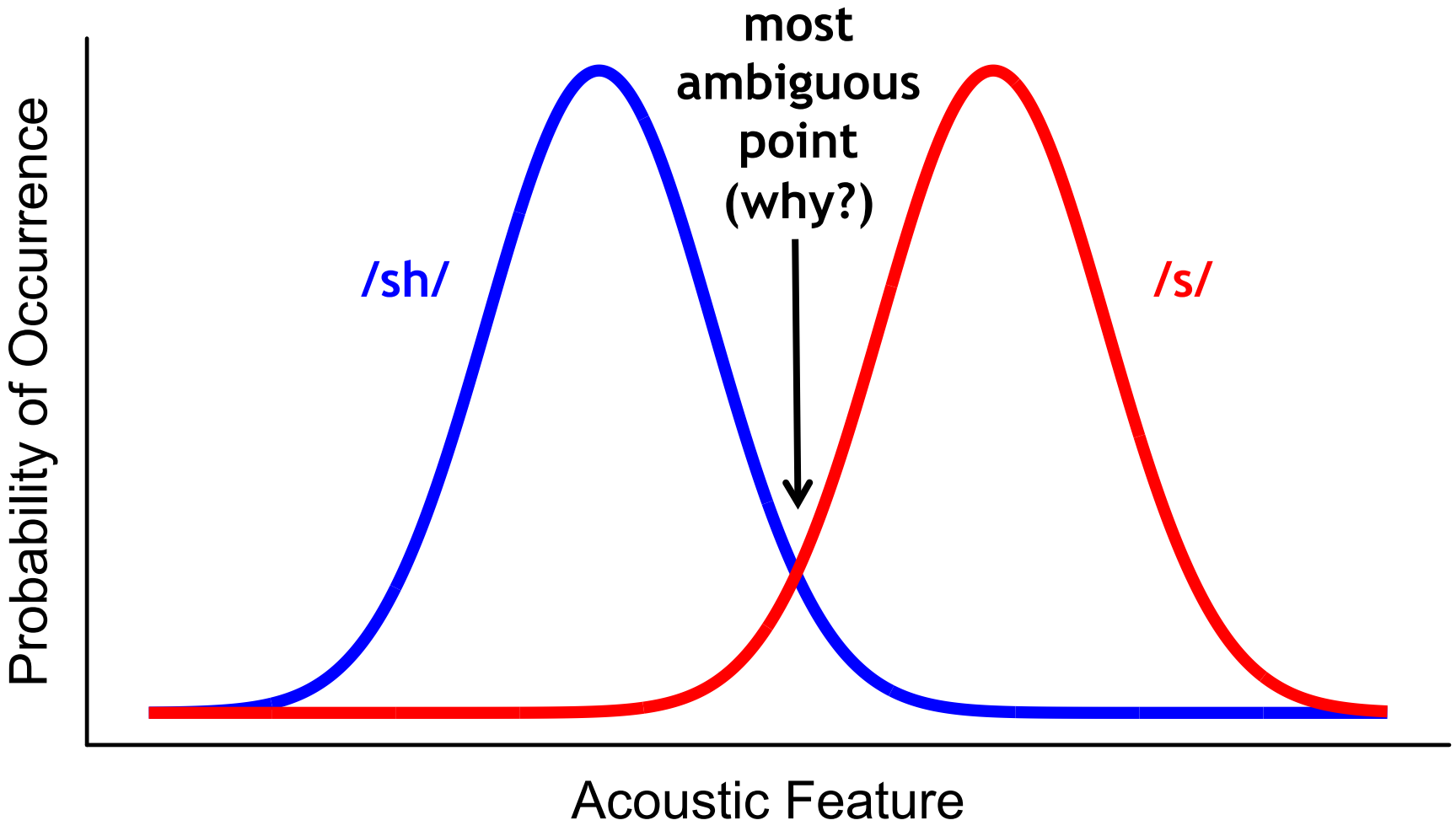
Expose participants to **full distribution** of a speaker's productions



Expose to **targeted point(s)** along acoustic continuum



Average English /s/ and /sh/

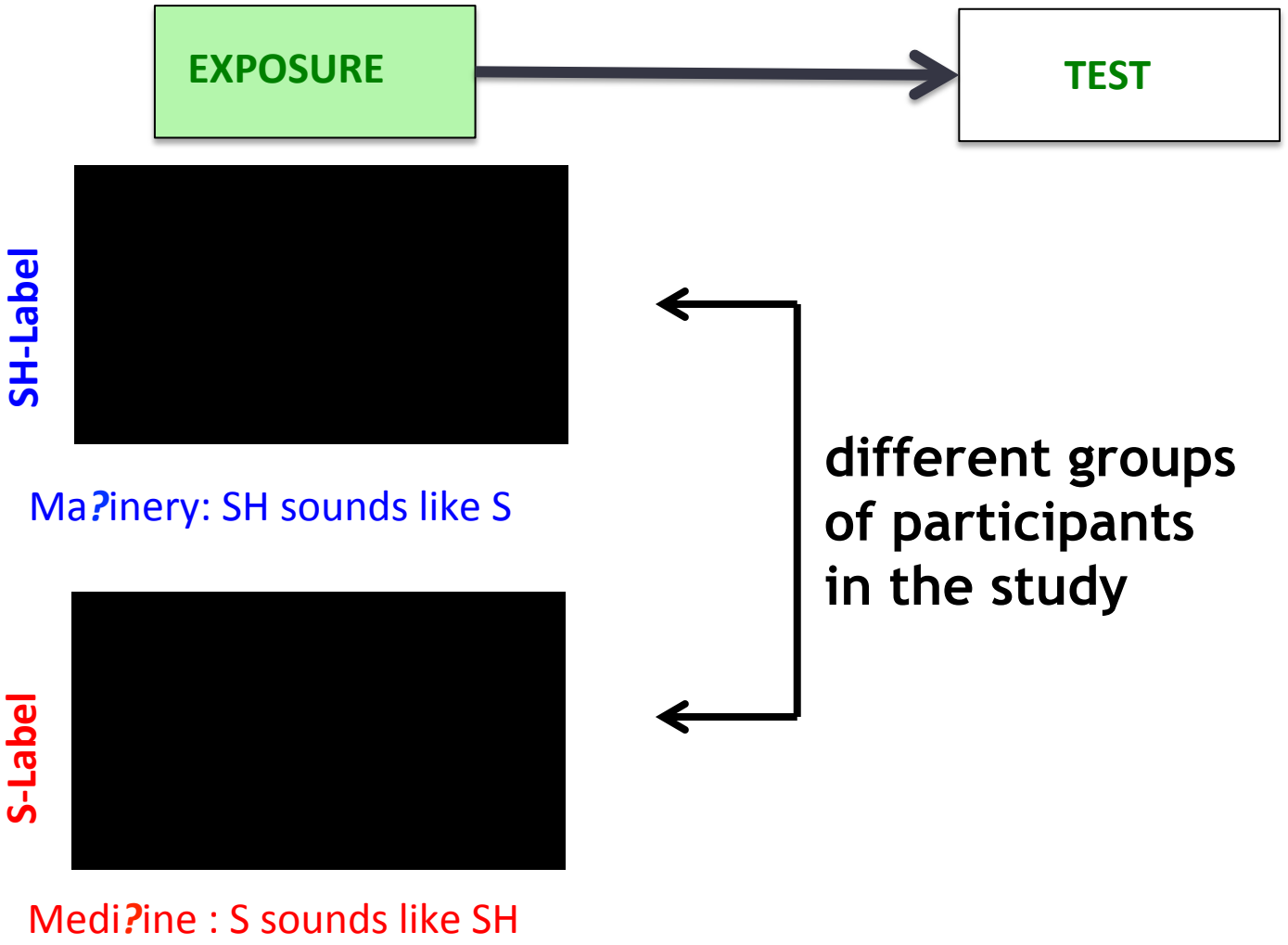


Exposure Phase

- Use the most ambiguous sound as exposure stimulus
- Manipulate the ***lexical context*** as a label for whether the sound is /sh/ or /s/
- medi?ine → ? more likely to be perceived as /s/ than /sh/
  - ***(/s/-label exposure)***
- ma?inery → ? more likely to be perceived as /sh/ than /s/
  - ***(/sh/-label exposure)***

