Exploring the Automaticity of Speech Perception and Adaptation

Rachel Sabatello[1], Shawn Cummings[2], & Florian Jaeger[1]

[1]University of Rochester, Department of Brain and Cognitive Sciences

[2]University of Connecticut, Speech, Language, and Hearing Sciences

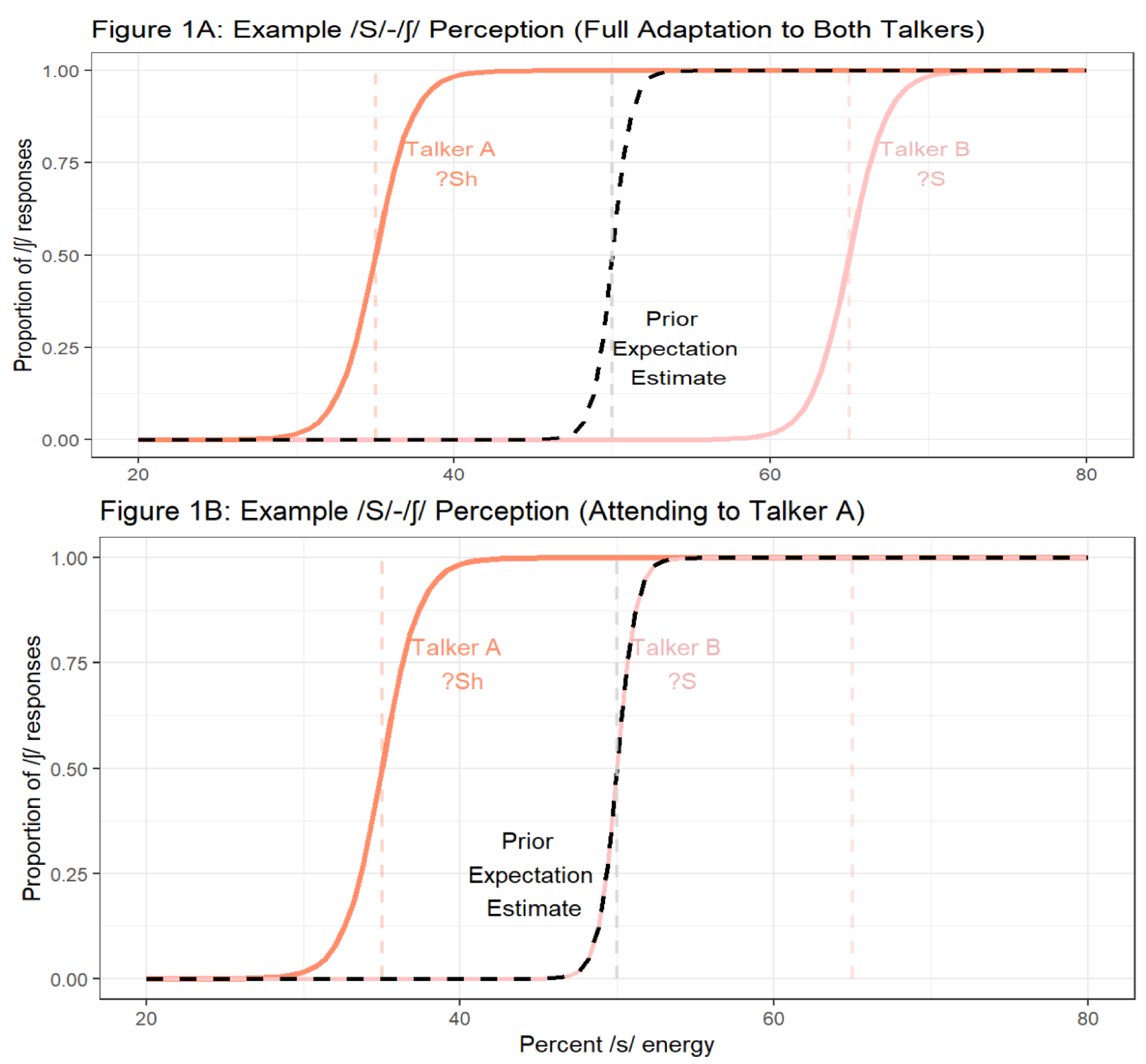
Background

Despite spoken language being highly variable, listeners can often understand newly encountered talkers when hearing them speak for the very first time. Variation in speech presents a unique challenge for cognitive processing that is solved seemingly automatically: Our brains learn how talkers speak, and then apply this information to construct expectations about speech they encounter in the future. This process often occurs without the listener even noticing. However, this phenomenon presents the question of **how automatic is speech perception adaptation?** Are we constantly processing any speech we happen to hear in our environment? In this study, we will explore the automaticity of speech perception and adaptation when participants' available attentional resources are limited. To achieve this, we will expose listeners to two talkers speaking simultaneously, and test the effects of **directing the listener's attention to *one talker*** on the listener's **ability to adapt to *both talkers*.**

Hypothesis

We hypothesize that speech perception adaptation is contingent upon attention. If there are limits to the automaticity of speech perception, then we expect listeners will adapt their perceived categorical boundary to align better with the speech of the talker they are instructed to attend to compared to the unattended talker.

