# The perceptual generalization of normalized cue distributions across speakers

Wei Lai (Vanderbilt University)
Aini Li (University of Pennsylvania)

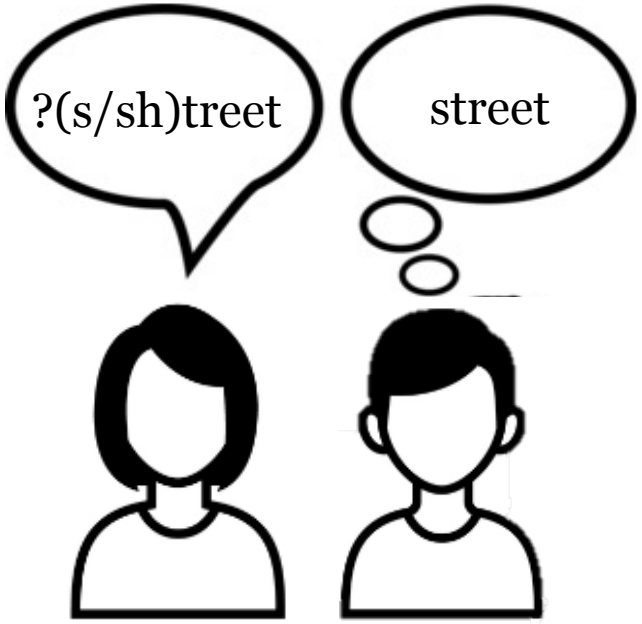
Jan 5-8, LSA 2023

# Perceptual learning and generalization

- Listeners make perceptual adjustments to adapt to talker-specific phonetic distributions. (Norris, McQueen, & Cutler, 2003)
- They also generalize the perceptual adjustments across different speakers. (Kraljic & Samuel, 2006; Reinisch & Holt, 2014; Xie et al., 2018).

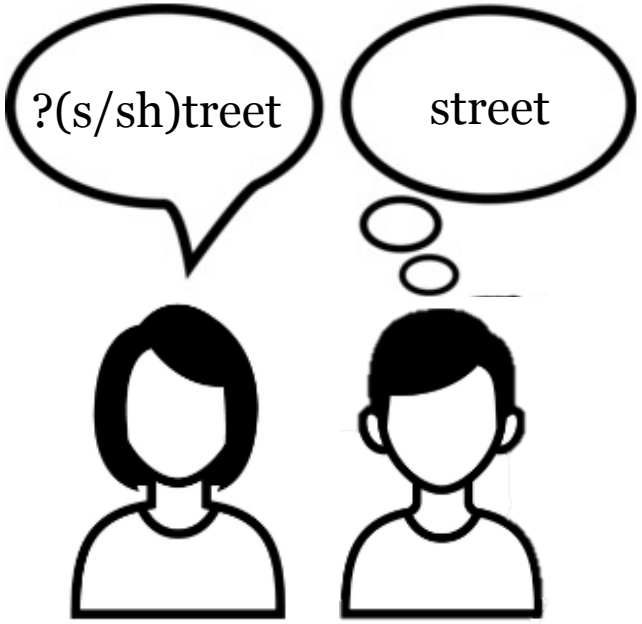
# Perceptual learning and generalization

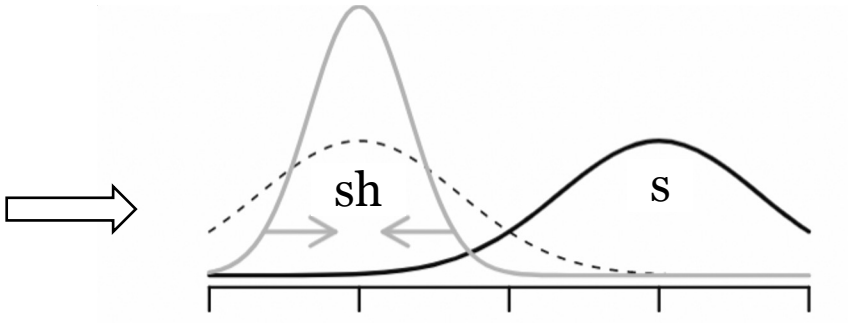
- Listeners make perceptual adjustments to adapt to talker-specific phonetic distributions. (Norris, McQueen, & Cutler, 2003)
- They also generalize the perceptual adjustments across different speakers. (Kraljic & Samuel, 2006; Reinisch & Holt, 2014; Xie et al., 2018).