Experiment Design

[Slides adapted from Liu & Jaeger, 2015-AMLaP]



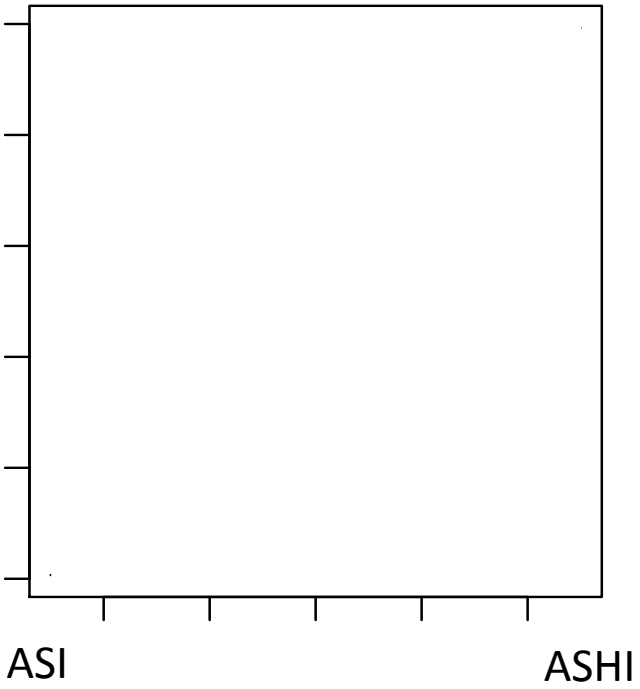
Experiment Design

SH-Label

Ma?inery: SH sounds like S

S-Label

Medi?ine : S sounds like SH



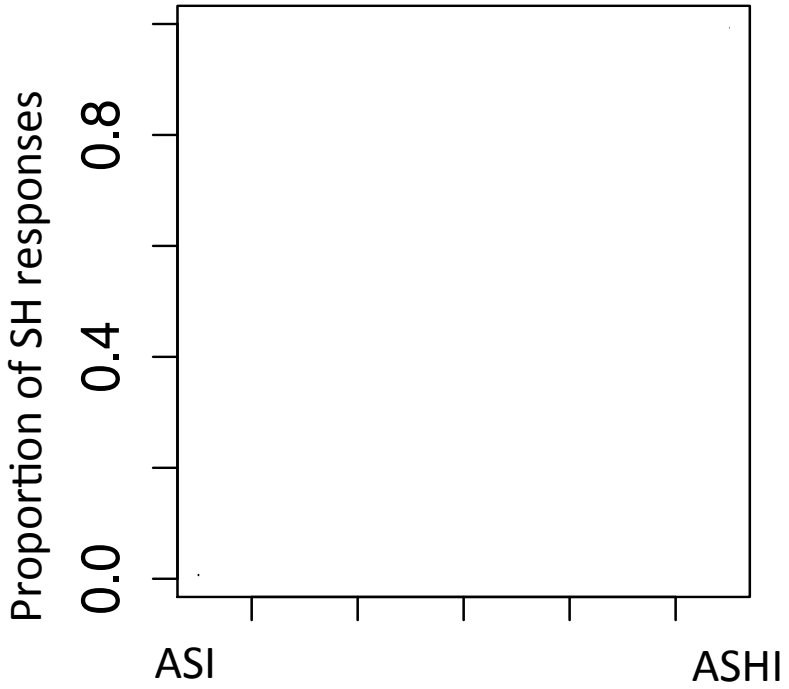
Experiment Design

SH-Label

Ma?inery: SH sounds like S

S-Label

Medi?ine : S sounds like SH



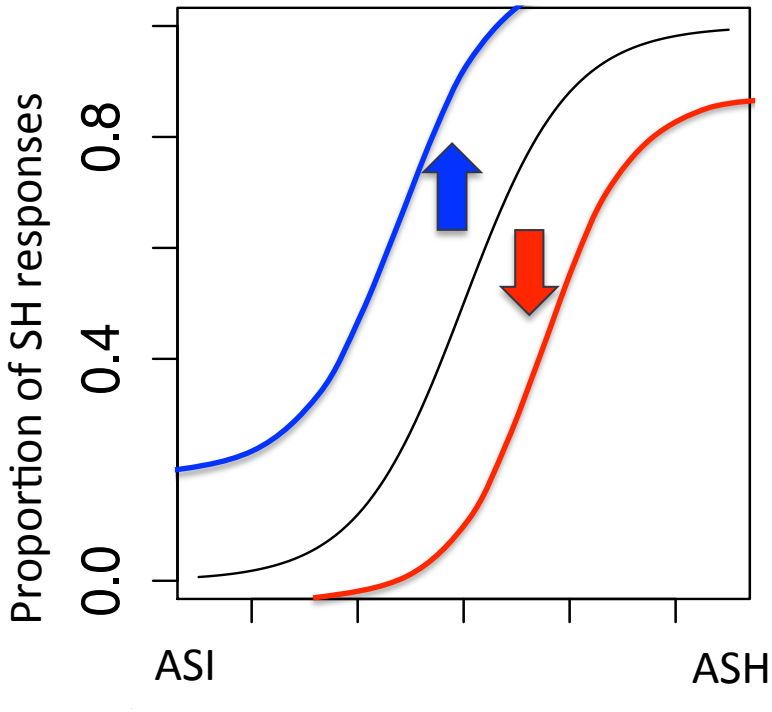
Experiment Design

SH-Label

Ma?inery: SH sounds like S

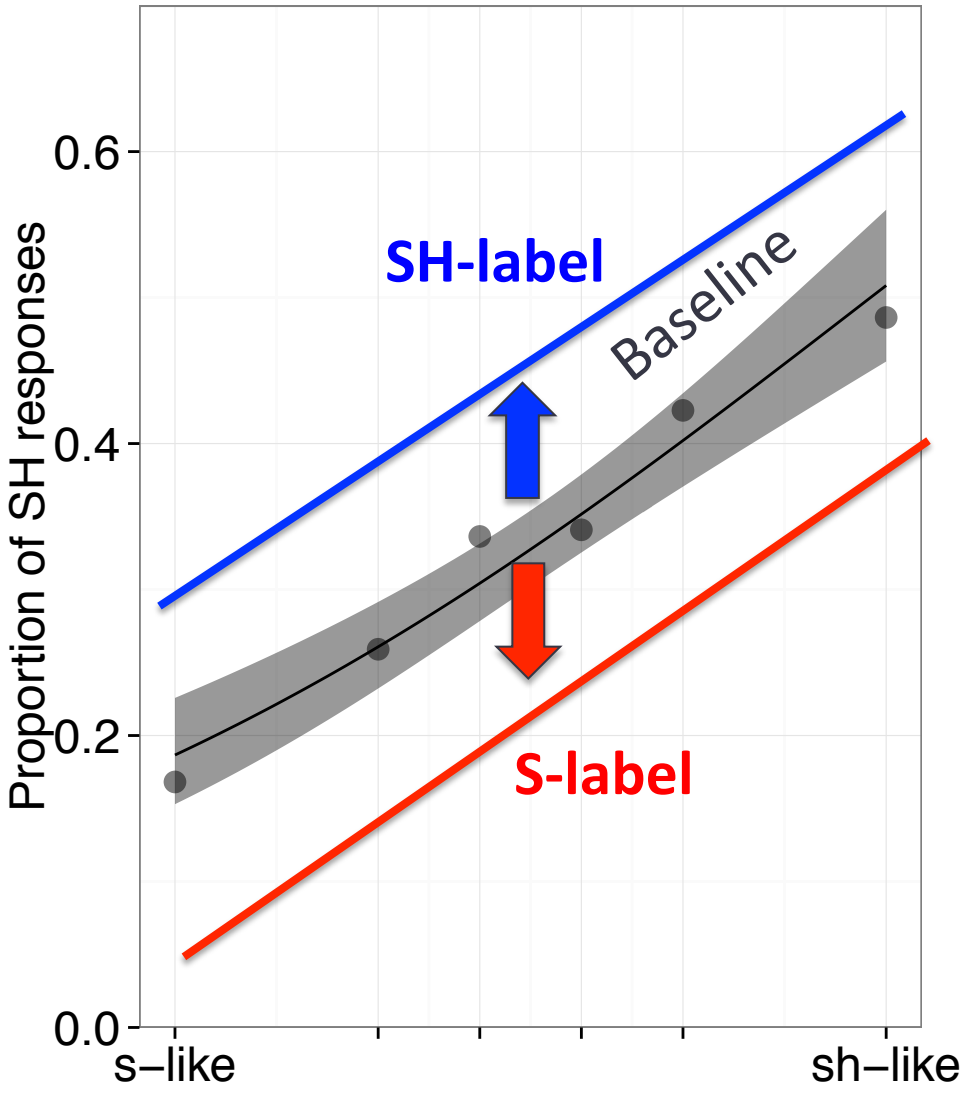
S-Label

Medi?ine : S sounds like SH

