**Are You Paying Attention?**

**The Cognitive Costs of Speech Perception Adaptation**

Rachel Sabatello1

1Department of Brain & Cognitive Sciences, University of Rochester

**Abstract**

**Are You Paying Attention?**

**The Cognitive Costs of Speech Perception Adaptation**

[Introduction to the lack of invariance problem]

Despite being one of the most prevalent forms of human communication in daily life, spoken language is highly variable. Even talkers with similar language backgrounds tend to differ in how they produce speech sounds, for example how they may distinguish /s/ (the “S” sound) from /ʃ/ (the “Sh” sound) (X). Still, listeners can often understand newly encountered talkers when hearing them speak for the very first time (X). How we are able to flexibly categorize speech sounds despite the presence of natural variation is known as the *lack of invariance problem* (X). 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 (Kleinschmidt & Jaeger, 2015). This cognitive process often occurs without the listener even noticing. However, the absence of awareness brings into question potential limitations of this ability: do listeners passively sponge information from speech in their environment, or must they direct their attention towards a talker to learn how they speak?

[Discussion of research that supports the proposal of this question]

(add contrary that inspired Samuel 2016?)

A large body of research suggests that perceptual learning does not require conscious effort, nor is it inhibited by environmental distractions (Zhang & Samuel, 2014) or exposure to multiple talkers (Cummings & Theodore, 2022). However, earlier research has also found that listeners may use the context of speech when adapting, for instance the effects of stereotypes on initial expectations and labeling in phoneme mapping (Zhang & Samuel, 2014). There is even evidence that listeners consider causality when learning how talkers speak—e.g., accommodating a talker chewing on a pen while talking (Kraljic & Samuel, 2011). These latter findings could suggest that listeners store information based on a perceived utility. Therefore, in our experiment when there is competition for auditory processing resources, we expect that processing of “useful” information— information framed as being more relevant to the participant’s task—will take precedence.

[Samuel 2016: findings, similarities]

This theory is echoed in the findings of Dr. Samuel’s 2016 paper, which strongly supports 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/-/ʃ/ 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 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.

[Our Study]

In our study, we will investigate the role of attention in speech perception and adaptation by limiting the participant’s available attentional resources. To this end, we will expose listeners to two simulated 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.

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. Both talkers have been engineered to have distinct voices, and also produce inversely atypical “accents” on the S-ʃ continuum in their speech (i.e., one talker produces their /s/ sounds more like “Sh,” and the other talker produces their /ʃ/ sounds more like “S”). Participants will be instructed to attend to one of the two talkers throughout the exposure phase of the experiment, and then select if that talker is saying a word or a nonword in a forced-choice lexical decision task. As a result, we will be able to compare adaptation between talkers.

We will then test the effects of directing the participants’ attention during exposure to one talker on the participants’ ability to adapt their speech perception to both talkers. During the test phase, we will measure how participants categorize sounds on the S-ʃ spectrum using a set of categorization tasks that feature a S-ʃ test continuum. Then we will compare how participants respond to these tasks for each of the two simulated talkers. 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. No matter which of these theories is supported by our results, this study has the potential to pioneer future paradigms for studying speech perception adaptation and develops a foundation for further research on how human cognition prioritizes speech processing.

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?

***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. If perceptual adaptation does not require attention directed towards a given talker, then we would expect participants to exhibit adaptation for both talkers. 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, however this is unlikely due to precautions described in the methodology.

**Methods**

In this study, we measure participants’ perceptual adaptation to two simulated talkers’ /s/-/ʃ/ productions. /s/-/ʃ/ sounds exist on a continuum, spanning from an “s” sound (e.g., “**S**ock”) to an “sh” sound (e.g., “**S**hock”). This continuum is determined by spectral energy, where /ʃ/ is produced with more spectral energy than /s/ in English (X). However, the perceptual boundary between /s/-/ʃ/--when a “s” sound begins to be perceived as a “sh” sound and vice versa--is variable across listeners (Norris, McQueen, & Cutler, 2003), and potentially even across talkers. 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). This quality also prevents cross-talker contamination in adaptation, allowing us to potentially simulate two distinct talkers with different /s/-/ʃ/ productions during the same experimental exposure (Cummings & Theodore, 2022; Kraljic & Samuel, 2007).