# Perceptual learning and generalization

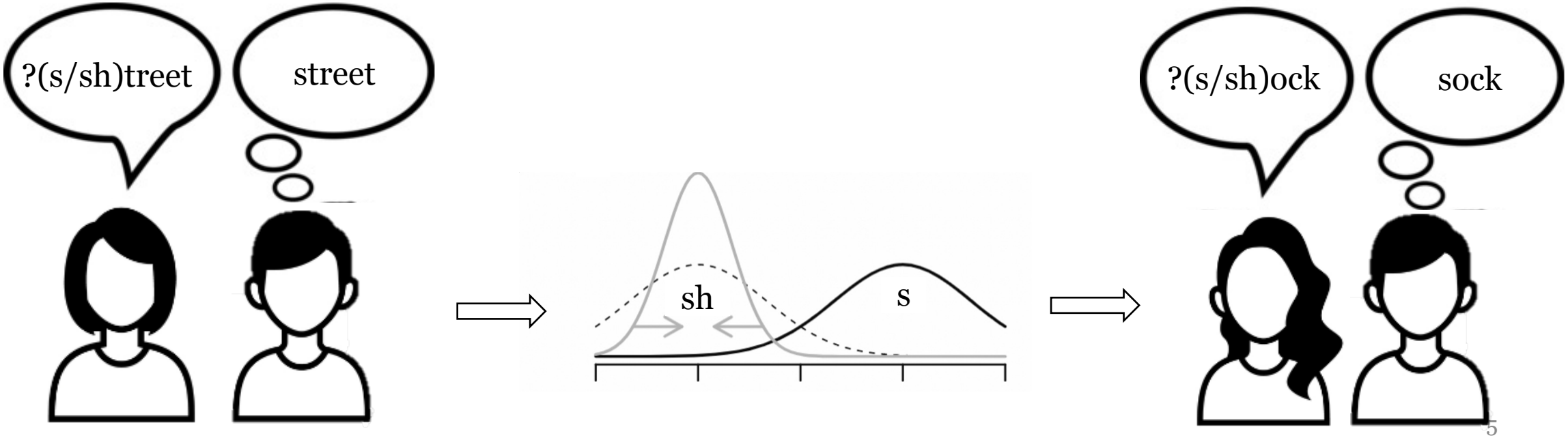
- Listeners make perceptual adjustments to adapt to talker-specific phonetic distributions. (Norris, McQueen, & Cutler, 2003)
- They also generalize the perceptual adjustments across different speakers. (Kraljic & Samuel, 2006; Reinisch & Holt, 2014; Xie et al., 2018).



# Perceptual learning and generalization

- Listeners make perceptual adjustments to adapt to talker-specific phonetic distributions. (Norris, McQueen, & Cutler, 2003)
- They also generalize the perceptual adjustments across different speakers. (Kraljic & Samuel, 2006; Reinisch & Holt, 2014; Xie et al., 2018).

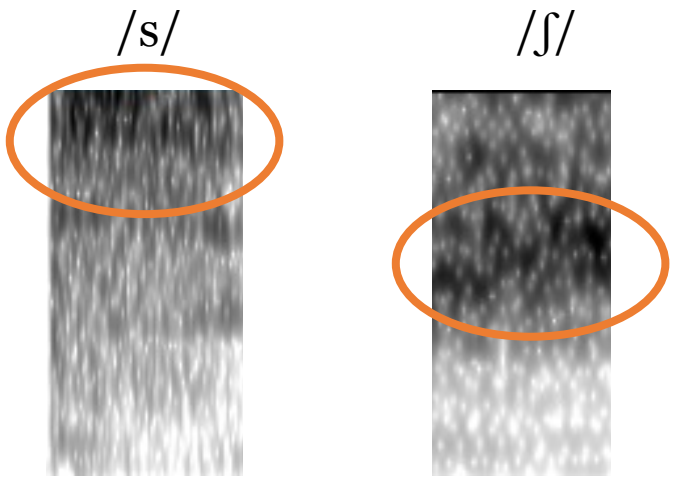


# Speech normalization

- Phonemic categorization is not only informed by raw phonetic distributions, but also *relative contextual cues* from the talker's speech.
  (e.g., Johnson, 1990, 2018; Port, 1979; Summerfield, 1975)

