Northwestern

Delayed effects of speech and non-speech stimuli on sibilant categorization

Eleanor Chodroff[1] and Colin Wilson[2]

[1]Department of Linguistics, Northwestern University, [2]Department of Cognitive Science, Johns Hopkins University

JOHNS HOPKINS UNIVERSITY

Introduction

Adaptation to the speech of a novel talker can involve at least two types of mechanism: **perceptual adaptation to phonetic properties** and **spectral contrast effects**.

Previous studies have demonstrated that listeners can adapt to talker- or dialect-specific properties of fricatives e.g., Norris et al. 2003, Kraljic & Samuel 2005, Eisner & McQueen 2006, vowels e.g., McQueen & Mitterer 2005, Maye et al. 2008, Reinisch & Sjerps 2013, Chladkova et al. 2017, and stops e.g., Kraljic & Samuel 2006, Nielsen 2007, Theodore et al. 2010.

Moreover, perceptual adaptation to properties of speech has been shown to persist over long periods of time.

- 25 minutes between exposure and test: [s]-[ʃ] Kraljic & Samuel 2005
- 12 hours between exposure and test: [s]-[f] Eisner & McQueen 2006

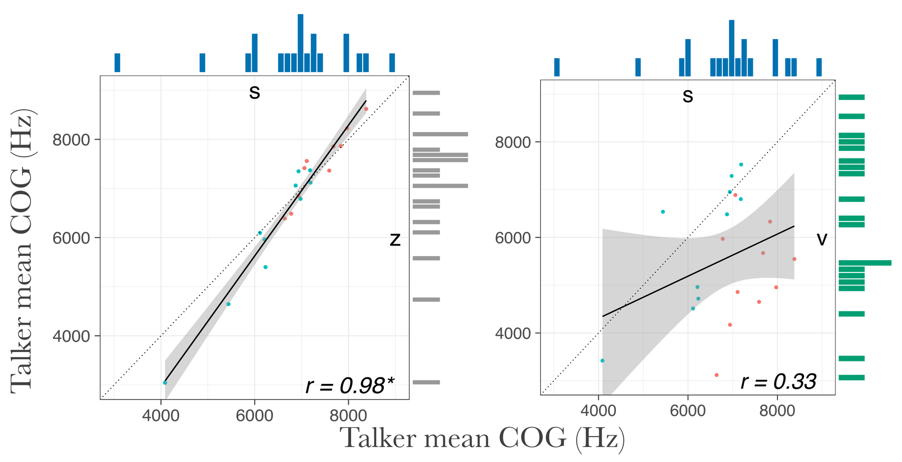
Strong effects of nonspeech stimuli have also been found on perception of following speech sounds e.g., Lotto & Kluender 1998, Holt 2005, Laing et al. 2012, Huang & Holt 2012

The longest attested period for nonspeech effects on speech adaptation is 1.3 seconds, but longer periods have not yet been tested. Holt 2005

Accounts of phonetic- and auditory- based adaptation make similar predictions regarding the expected direction of adaptation:

Covariation-based adaptation

- Listeners infer talker-specific parameters for each sound in a way that takes into account covariation of category cues. Ex. If observe high COG [z], infer high COG [s]