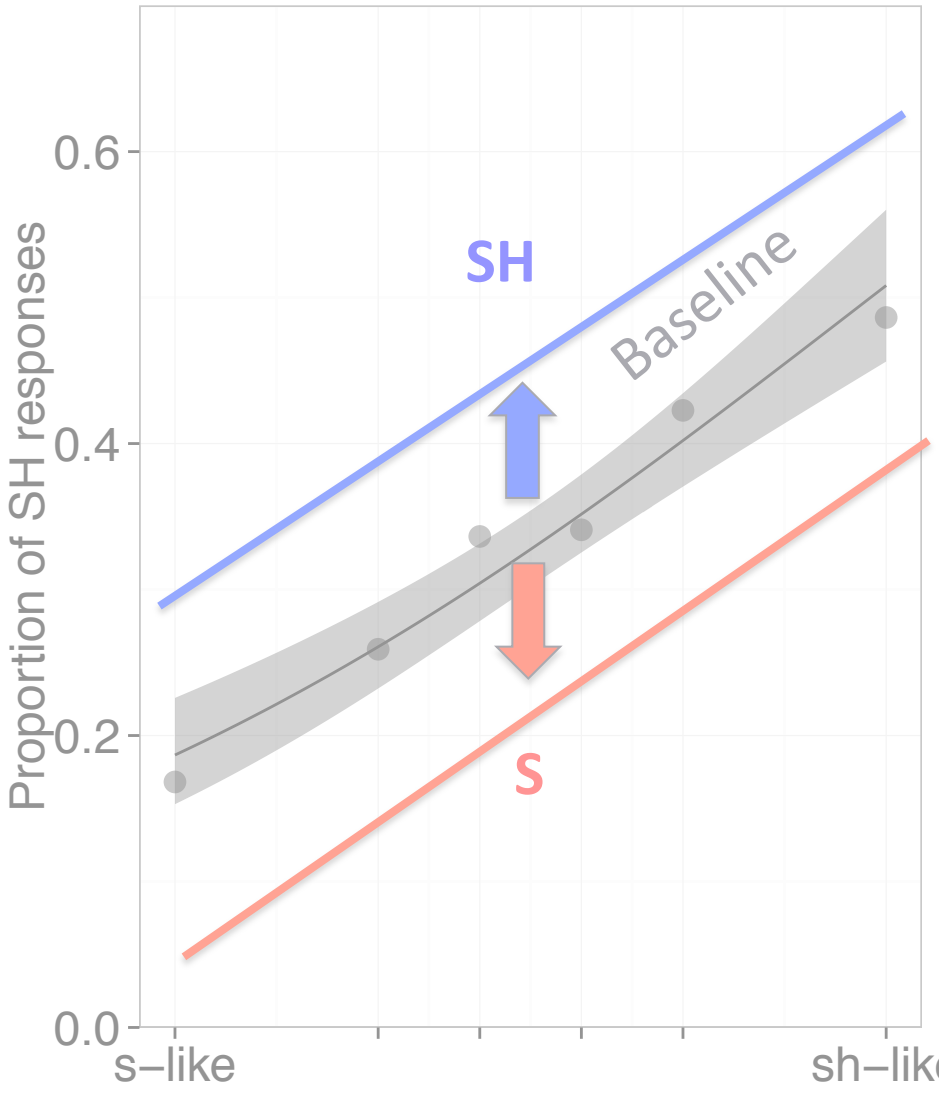


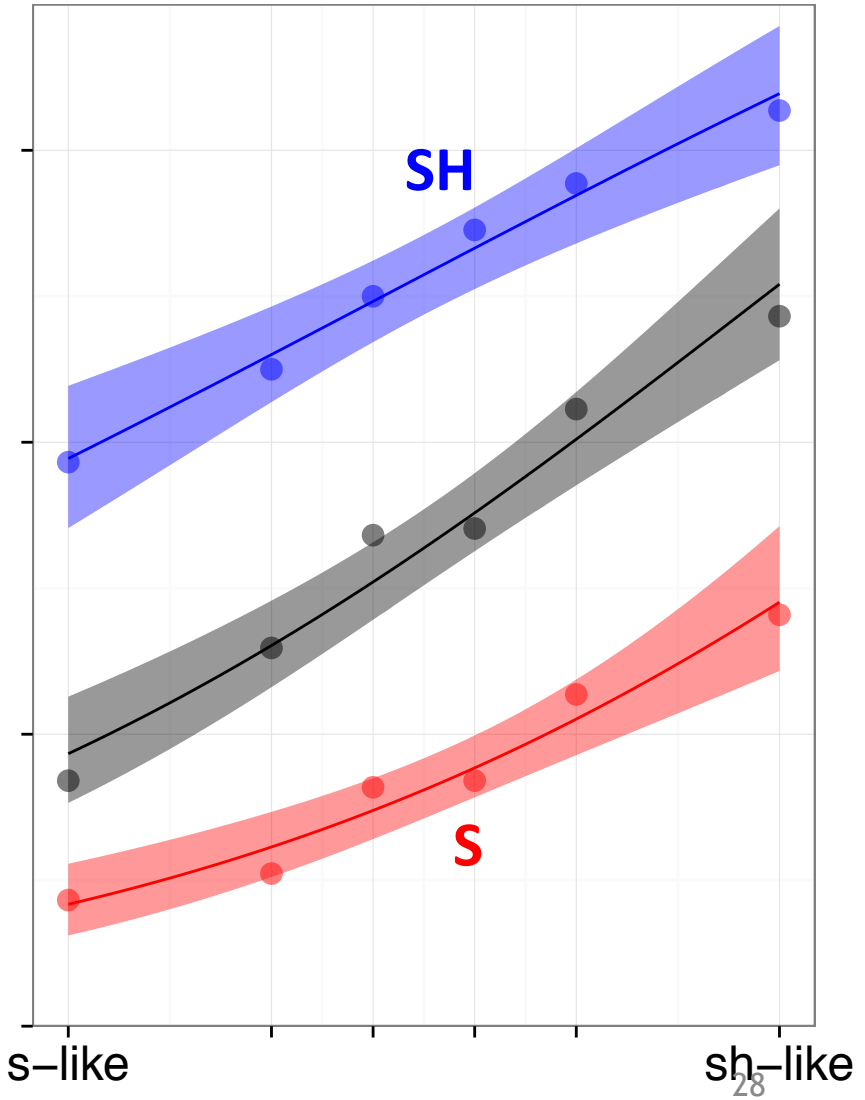
(e.g. Norris et al. 2003, Kraljic & Samuel, 2005)

Prediction


Proportion of SH responses
0.6
0.4
0.2
0.0
SH-label
Baseline
S-label
s–like
sh–like

Results (Liu et al., 2018)





Similar results for other speech sounds

- /s/-/sh/
- /d/-/t/ (Kraljic & Samuel, 2006)
- /b/-/p/
- /f/-/v/
- /f/-/s/
- …
- **Adaptation seems to be a general strategy listeners employ during speech perception**

To what extent does adaptation generalize?