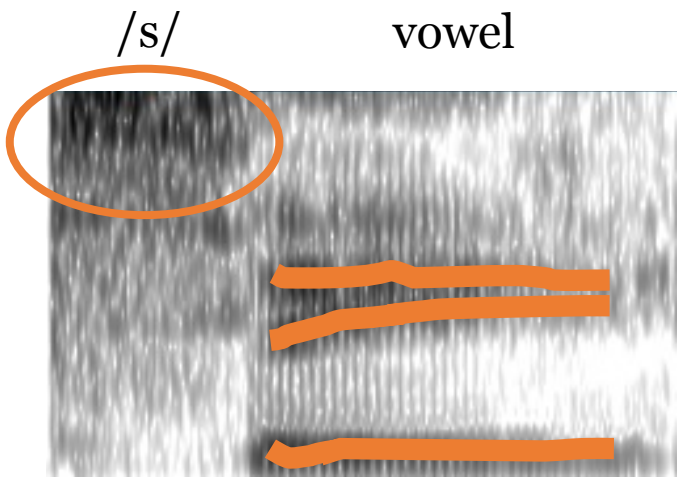
# Speech normalization of spectral cues

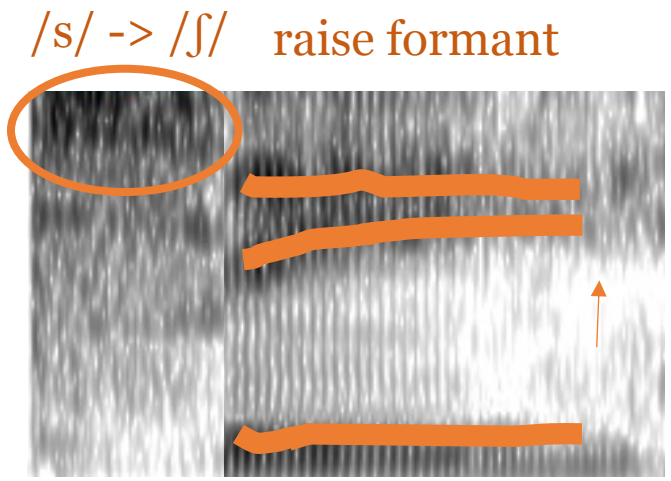
- The categorization of /s-ʃ/ varies with contextual vowel formants (Johnson, 1990, 2018)



# Speech normalization of spectral cues

- The categorization of /s-ʃ/ varies with contextual vowel formants (Johnson, 1990, 2018)



# Speech normalization of temporal cues

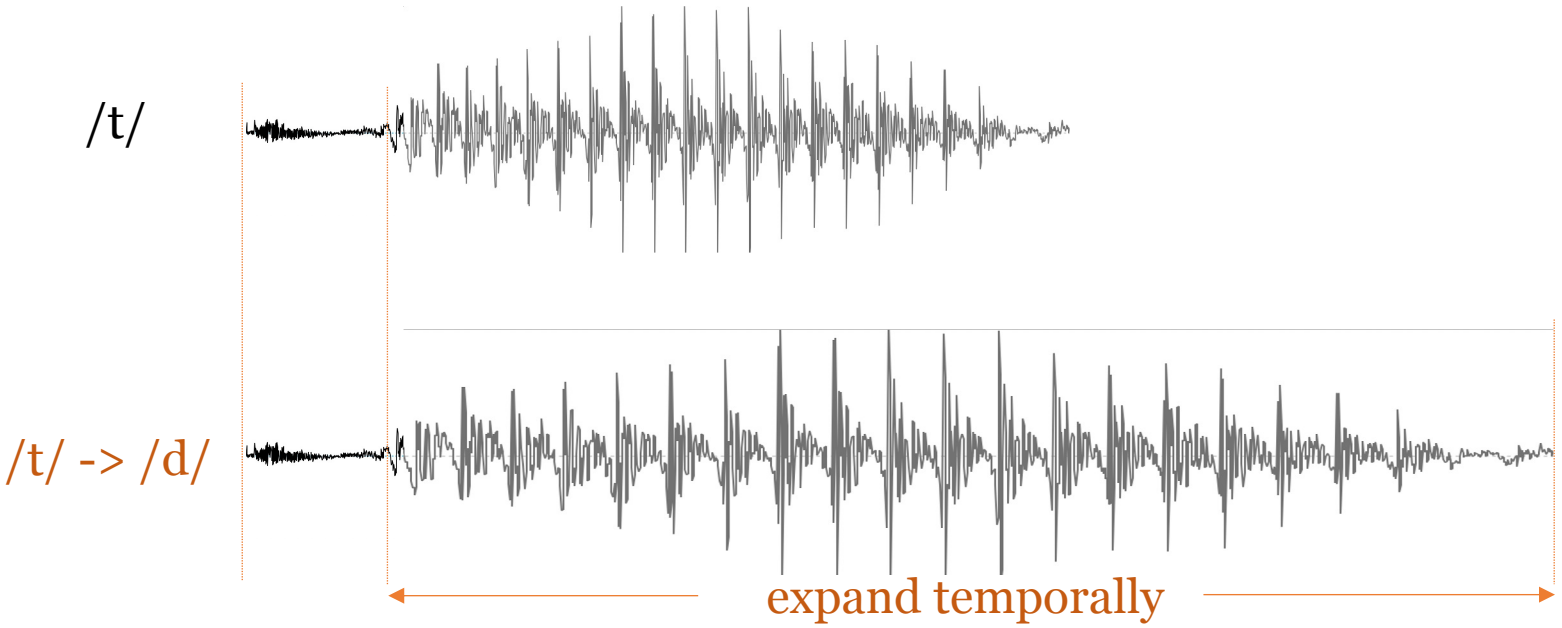
- The categorization of /t-d/ varies with contextual vowel duration
  (Summerfield, 1975; Port, 1979)

/t/ 

/d/ 

# Speech normalization of temporal cues

- The categorization of /t-d/ varies with contextual vowel duration
  (Summerfield, 1975; Port, 1979)



# Research Question

- In perceptual learning, do listeners learn and generalize raw phonetic cues or normalized cue distributions within a speaker's acoustic space?
  - Raw-distribution hypothesis
  - Normalized-distribution hypothesis

# Research Question

- In perceptual learning, do listeners learn and generalize raw phonetic cues or normalized cue distributions within a speaker's acoustic space?
  - Raw-distribution hypothesis
  - Normalized-distribution hypothesis

- The current study:
  - Experiment 1: spectral cues /s-ʃ/
  - Experiment 2: temporal cues /t-d/