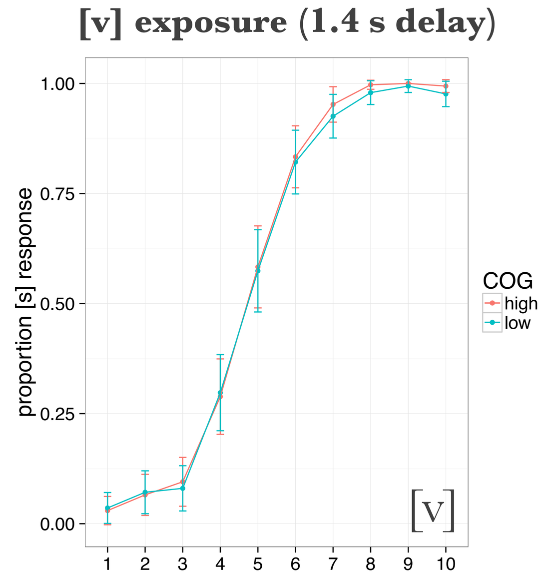
Data from Jongman et al., 2000

Contrast-based adaptation

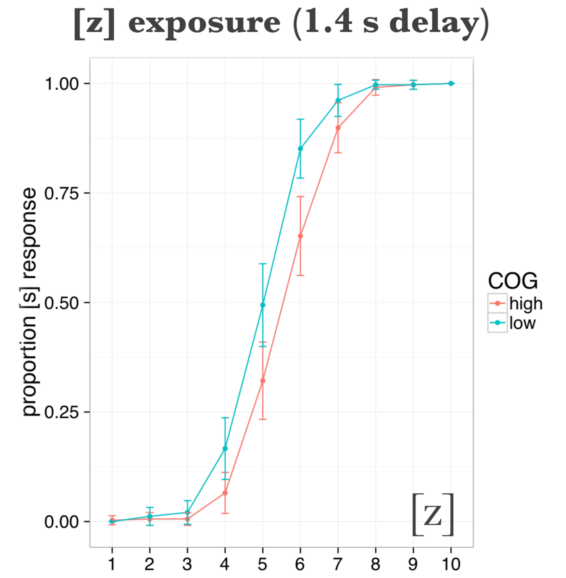
- *High* frequency energy in a preceding sound should enhance *low* frequency energy present in a subsequent sound (and vice versa), shifting perception *contrastively*
- Adaptation should occur only when context sounds have energy in the frequency ranges that are relevant for perception (discrimination or categorization) of targets
- Non-speech contexts should elicit the same effects as matched speech context

Chodroff 2017, Chodroff & Wilson, in prep

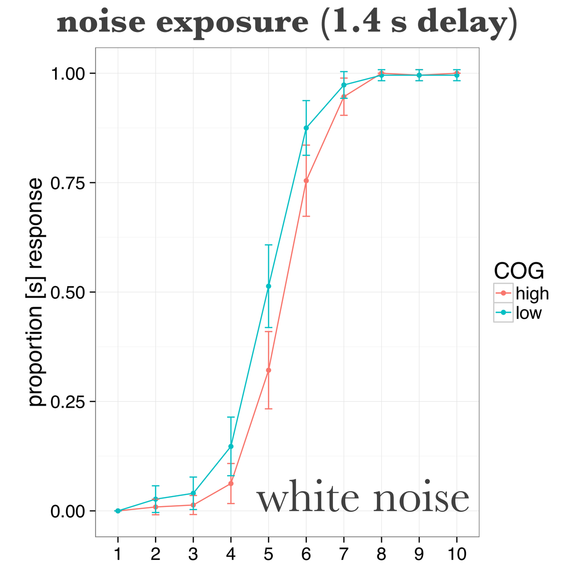
[ʃ]-[s] categorization after [v] exposure (1.4 s delay)



[ʃ]-[s] categorization after [z] exposure (1.4 s delay)



[ʃ]-[s] categorization after white noise exposure (1.4 s delay)



Can phonetic and auditory mechanisms be distinguished by introducing a substantially longer delay between exposure and test in adaptation to the [s]-[ʃ] contrast?

Methods

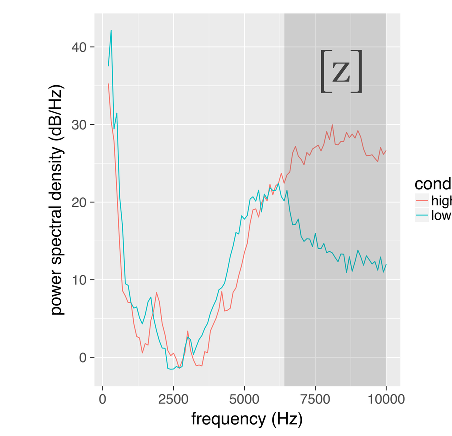
Speech stimuli: [z]-initial CVC syllables created through concatenation of natural recordings from 3 female speakers in a laboratory corpus

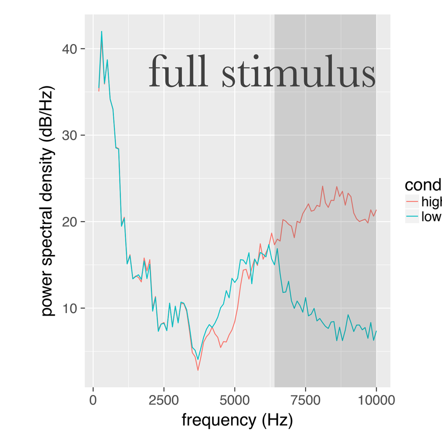
Speaker 1	z – high COG	Speaker 3: vowel [i ɪ eɪ ɛ æ ʌ ɑ ɔ o u] 65 dB	t
Speaker 2	z – low COG		

65, 85, 125 ms
65 dB
different [z] for each vowel

[z] only
Mean COG > 550 Hz
High: 8576 (587)
Low: 6217 (785)