Predictions



Figures 1A & 1B: Graphical representations of listeners responses to the asi-ashi test continuum during the Test Phase. As the percent of /s/ energy in the stimulus increases, the more likely participants are to respond "ashi" (/ʃ/), rather than "asi" (/s/). If perceptual adaptation is dependent on attentional resources, we anticipate listener' adjustment to the unattended talker will be constrained.

Implications

A listener's perceptual boundary changing more to fit the attended talker's speech than the unattended talker's speech would suggest there are limits to the automaticity of speech perception. Conversely, complete adaptation to both talkers would suggest that humans automatically adapt their perception to any speech in their environment.

Design

In this study, we will be measuring listeners' perceptual adaptation to two simulated talkers' S-ʃ production.

S-ʃ sounds exist on a continuum, spanning from /s/ as in "Sock" to /sh/ as in "Shock." Earlier research suggests that listeners' adaptation to S-ʃ production is **talker-specific,** meaning that listeners adjust their perceived boundary between S-ʃ for each talker (Kraljic & Samuel, 2005). In contrast, listener's judgement of other sound categories can be influenced by and applied to multiple talkers. **This quality of how the S-ʃ is perceived could allow us to simulate two distinct talkers with different S-ʃ productions during the same experimental exposure** (Cummings & Theodore, *accepted).*

Exposure Phase

Critical Trials

Our critical stimuli are created from 40 recordings of S/ʃ words, each **spoken typically (S, Sh)** and **accented (?S, ?Sh)** (Kraljic & Samuel, 2005).

All recordings were processed using Praat (Boersma, 2002) to simulate the words being **spoken by a male talker** and **a female talker** (Luthra et al., 2021).

These words were split in half to **create two sets of words representing two talkers**:10 unique S words and 10 unique ʃ words were allocated to each talker (*see below).*

Talker A

ʃ	S
Ambition	Pregnancy
Machinery	Democracy
Brochure	Embassy
Official	Legacy
Crucial	Reconcile
Pediatrician	Personal
Flourishing	Eraser
Reassure	Episode
Graduation	Literacy
Vacation	Coliseum

Talker B

ʃ	S
Initial	Parasite
Beneficial	Obscene
Neogtiate	Medicine
Commercial	Tennessee
Parachute	Peninsula
Efficient	Hallucinate
Publisher	Arkansas
Glacier	Compensate
Refreshing	Dinosaur
Impatient	Rehersal

Figure 2: The lists of s and ʃ words that will be produced in Talker A's voice (left) and Talker B's voice (right).

Each experiment will have a male and a female talker:
If Talker A is female, then Talker B is male.
If Talker A is male, then Talker B is female.

Talker A and Talker B recordings were paired to create Materials A and Materials B. Half of the participants will hear the words in Materials A with the simulated accent (?S, ?Sh), and the words in Materials B without the accent (S, Sh). The other half of the participants will hear the inverse, meaning Materials B will be accented and Materials A will not be *(See below).*

The word pairings shown horizontally across in Materials A & B were then spliced together to **create stereo audio files** where one talker is played in the left ear, and the other in the right ear. Like talker gender and which materials are accented, ear assignment will be counterbalanced across participants.