# Experiment 1: /s-ʃ/

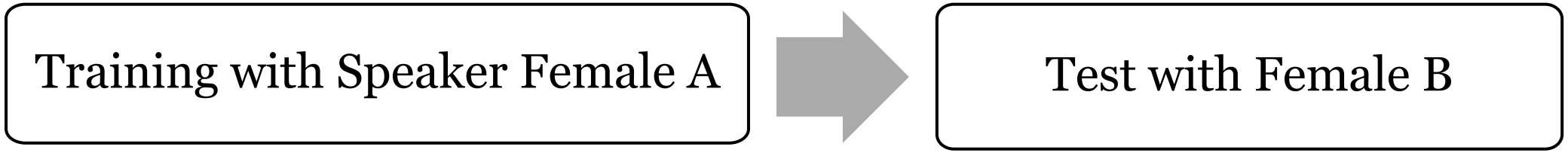
Question:

- Would changing *contextual vowel formants* of a training speaker affect listeners' categorization of */s-ʃ/* in a test speaker's speech?

Subject:

- 45 monolingual English speakers (20 men and 25 women) recruited through Prolific to participate online.
- Experiment implemented through PennController Ibex.

# Experiment 1: Method

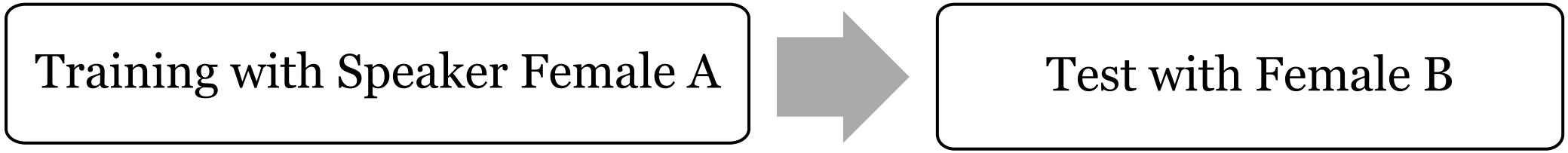


# Experiment 1: Method

Training with Speaker Female A → Test with Female B

- 51 trials of spoken word identification

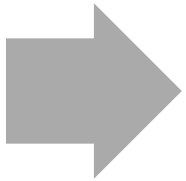
# Experiment 1: Method




- 51 trials of spoken word identification
- - 17 words containing /s/
- -17 words containing /ʃ/
- -17 fillers with no /s ʃ/

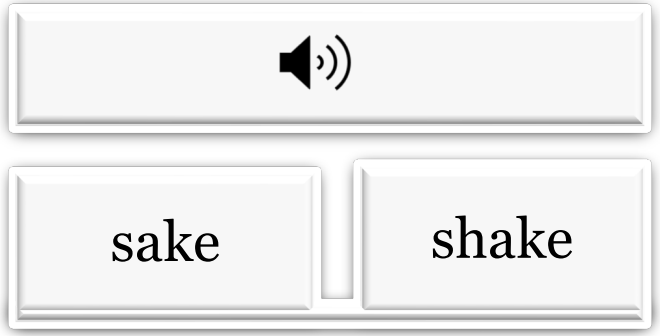
# Experiment 1: Method

Training with Speaker Female A



Test with Female B

- 51 trials of  word identification



# Experiment 1: Method

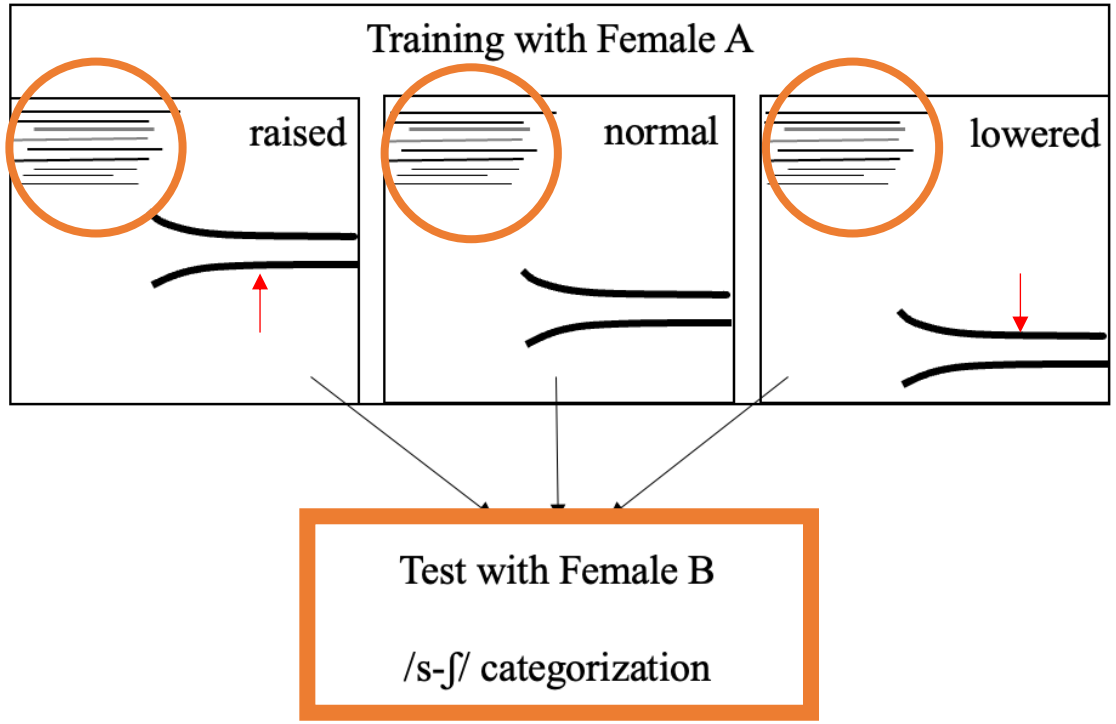
Training with Speaker Female A → Test with Female B