Full stimulus
mean COG (Hz) > 550 Hz
High: 5920 (2271)
Low: 2552 (1292)





Noise stimuli: white noise matched in LTAS, duration, and amplitude to CV portion of [z]-initial syllable

Synthesized [ʃ] - [s] continuum: 10-step Bark interpolation between spectral peaks and slopes for ambiguous [ʃ] and [s]

Shaded regions correspond to Bark critical bands 20-21: potential critical region for contrast-based effects

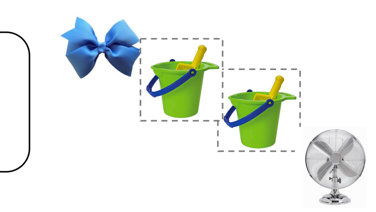
Procedure:

1) Exposure
One-back repetition task

High/low speech
High/low noise
No exposure

6 blocks of 30 trials | Rep. rate of 0.1 | ~7 minutes

2) Image Task
One-back repetition task



200 images | Rep. rate of 0.1 | ~14 minutes

3) s/ʃ Test
2AFC Categorization

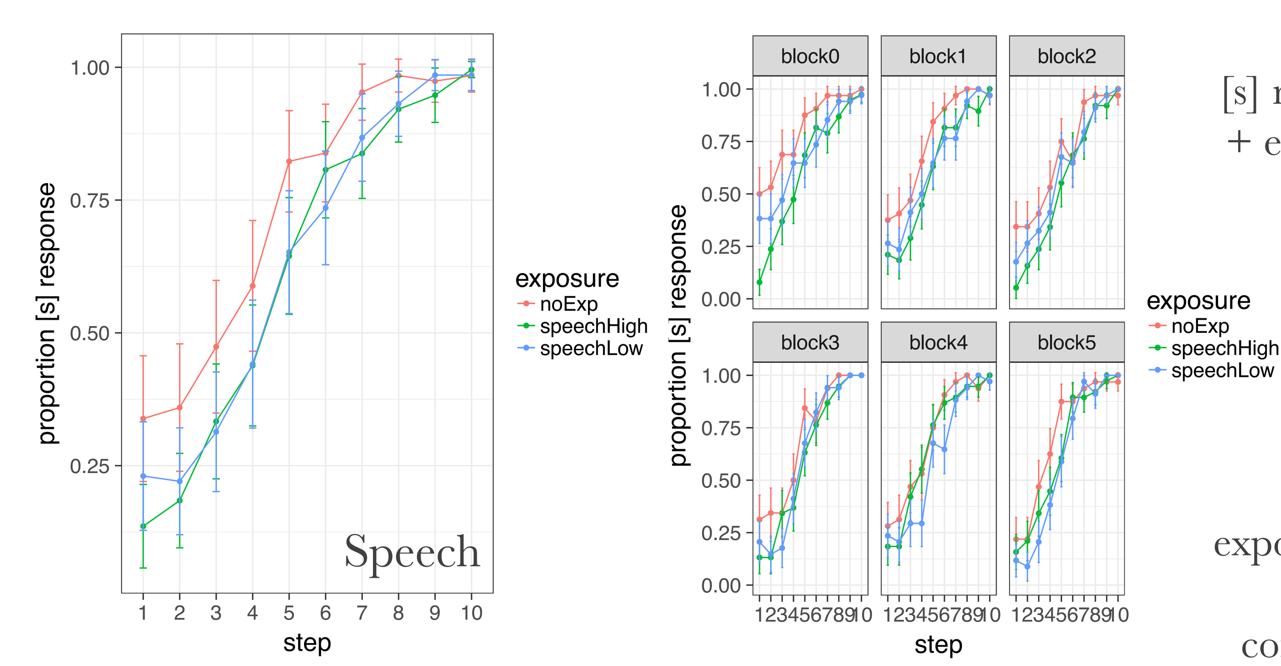
s ʃ

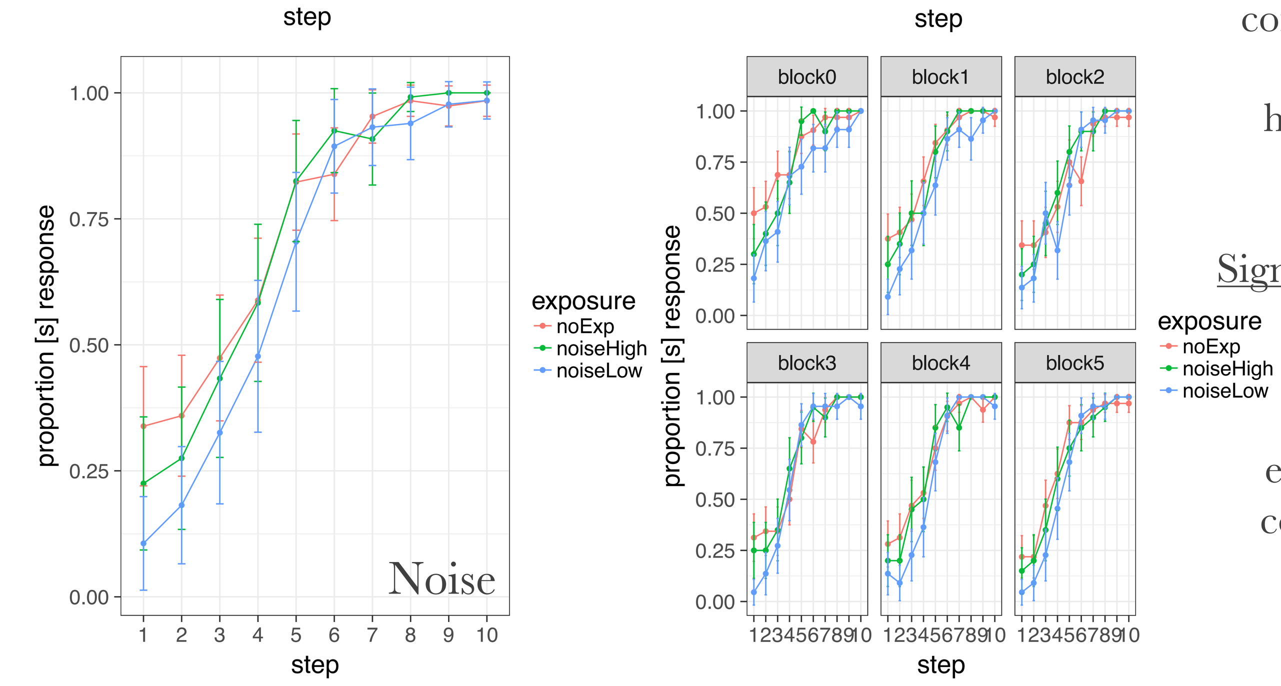
seat-sheet
suit-shoot

6 blocks of 20 trials | ~4 minutes

Participants: 19 in high speech, 17 in low speech, 16 in no exposure, 10 in high noise, 11 in low noise

Results





Model structure

[s] response ~ 1 + step + vowel + exposure + condition + half + exposure*half + condition*half + (1|participant)

step (scaled)
vowel ([i] 0.5, [u] -0.5)
exposure (speech 0.5, noise -0.5, none 0)
condition (high 0.5, low -0.5, none 0)
half (first 0.5, second -0.5)

Significant effects / interactions

step | β = 2.69
vowel | β = -1.52
half | β = 0.19
exposure*half | β = -0.34
condition*half | β = -0.43

Listeners less likely to respond [s] after speech than noise in the first half.
Listeners less likely to respond [s] after high than low COG exposure in the first half.

A speech-only model revealed that listeners were significantly less likely to respond [s] after high COG [z] exposure than low COG [z] exposure in the first half.