Adaptation to one sound contrast (e.g., /b/-/p/) generalizes to other sounds that contrast along the same feature (e.g., /d/-/t/)

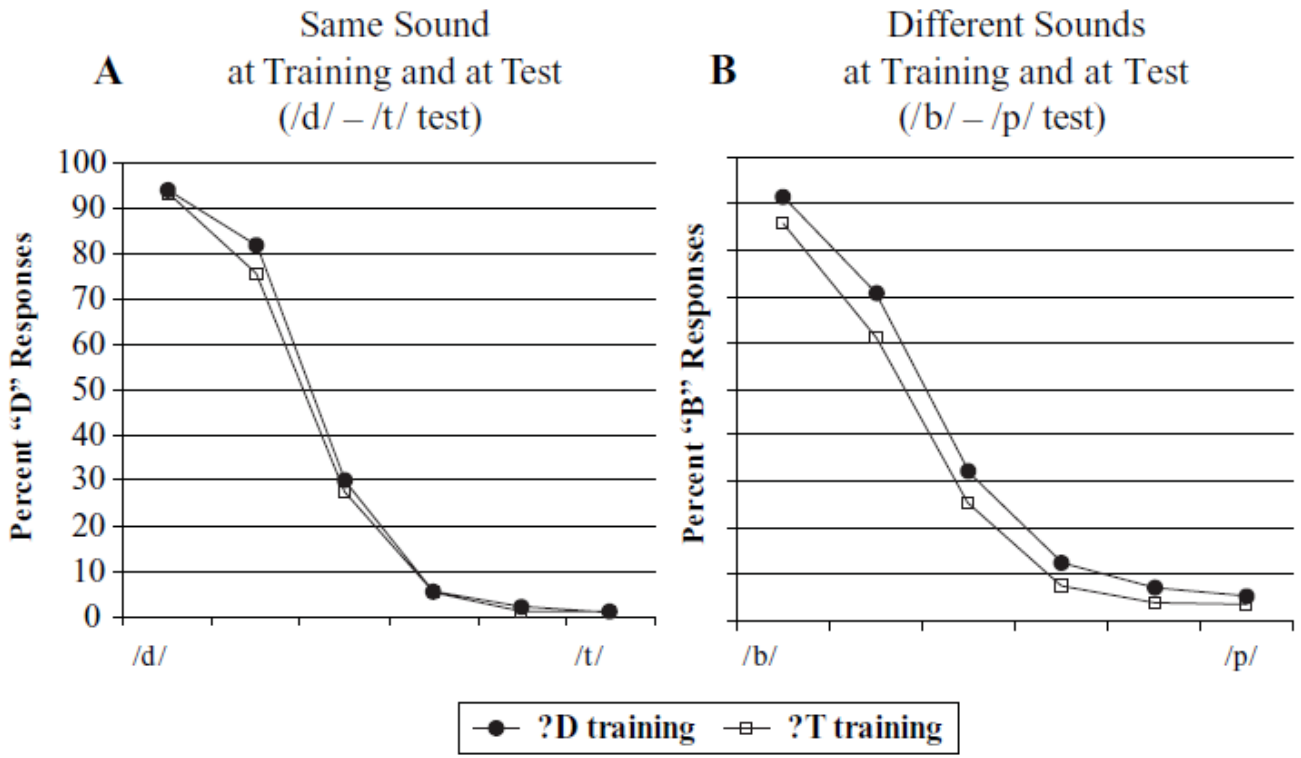


Figure 3. The training effect also generalizes to new phonemes: (A) Training and test on same consonant (/d/–/t/). (B) Training and test on different consonant (/b/–/p/)

Fig 3 from Kraljic and Samuel (2006)

Generalization also occurs over different speakers

(is this good or bad?)

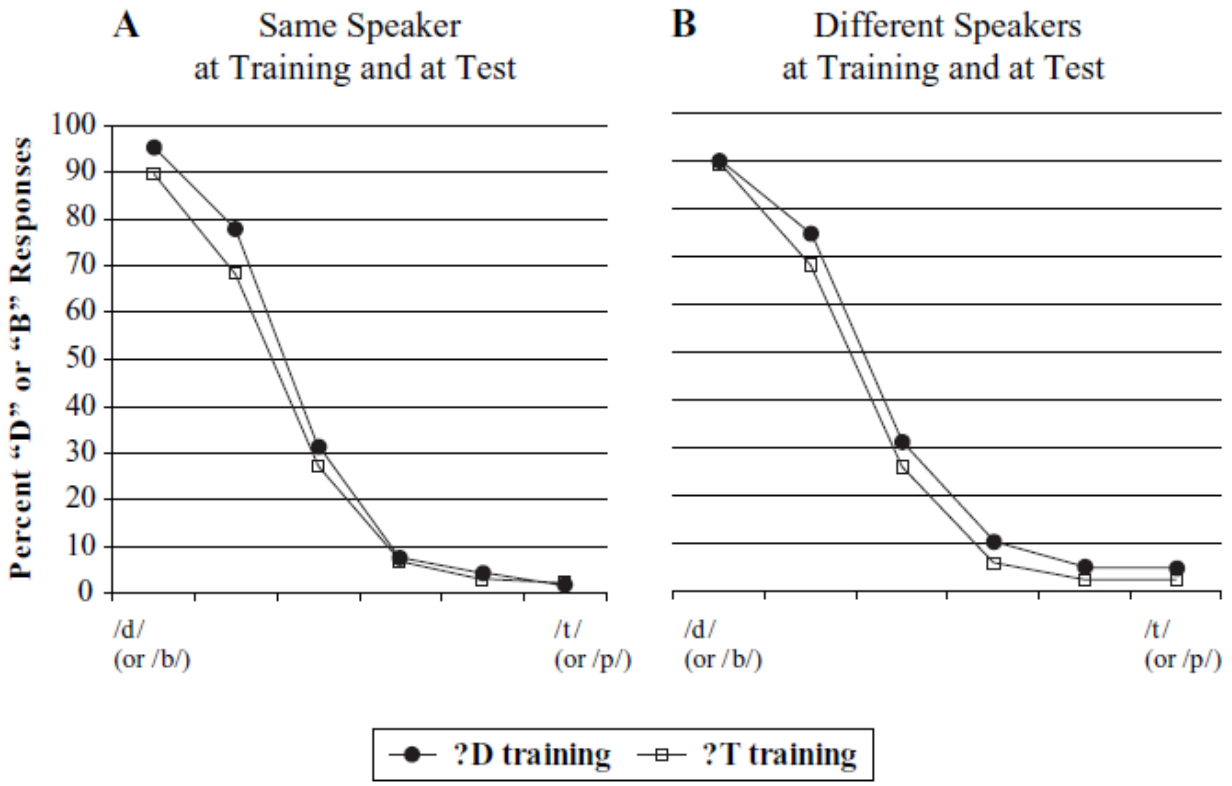


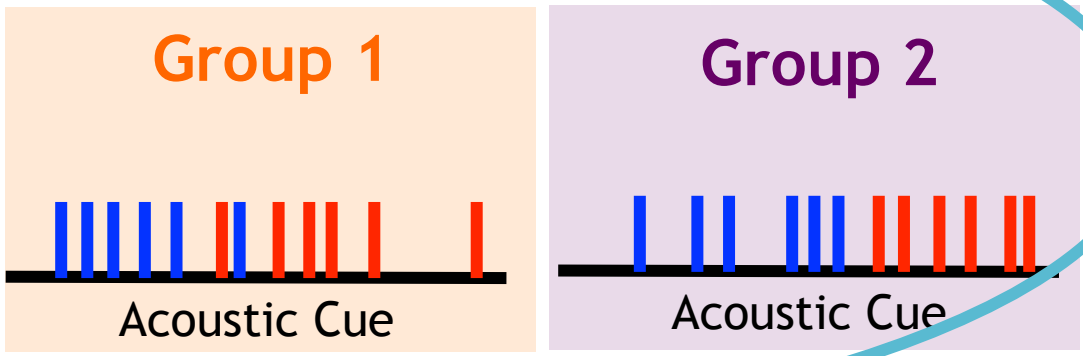
Figure 2. The training effect generalizes to a new speaker: (A) Training effect when participants are tested on the voice to which they were exposed in the lexical decision task. (B) Training effect when participants are tested on a voice different from the one to which they were exposed.