- 51 trials of  word identification
- - 35 /s ʃ/ minimal pairs
  - 5 steps x 7 words
  - *same, sign, seat, shelf, shake, shell, shy*
- - 16 filler trials with no /s ʃ/

# Experiment 1: Method

Participants assigned to 3 experiment conditions (N=15 on each condition):

- identical test phase
- identical /s ʃ/ in the training stimuli
- *different* contextual vowel formants of the training stimuli:
  - ➢Normal: unaltered
  - ➢Raised: scale formants by 1.2
  - ➢Lowered: scale formants by 0.8



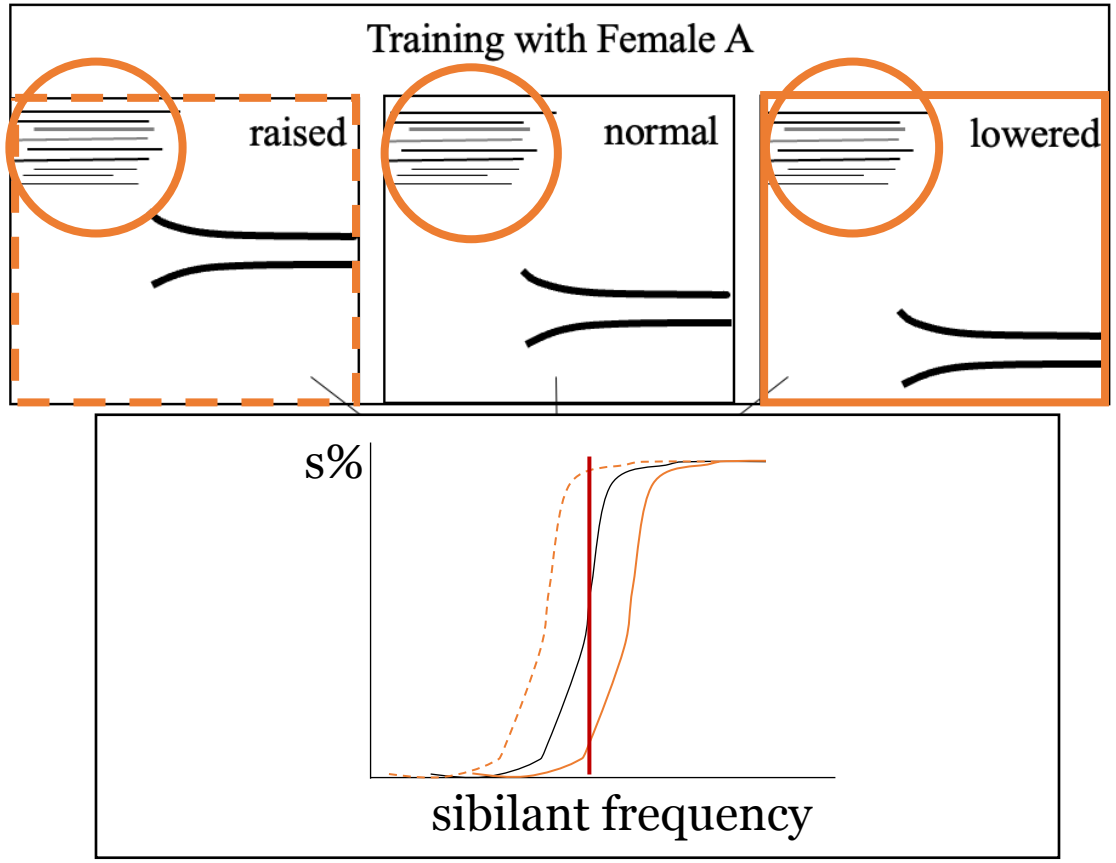
# Experiment 1: Hypotheses

Raw-distribution hypothesis:

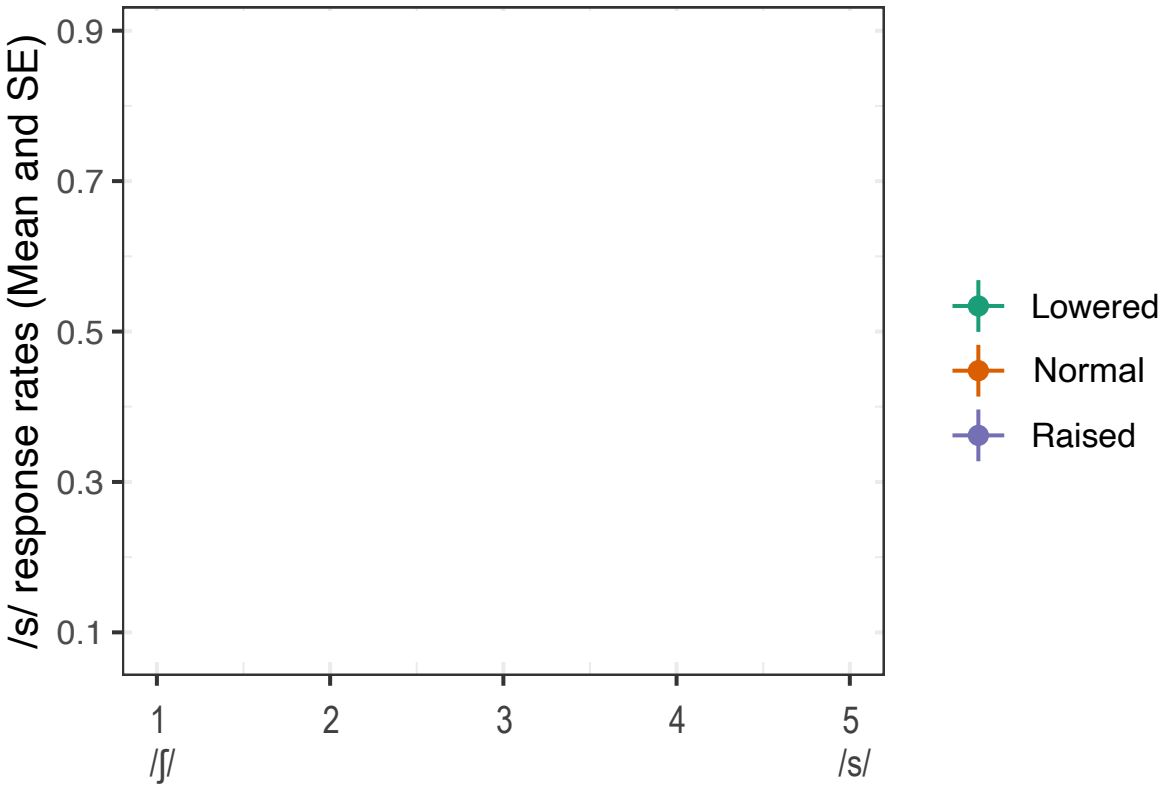
- Predicts that participants across conditions do not differ

Normalized-distribution hypothesis :

- The proportion of /s/: raised > normal > lowered



# Experiment 1: Results




- /s/ response rate: raised > normal > lowered
- Normalized distribution hypothesis ☺

- *response ~ step*condition+ (step|subject)+ (step|word)*
  - Step: β = 1.82, p < 0.001
  - Condition (raised-lowered): β = 1.54, p = 0.02

# Intermediate summary

- In the perceptual generalization of sibilants across speakers, changing contextual spectral cues of a training speaker would affect listeners' sibilant categorization of a test speaker
- The perceptual learning of *spectral* cues involves some degree of knowledge and computation about speaker-normalized distributions.
- Will the pattern hold for *temporal* cues?

# Experiment 2: /t-d/

Question:

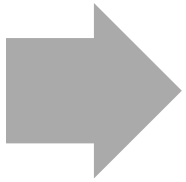
- Would changing contextual temporal cues of a training speaker affect listeners' categorization of /t-d/ in a test speaker's speech.

Subject:

- 45 English monolinguals (23 men and 22 women) recruited through Prolific to participate in the experiment online.
- Experiment implemented through PennController Ibex.

