Are you paying attention?
The cognitive cost of speech perception adaptation

Rachel Sabatello

Introduction
Despite spoken language being highly variable, listeners tend to understand most talkers when hearing them speak for the first time. Variation in speech production 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 the speech they will encounter in the future (Kleinschmidt & Jaeger, 2015). This is a cognitive process that often occurs without the listener noticing. However, the absence of awareness brings into question the potential limitations of this ability: are listeners always passively sponging information about talker's speech production from their environment, or must they be directing their attention towards a given talker? The goal of this experiment is to investigate the automaticity of speech perception adaptation; specifically, how does directing attention to one talker compete with adaption to a second talker when both are speaking at the same time.

Background
A large body of research suggests that perceptual learning is not inhibited by lack of intention, distractions (Zhang & Samuel, 2014), or exposure to multiple talkers (Cummings & Theodore, 2022). However, earlier research has also found that listeners use context to moderate their adaptation, such as applying stereotype-specific expectations and using labeling to adjust their expectations (Zhang & Samuel, 2014). Additionally, there is evidence that listeners consider causality when learning how talkers speak—e.g., accommodating a talker chewing on a pen while talking (Kraljic & Samuel, 2011). The latter processes indicate that listeners may store information based on its perceived relevance. Therefore, in our experiment when there is competition for auditory processing resources, we expect the processing of "useful" information to take precedence.

This theory is echoed in the findings of Dr. Samuel's 2016 paper. Using a structurally similar paradigm to what we propose, his findings strongly support that adaptation is involuntary and robust to distractions unless the competing task requires some form of categorization of the auditory information. In Experiment 1, participants were exposed to two simulated talkers. One talker produced speech with a phonetically shifted s/sh and was always presented before the 2nd talker. The 2nd talker did not produce a phonetic shift. Samuel found that when the onset of the 2nd talker interrupted the 1st talker and the participant performed a lexical recognition task for the 2nd talker's speech, the participant did not exhibit adaptation. This was also true in Experiment 3, where the 2nd talker was replaced with an environmental sound (e.g., a doorbell) and the participant was asked to identify this sound. However, when the lexical recognition task was performed for the 1st talker's speech in Experiment 1, the participant did exhibit adaptation despite still being interrupted by the 2nd talker (Samuel, 2016). By engaging in a categorization task that requires information from one of multiple competing audial streams, the perceived utility of that stream is elevated above the others. This suggests that adaptation is only automatic given that cognitive resources are available—i.e., not occupied with processing another talker's speech.

Specific Aims/Hypotheses
We hypothesize that perceptual adaptation to speech is contingent on a listener's attention being directed towards a given talker. In this experiment, we therefore aim to simulate two distinct talkers that a listener will hear speak simultaneously. The listener will perform a lexical recognition task for one of these talkers (referred to as the Attended Talker henceforth), and then we will compare participants' adaptation to **both** the Attended Talker and the Unattended Talker. If perceptual adaptation requires a listener to be attending to the talker, then we would expect participants to only adapt to the attended talker *(Figure 1A)*. If perceptual adaptation does not require attention directed towards a given talker, then we would expect participants to exhibit adaptation for both talkers *(Figure 1B)*. We would not expect to see this trend in the data based on the results in Samuel 2016. It is also possible that hearing simultaneous speech may obstruct adaptation if the listener cannot separate the two verbal streams (*Figure 1C),* however this is unlikely due to precautions described in the methodology.