Fig 2 from Kraljic and Samuel (2006)

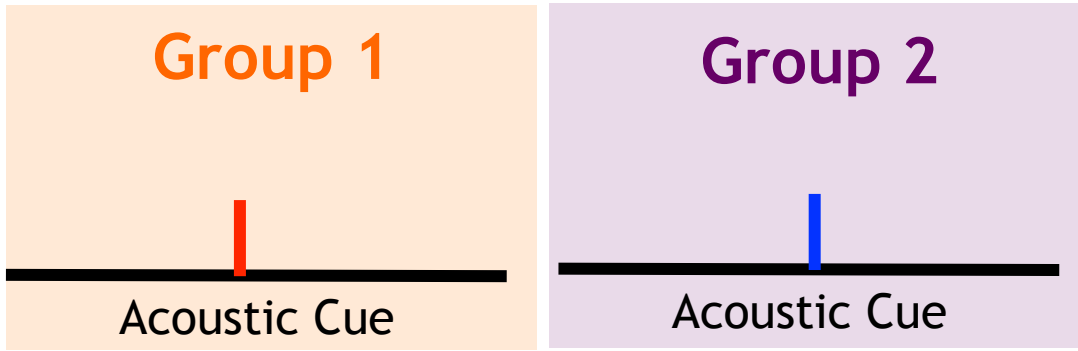
2 Methods of Exposure

/s/
/sh/

Expose participants to **full distribution** of a speaker's productions

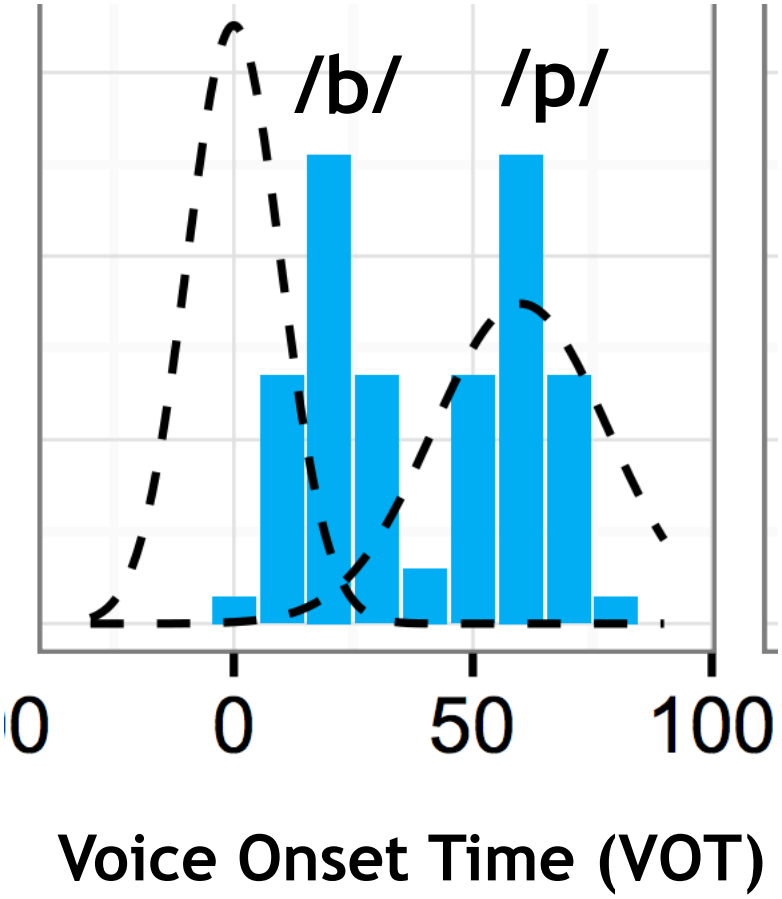


Expose to **targeted point(s)** along acoustic continuum



Full-Distribution Exposure

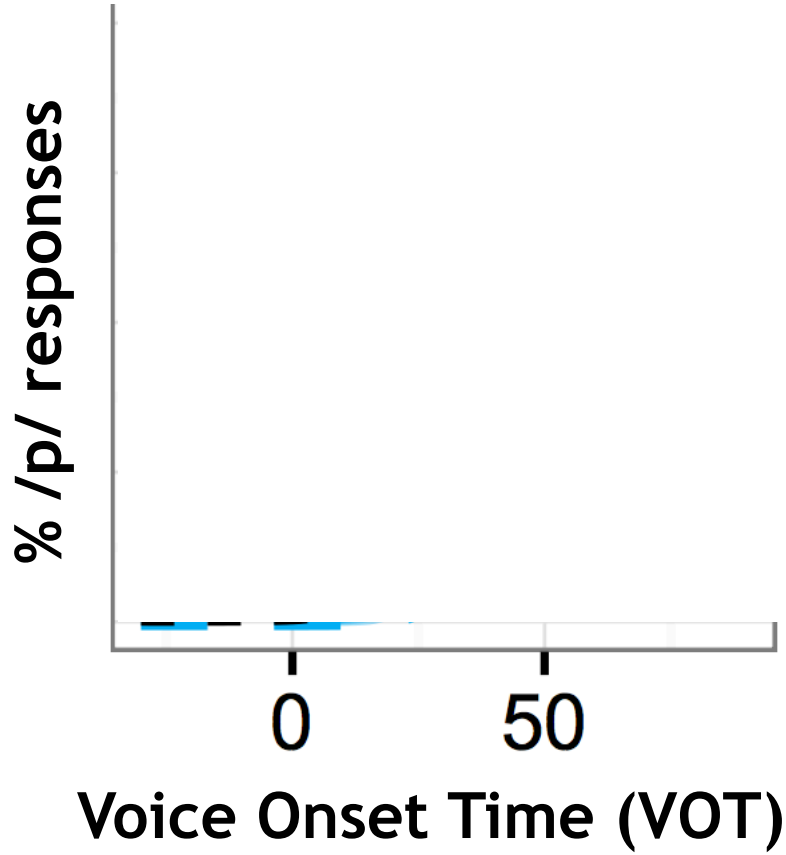
- Participants hear a distribution of productions rather than targeted towards the most *a priori* ambiguous points
- Words can be labeled *or* unlabeled
  - unlabeled data: unsupervised learning



from Kleinschmidt & Jaeger (2016)

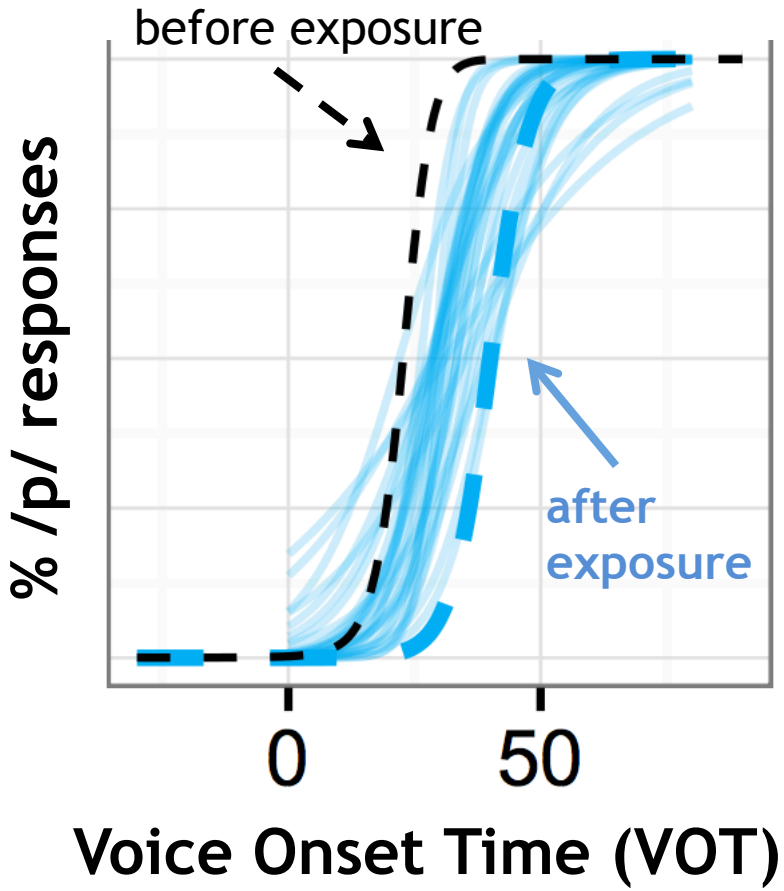
Experiment Design

?each (20ms VOT)
?it (70 ms VOT)
?eak (20ms VOT)
?each (25ms VOT)
?ill (75ms VOT)
?eak (70ms VOT)
…