A noise-only model revealed no significant effect of condition (high vs low COG noise) in either half of the experiment.

Discussion

Evidence for phonetic covariation-based adaptation over an extended time period (14 min)

- Significant effect of speech condition (high vs low) present within the first half of the experiment
- Direction was consistent with covariation-based adaptation: listeners with exposure to a high COG [z] were *less likely* to respond [s] (i.e., expectation that the talker has a very high COG [s])

Evidence for general auditory contrast-based adaptation over short time periods (msec to sec)

- Effect of condition and exposure disappeared in the second half of the experiment

Implications and future directions:

- Design of phonetic learning and categorical perception studies (beware general auditory effects!)
- At what time period do linguistic and non-linguistic mechanisms start to diverge?
- How do linguistic mechanisms emerge if contrast-based adaptation is constantly present?
- Need for additional participants

References (selected)

Chodroff, E. (2017). Structured variation in obstruent production and perception. PhD Thesis. / Chodroff, E., & Wilson, C. (in prep). Auditory and acoustic-phonetic mechanisms of adaptation in the perception of sibilant fricatives. / Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time (L). *JASA*, *119*(4), 1950–1953. / Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psych. Sci.*, *16*(4), 305–312. / Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cog. Psych.*, *51*(2), 141–178. / Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619. / Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cog. Psych.*, *47*(2), 204–238.