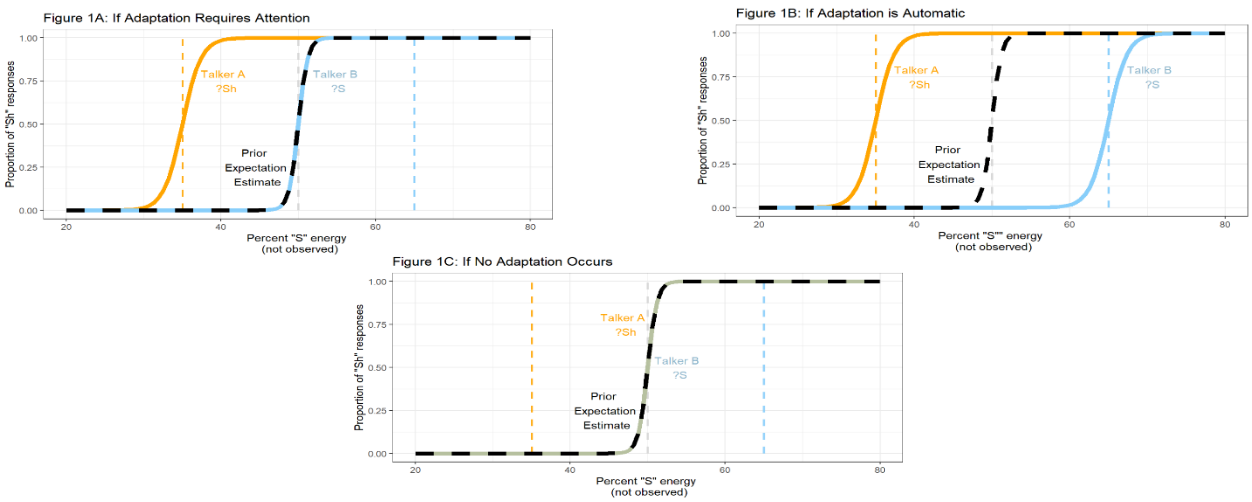

Figure 1: Representations of listeners' potential responses to the asi-ashi test continuum during the Test Phase. As the percent of /s/ energy in the stimulus increases, the more likely participants are to respond "ashi" ("sh"), rather than "asi" ("s"). Here Talker A represents the Attended Talker and Talker B represents the Unattended Talker.

Significance

This study would expand on the findings of Samuel 2016 in several ways. Our paradigm removes the stimulus onset asynchrony (SOA) previously employed in Samuel 2016, resulting in both talkers' speech overlapping perfectly. This should increase the difficulty of the task—when speech naturally coincides, it rarely occurs simultaneously and is not usually limited to one word—and will allow us to draw conservative conclusions about the role of attention in adaptation. By doing so, we can also investigate listeners' ability to separate the talkers and maintain attention, whereas Samuel 2016 focuses primarily on the latter. Removing the SOA also negates the potential influence of order on listeners' ability to hone in on a stimulus.

Our paradigm also introduces an atypical talker in place of the 2nd talker in Samuel 2016. As a result, we will be able to compare adaptation between talkers. If our hypothesis is not proven false, this paradigm could be used in the future to investigate features that may cause a listener to prioritize one verbal stream over another. By extension, this may have further implications for language learning and the possible effects of social biases on speech processing. If our hypothesis is proven false—contrary to our expectations based on prior research—then this novel finding would suggest that speech perception adaptation occurs automatically when the brain encounters any form of human speech. This could have informative implications for how we theorize the brain actively collects, stores, and uses information to formulate expectations.

Methods

In this study, we will be measuring participants' perceptual adaptation to two simulated talkers' s/sh productions. s/sh sounds exist on a continuum, spanning from "s" as in "**S**ock" to "sh" as in "**Sh**ock." Earlier research suggests that listeners' adaptation to s/sh production is talker-specific, meaning that listeners adjust their perceived boundary between s/sh for each talker (Kraljic & Samuel, 2005). This quality prevents cross-talker contamination in adaptation, allowing us to simulate two distinct talkers with different s/sh productions during the same experimental exposure (Cummings & Theodore, 2022).

Our stimuli were adapted from those developed in Kraljic & Samuel, 2005. Critical items include two versions a word that contain an s/sh sound: one with typical s/sh production, and one where the s/sh sound is replaced by an ambiguous s/sh sound (e.g., "dino**s**aur" becoming "dino**sh**aur," or "publi**sh**er" becoming "publi**ss**er"). Filler items are either real words or nonwords that followed the typical pattern of real words. They do not contain any s/sh sounds. We used the audio processing software Praat (Corretge, 2022) to change the gender of the voice in the recording (Luthra et al., 2021), and then to pair both the critical items and the filler items so one talker is heard in the left ear of a headset and the other in the right. This resulted in a set of audio files that simulates two distinct talkers: a female talker and a male talker in opposite spatial positions on either side of the participant. Items, gender, and positioning factors will be counterbalanced across participants.