Results

?each (20ms VOT)
?it (70 ms VOT)
?eak (20ms VOT)
?each (25ms VOT)
?ill (75ms VOT)
?eak (70ms VOT)
…



from Kleinschmidt & Jaeger (2016)

Adaptation to Single Sounds: Summary

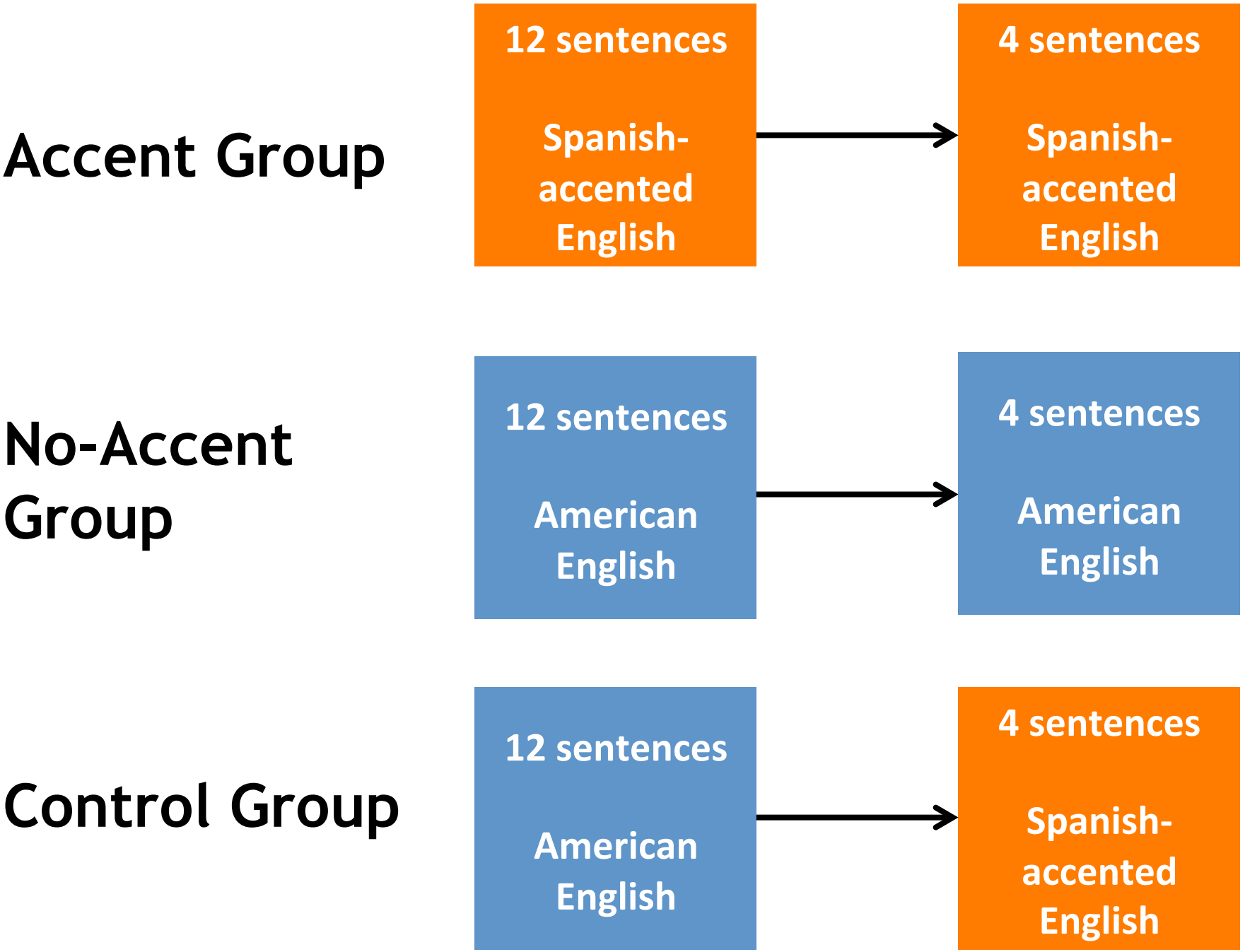
- Listeners are flexible: we don't use the same category boundary for every situation
- People infer categories of new speakers in both supervised and unsupervised contexts
- Adaptation generalizes to new sound contrasts along similar acoustic dimensions
- Adaptation studies give us insight into how we seem to effortlessly understand each other even though we all speak differently

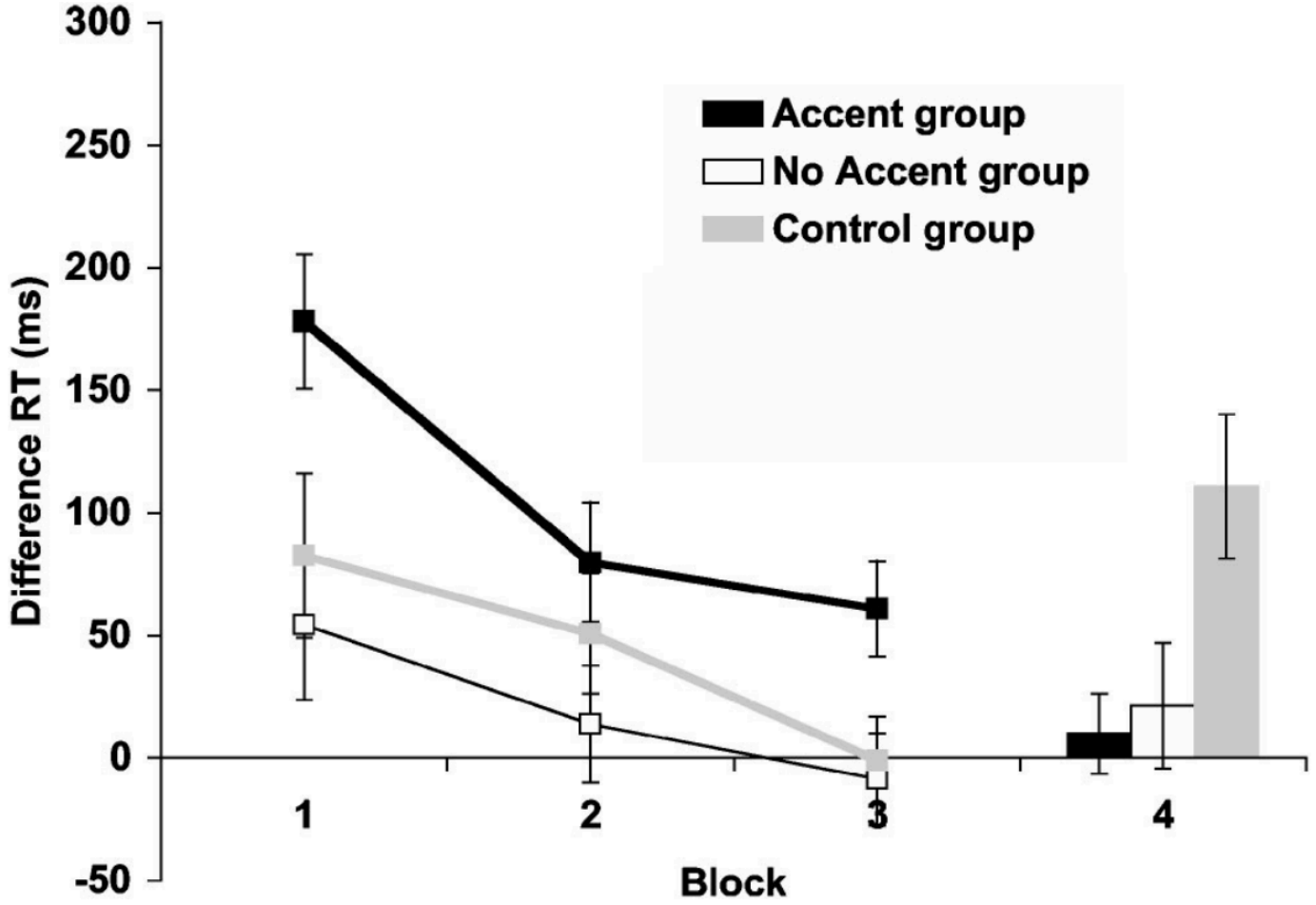
Adaptation to Accents & Dialects

- What happens when you need to adapt to many differently produced sounds at once?
- Experimentally harder to pinpoint what specifically listeners are adapting to
- Primary methodology: transcription tasks