# Experiment 2: Method

Training with Speaker Female A



Test with Female B

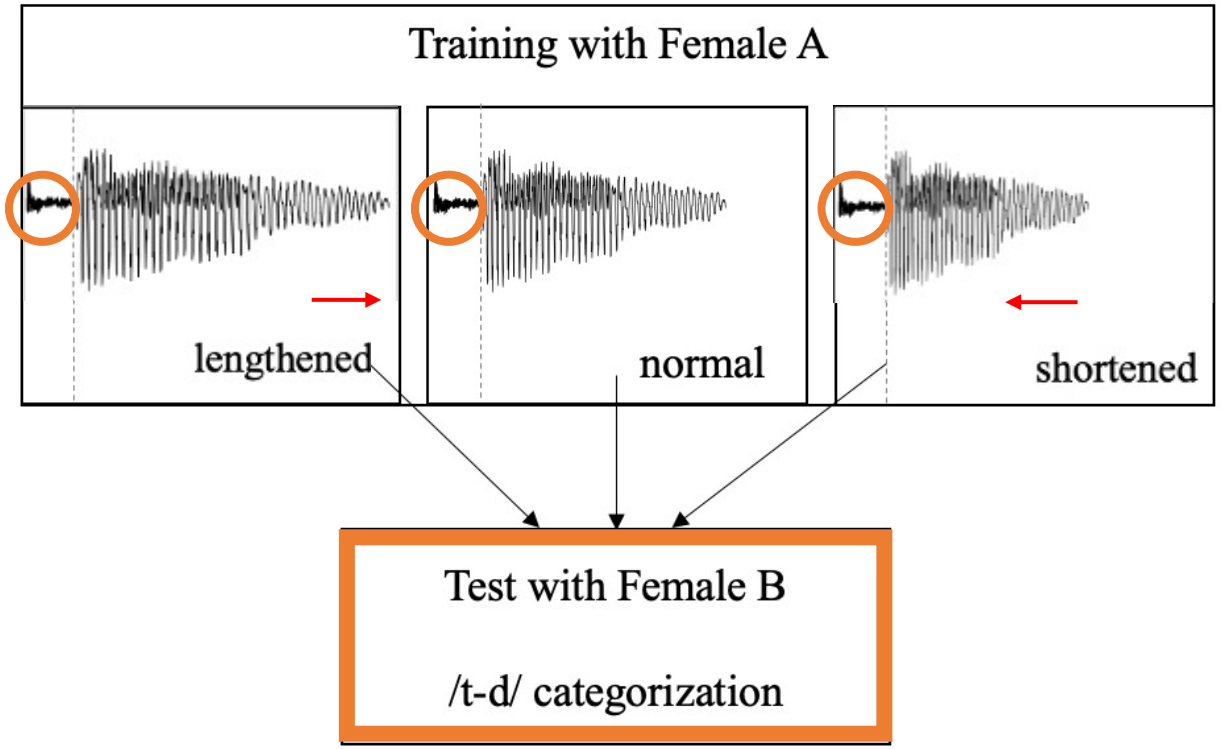
- 51 trials of spoken word identification
- - 17 words containing /t/
- -17 words containing /d/
- -17 fillers with no /t d/

- 51 trials of  word identification
- - 35 /t d/ minimal pairs
  - 5 steps x 7 words
  - *tear, tie, town, touch, time, tip, toes*
- - 16 filler trials with no /t d/

# Experiment 2: Method

Participants assigned to 3 experiment conditions (N=15 on each condition):

- identical test phase
- identical /t d/ in the training stimuli
- *different* contextual speech rates of the training stimuli
  - ➢ Normal: unaltered
  - ➢ Lengthened: temporally expanded by 1.7
  - ➢ Shortened: temporally compressed by 0.7