In each experiment, participants first experience an exposure phase that will consist of 80 randomized trials. 20 of these trials will be critical trials, and 60 trials will be filler trials. Participants will be instructed to attend to either the female talker or the male talker, and to then perform a lexical recognition task (see *Figure 2).* Participants will then experience a 72-trial test phase in which they will hear both talkers independently produce a six-increment s/sh test continuum 6 times. They will perform a 2AFC lexical decision task for each trial to determine the categorization boundary for both talkers. We will compare the responses for both talkers within-participant (see *Figure 1).* If there are limits to the automaticity of speech perception, then we can expect listeners will adapt their perceived categorical boundary to align better with the attended talker's speech compared to their adjustment for the unattended talker. Conversely, complete adaptation to both talkers could suggest that speech perception adaptation is automatically shaped by any speech in a listener's environment.

Figure 2: Experimental Design

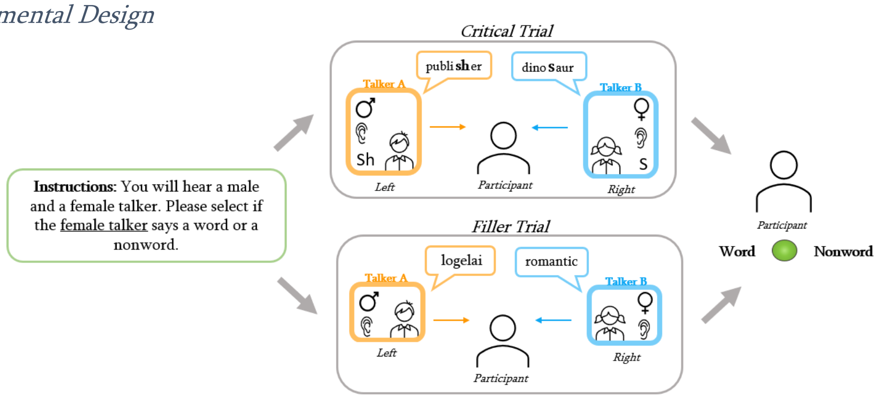

Figure 2: A visual diagram of a single trial. Participants will be instructed to attend to either the male or female talker at the beginning of the experiment (left, green box). Each subsequent exposure trial will feature two talkers, either in a critical trial (top) where both talkers produce a word that contains a s/sh sound, or a filler trial (bottom) where one talker produces a word and the other a nonword. The participant must then select if the Attended Talker produced a word or a nonword for each exposure trial (right).

Roles

I greatly appreciate my PI's help thus far with shaping the question I proposed and providing guidance in crafting the experimental design. Moving forward, my PI has agreed to run the necessary java coding portion of the experiment on Prolific, help me code my analyses in Rstudio, and provide feedback on my thesis.

Timeline

I began developing this project over the summer the with support of the Wiesman Fellowship. Since then, I have developed my question and methodology, and have created my stimuli. My target timeline is:

Week of:	Experiment Goals:	Thesis Goals:
01.08.23	Launch experiment by the end of January	Draft of Intro
01.22.23		
02.05.23	Debug experiment *(should take about 2 weeks)*	Draft of Methods
02.19.23		
03.05.23	Experiment Live by early March *(2-3 days needed for data collection)*	Revise Intro + Methods
03.12.23	data analysis, conclusions, and	
03.19.23		Draft Results + Conclusions
04.02.23		Revise Results + Conclusions
04.09.23		Polish Final Paper
04.14.23	**Thesis due at 5pm!**	

References

Corretge, Ramon (2022). Praat Vocal Toolkit. https://www.praatvocaltoolkit.com

Cummings, S. N., & Theodore, R. M. (2022). Perceptual learning of multiple talkers: Determinants, characteristics, and limitations. *Attention, perception & psychophysics*, 84(7), 2335–2359. https://doi.org/10.3758/s13414-022-02556-6

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. https://doi.org/10.1037/a0038695

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. Cognition, 121(3), 459-465.

Luthra, S., Mechtenberg, H., & Myers, E. B. (2021). Perceptual learning of multiple talkers requires additional exposure. *Attention, Perception, & Psychophysics*, 83(5), 2217–2228. https://doi.org/10.3758/s13414-021-02261-w

Samuel, A. G. (2016). Lexical representations are malleable for about on second: Evidence for the non-automaticity of perceptual recalibration. *Cognitive Psychology*, 88, 88–114. https://doi.org/10.1016/j.cogpsych.2016.06.007

Zhang, X., & Samuel, A. G. (2014). Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 200–217.