Clarke & Garrett (2004)

- How well do listeners adapt to a new accent just by listening?
- Task:
  - listen to sentence
  - after sentence, visually presented word appears
  - judge whether the word matched the last word of the sentence
  - measure reaction time

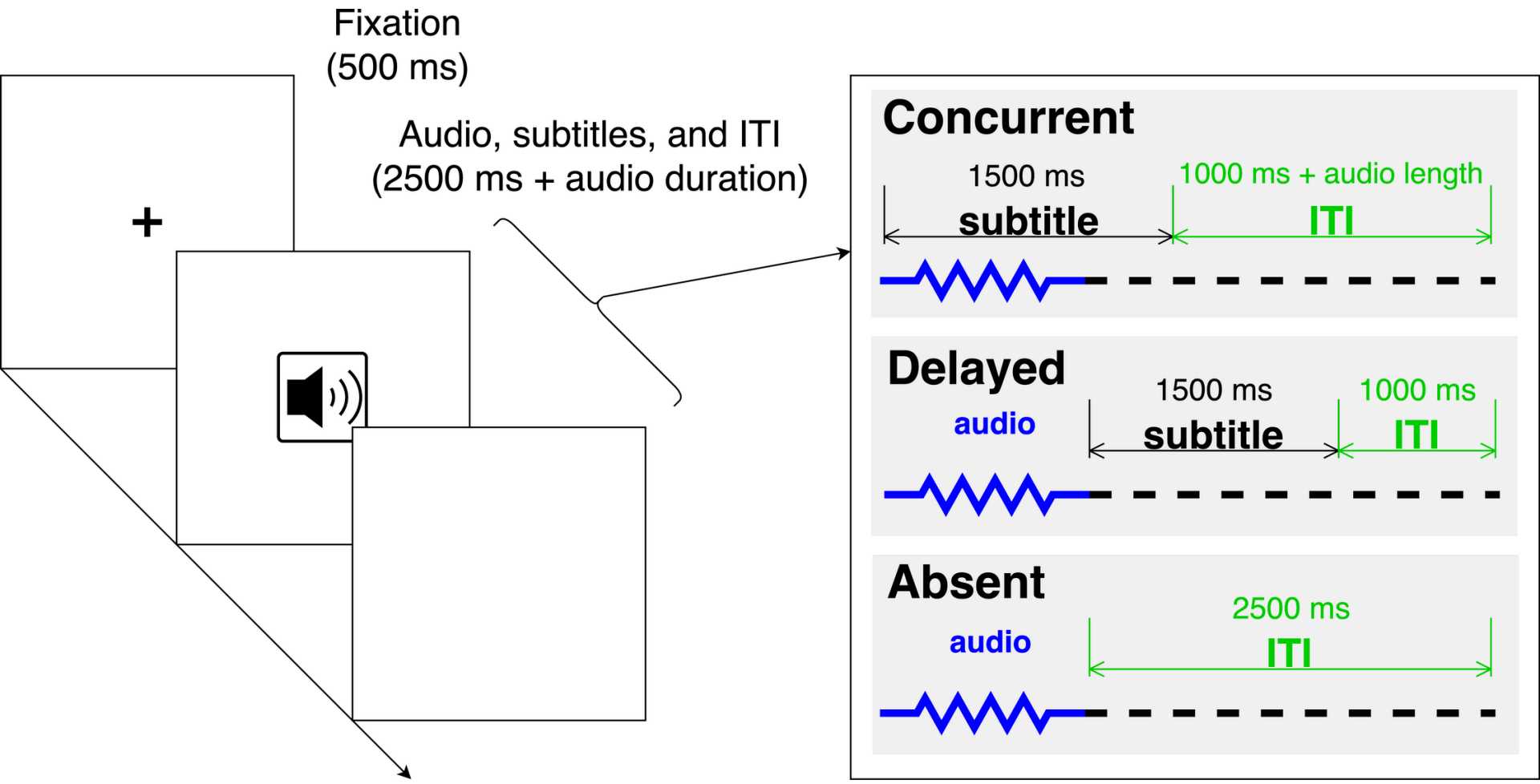

Accent Group
12 sentences
Spanish-accented English
4 sentences
Spanish-accented English
No-Accent Group
12 sentences
American English
4 sentences
American English
Control Group
12 sentences
American English
4 sentences
Spanish-accented English



- Exposure to Spanish-accented English → better performance than no exposure
- Performance on Spanish-accented talker identical to American English-accented talker after only 12 sentences!

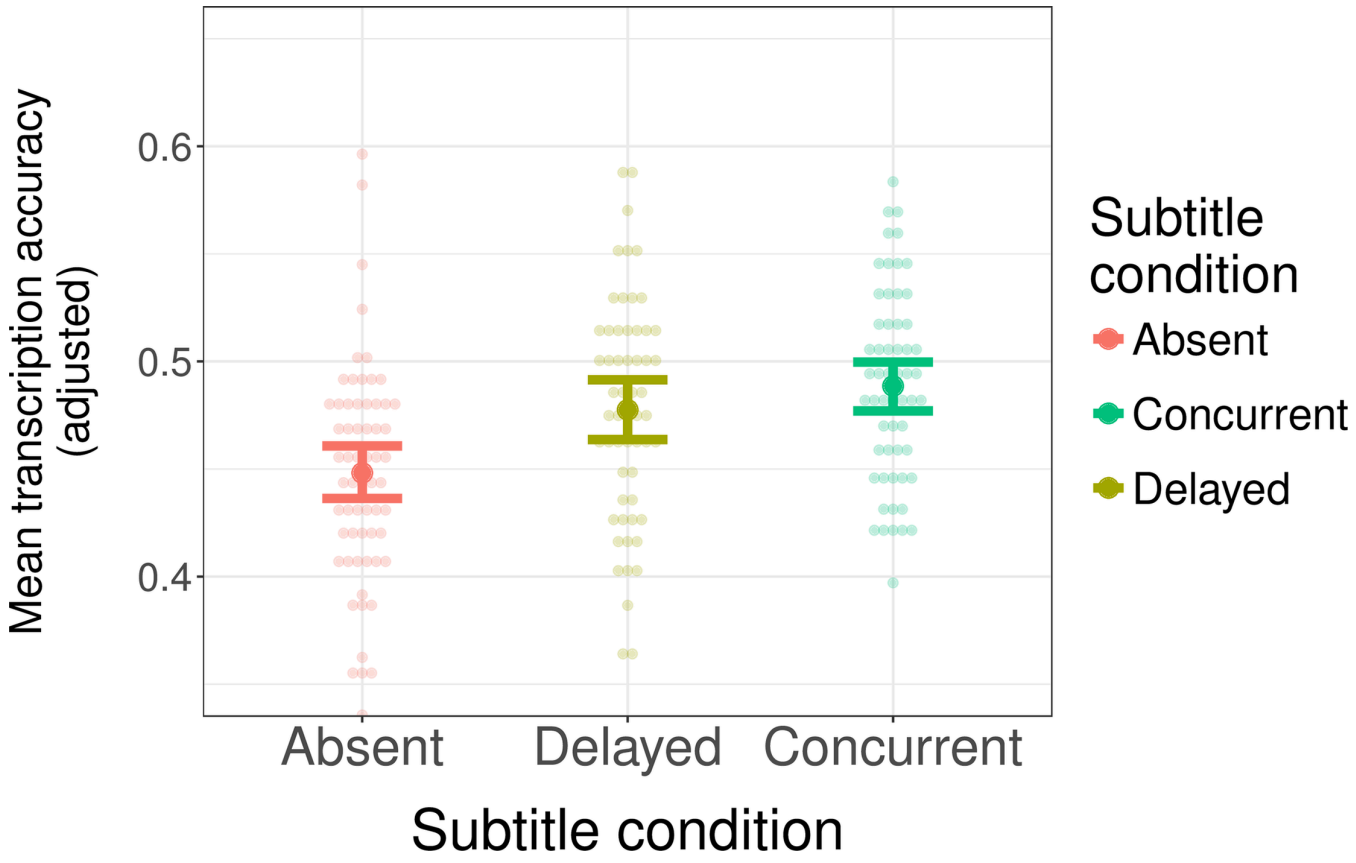
- People are good at adapting to a new accent just through raw exposure
  - equivalent performance (in RTs) to native accented speech within 3 minutes!
- But accuracy doesn't quite reach native-accented speech levels
- What factors might improve adaptation?