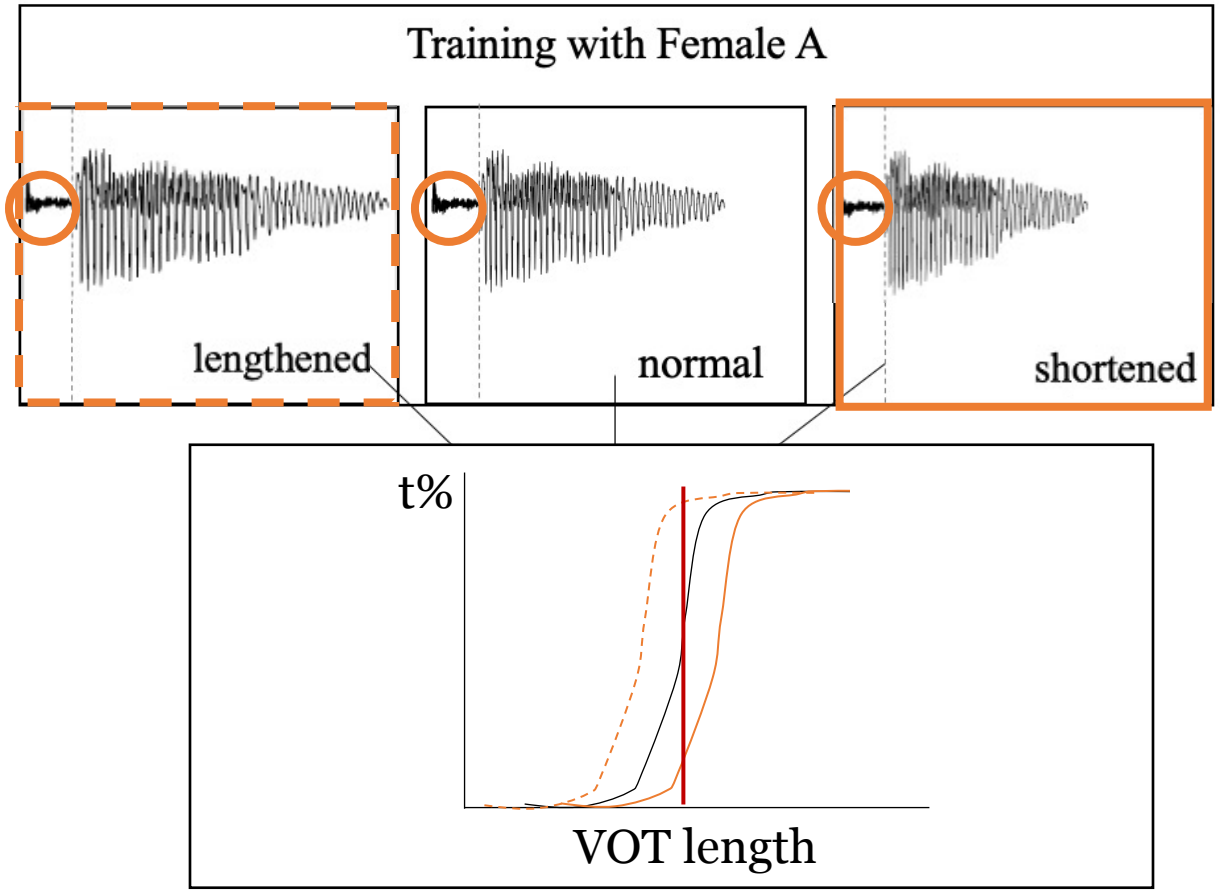
# Experiment 2: Hypotheses

Raw-distribution hypothesis:

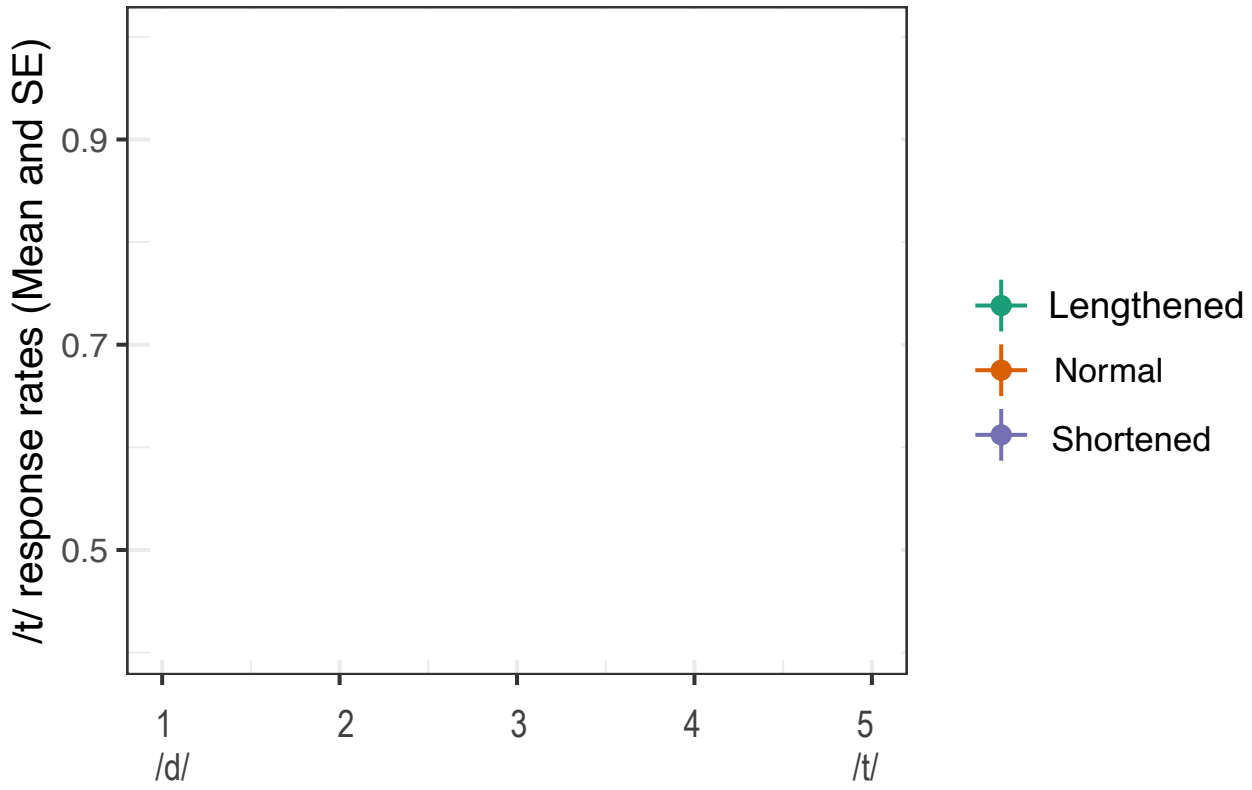
- Predicts that participants across conditions do not differ

Normalized-distribution hypothesis :

- The proportion of /t/: lengthened > normal > shortened



# Experiment 2: Results




- /t/ response rate: lengthened > normal > shortened
- Normalized distribution hypothesis ☺

- *response ~ step*condition+ (step|subject)+ (step|word)*
  - Step: $\beta = 1.18$, $p < 0.001$
  - Step: Condition (lengthened-shortened): $\beta = 0.54$, $p = 0.01$

# Discussion

- Studies on perceptual learning and speech normalization were usually discussed in the lines of different theoretical frameworks.
- The study provides preliminary evidence of their interaction, i.e., listeners learn and generalize speaker-normalized distributions.
- Our findings shed lights on the possibility of incorporating speech normalization mechanisms into current perceptual learning models (e.g., Kleinschmidt & Jaeger, 2015)

# Thank you!

- Reach us at

*weilai.phonetics@gmail.com*

*liaini@sas.upenn.edu*