For the purpose of this experiment, we will assign an atypical production to both simulated talkers: one talker will produce a typical “s” sound and an atypical “sh” sound (“?sh”, as in Publi**ss**er), while the other will produce a typical “sh” sound and an atypical “s” sound (“?s”, as in Dino**sh**aur). The “?s” and “?sh” sounds were created by combining /s/ and /ʃ/ recordings (Kraljic & Samuel, 2005).

Many earlier studies (Cummings & Theodore, 2022; Kraljic & Samuel, 2007; Trude & Brown-Schmidt, 2012; Luthra et al., 2021) have simulated two unique talkers within the same experimental exposure by presenting one talker as female-sounding, and another talker as male-sounding. We adopted this same strategy to distinguish the two talkers. Additionally, we separated the vocal streams of the talker by ear

**Materials**

The stimuli used in this experiment were adapted from the female stimuli set developed by Dr. Tanya Kraljic and Dr. Arthur Samuel (Kraljic & Samuel, 2005). These stimuli included 20 critical words that included a /s/ sound (e.g., Para**s**ite, pronounced /pærə**s**aɪt/) and 20 critical words that included an /ʃ/ sound (e.g., Ambi**ti**on, pronounced /æmbɪ**ʃ**ən/). These same 40 words were also produced with an ambiguous sound that, to the typical Native U.S. speaker, is perceived as if s/sh are being swapped (e.g., Para**sh**ite, pronounced /pærə**ʃ**aɪt/; Ambi**s**on, pronounce /æmbɪ**s**ən/). Additionally, we used 64 filler words that contained no /s/, /ʃ/, // sounds, and 98 nonwords that followed the typical structure of English words and contained no /s/-/ʃ/ sounds. A full list of all the words/nonwords used in this experiment is available in *Appendix A*.

We are using the 6-step continuum from Cummings et al, Experiment 1 rather than the original Kraljic & Samuel stimuli. We do so because the original stimuli seem to elicit more ASHI than ASI responses and the Cummings et al stimuli are supposedly more balanced in this regard. Using steps 8, 12, 13, 14, 15, 19 (Cummings et al., 2022)

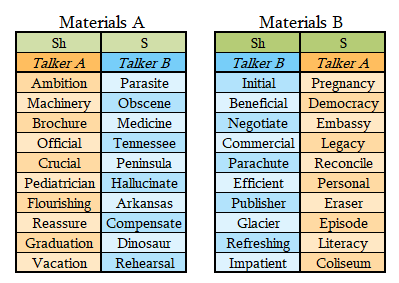
200ms of silence was added to all items (Hirst et al., 2018). (critical, filler, and test) used in Experiment 1 due to issues with the initial segments of the audio being cut off when listened to using Bluetooth headphones in early tests.

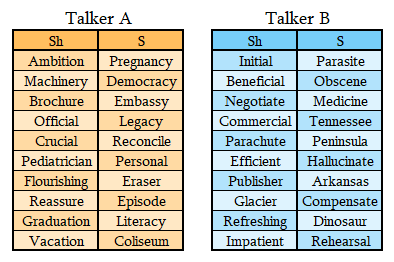
Critical items include two versions a word that contain a /s/-/ʃ/ sound: one with typical /s/-/ʃ/ production, and one where the /s/-/ʃ/ sound is replaced by an ambiguous /s/-/ʃ/ sound (e.g., “dinosaur” becoming “dinoshaur,” or “publisher” becoming “publisser”). Filler items are either real words or nonwords that followed the typical pattern of real words. They do not contain any /s/-/ʃ/ sounds.