Lexical Information



Burchill et al. (2018)

Lexical Information

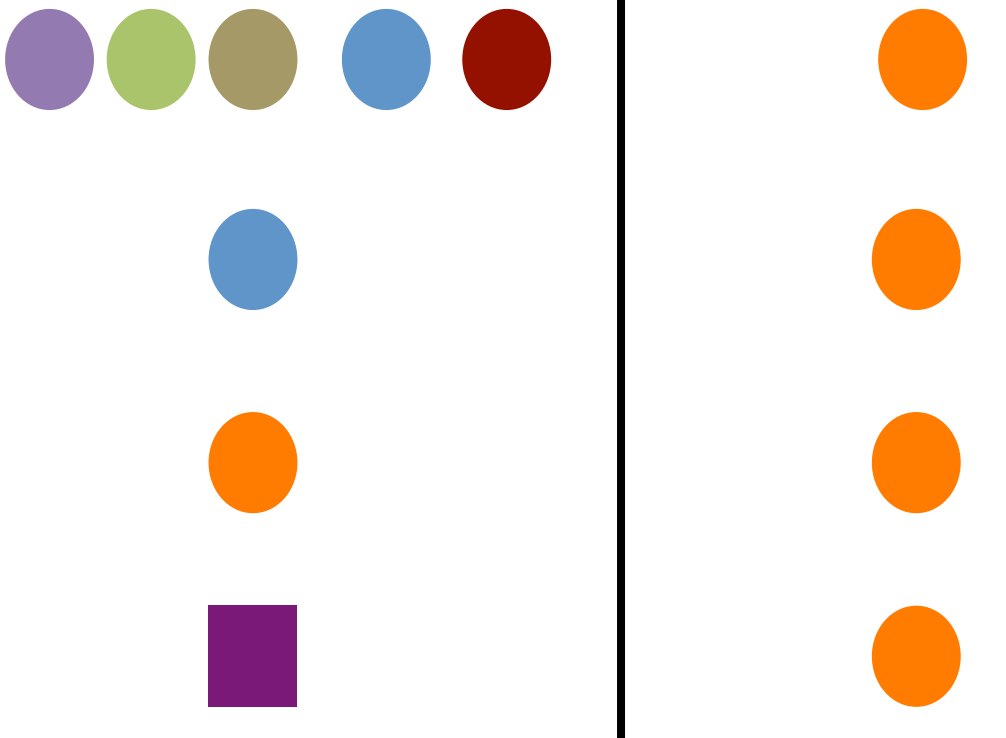


Burchill et al. (2018)

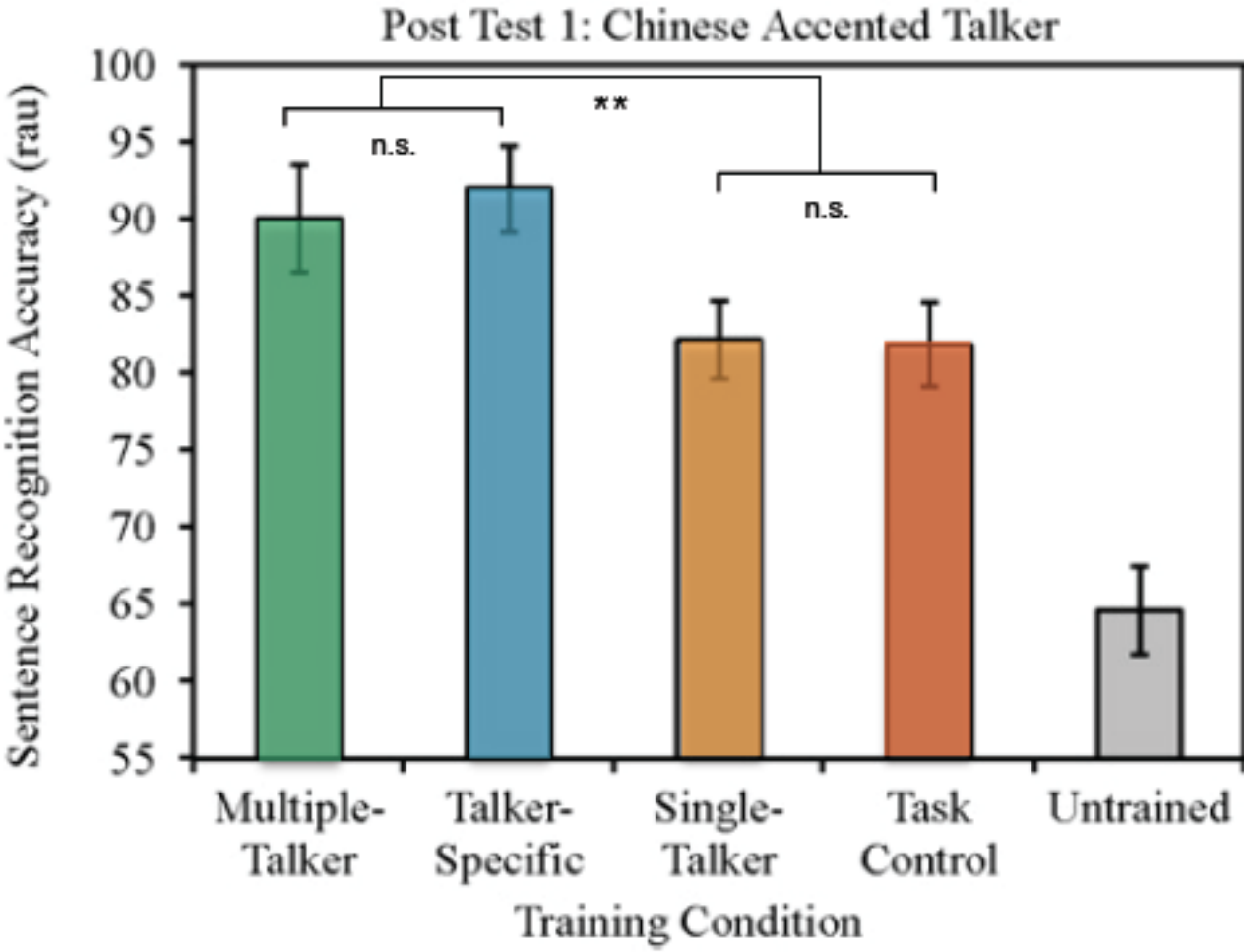
Exposure to other talkers with the same accent (Bradlow & Bent, 2008)

= Chinese-accented talker

= American English-accented talker



Exposure to other talkers with the same accent (Bradlow & Bent, 2008)



- Exposure to multiple different talkers of same accent just as beneficial for understanding a new talker as exposure to that new talker

- Single-talker exposure no different than control
  - suggests people need some variability in order to generalize

Preview to next week...

- How do we use our world knowledge and linguistic knowledge during adaptation?
- Is all of speech perception just continuous adaptation?
- How does the brain represent speech sounds? Can these representations change with short-term adaptation?