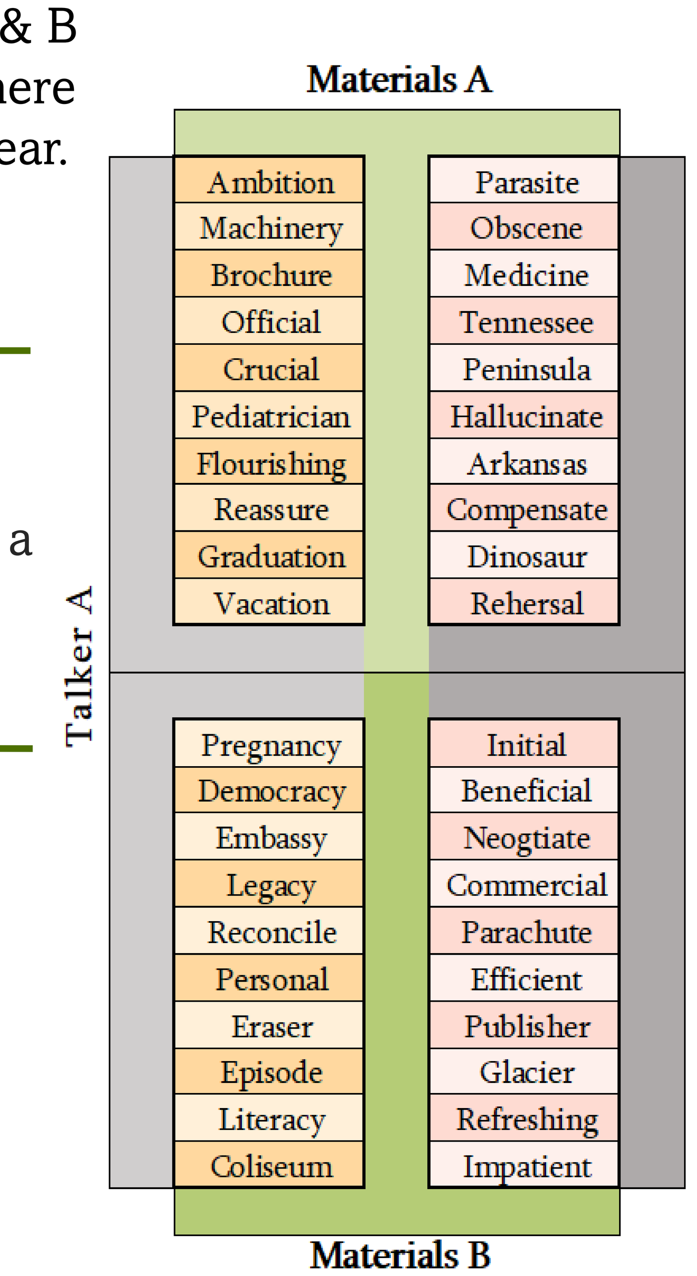


Figure 3: A visual illustrating how the words spoken by Talker A and Talker B will be paired to produce two sets of materials.

Filler Trials

Each experiment consists **of 80 total exposure trials**, including **20 critical** trials and **60 filler** trials. During filler trials, one talker will say a word and the other talker will say a nonword. The attended talker will have a 50% chance of saying a nonword.

Paradigm

Participants will be instructed to attend to either the female talker or the male talker. They will then perform a series of 2-option forced-choice lexical decision tasks, in which they will hear a recording and then select if the attended talker said a word or a nonword (*see below).*

Figure 4: A static representation of how a participant will progress through a trial. Each trial will begin with the participant hearing an audio file, and then selecting either "Word" or "Nonword".

Test Phase

After the Exposure Phase, participants will hear the asi-ashi test continuum across trials in each talker's voice. This continuum is used to gauge when listener's shift from perceiving a sound as "Sh" to "S" (*see right*). Each trial will only play a recording from a single voice at once. Participants will select if the audio they heard was "asi" or "ashi" for each trial to produce results like the predictions, shown to the left.

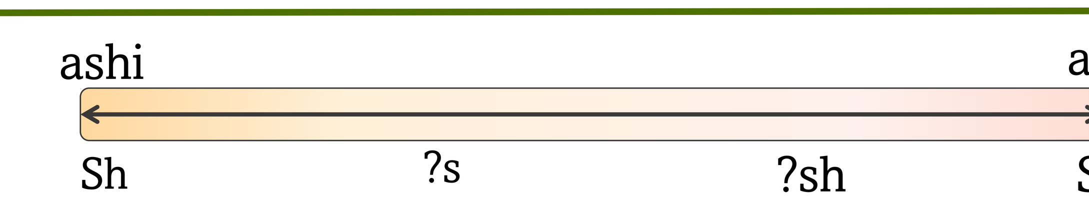


Figure 5: "S" and "Sh" sounds exist on a spectrum, where "asi" can be altered to sound like "ashi" by changing the percentage of /s/ energy.

References

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Cummings, S. N. & Theodore, R. M. *(accepted).* Perceptual learning of multiple talkers: Detriments, characteristics, and limitations. *Attention, Perception, & Psychophysics.*

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive psychology*, *51*(2), 141-178.

Luthra, S., Mechtenberg, H., & Myers, E. B. (2021). Perceptual learning of multiple talkers requires additional exposure. *Attention, Perception, & Psychophysics*, *83*, 2217–2228.

Tzeng, C. Y., Nygaard, L. C., & Theodore, R. M. (2021). A second chance for a first impression: Sensitivity to cumulative input statistics for lexically guided perceptual learning. *Psychonomic Bulletin & Review, 28*, 1003–1014.

Acknowledgements

Thank you to Dr. Tanya Kraljic and Dr. Arthur Samuel for their permission to use the stimuli they developed (Kraljic & Samuel, 2005), the 2022 Meliora Mentors, the HLP Lab and the University of Rochester Brain & Cognitive sciences department.

This project was funded by the University of Rochester Wiesman summer fellowship in brain and cognitive sciences.