We used Praat (Boserman, 2002) Vocal Toolkit (Corretge, 2022) to change the gender presentation of the voice in the recordings following the procedure for synthesizing a typical male voice and a typical female voice described in Luthra et al. (2021): To transform a recording to a typical male voice, we

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 ear positioning factors will be counterbalanced across participants.

**Design**





The experiment is divided into an Exposure Phase and a Test Phase. The exposure phase includes 10 blocks of 8 trials

*Trial types*

This experiment is separated into an exposure phase and a test phase.

*Exposure*

Each participant experiences a total of 80 trials in the exposure phase, divided evenly into 10 blocks.

*Test*

Each participant experiences a total of 72 test trials in the exposure phase, divided evenly into 12 blocks.

*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.*

**Participants**

A total of 64 participants (M = 32; age range 18-68, avg. = 34.483) were recruited from the online crowdsourcing program Prolific. Participants were self-reported English monolinguals, who were native to and spent most of their time in the United States before the age of 18. Additionally, participants had not participated in any other experiments launched by the Human Language Processing Lab on Prolific. The purpose of this criteria was to minimize variation in the perceptual boundary between S-Sh relative to the simulated talker’s shifts.

Before proceeding to the experiment, participants were asked to confirm that they spoke United States English most of the time before they were ten years of age in the experiment instructions. Participants were instructed to complete the study in one sitting, in a quiet room with minimal distractions. They were informed that they could not take the experiment multiple times or reload the experiment and were provided with reasons their work could be rejected. Each participant was provided with the participant consent form and were required to agree to proceed with the experiment.

Additionally, participants were informed that over-ear or in-ear speakers (e.g., headphones) would be necessary for this experiment, oppose to external speakers (e.g., laptop speakers). These instructions were followed by a 6-trial 3-AFC tone comparison task (Woods et al., 2017) to confirm that participants were using proper, functional hardware. To continue to the experiment, participants were required to correctly identify the softest tone in five of the six trials.

***Data Collection***

**Table 1: Participant recruitment break-down**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Recruitment  (date) | Start  Time | Experiments  Completed | Experiments  Returned | Experiments  Timed Out | Median  Time | Notes |
| 02.19 | 15:45 | 10 | 15 | 0 | 16:18 | Pilot 1: Data collection error: response not recorded in CSV file |
| 03.06 | 14:45 | 1 | 0 | 0 | 15:00 | Pilot 2a: Correct recording error |
| 03.06 | 16:48 | 1 | 2 | 0 | 15:00 | Pilot 2b: Gender selection balance |
| 03.06 | 20:23 | 53 | 48 | 1 | 16:11 | Experiment crashed on 1 participant before they were able to begin the experiment |

*Table 1: When participants were recruited, how many participants completed the experiment during a recruitment period, and how many participants did not complete the experiment. Also included is the medium time it took participants to complete the experiment during each recruitment and notes about the purpose and problems encountered during the recruitment period.*

***Exclusions prior to analyses***

A total of 131 participants engaged with this study on Prolific. Of these participants, 64 completed the experiment. 66 participants requested to take the experiment but decided not to complete it. Reasons for this may be that the task was regarded as too difficult for the reward, the participants lost interest in the task, or the participants failed the 3-AFC equipment check. 1 participant was excluded on 03.06 because they failed to complete the experiment in 56 minutes, the maximum time allotted by Prolific based on our estimated completion time of 15 minutes. 1 participant contacted us and was excluded on 03.06 due to a technical malfunction that resulted in them being unable to begin the experiment.

**Procedure**

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/-/ʃ/ test continuum 6 times. They will perform a 2-AFC lexical decision task for each trial to determine the categorization boundary for both talkers. We will compare the responses for both talkers within-participant.

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.

***Practice***

***Exposure***

***Test***

**Exclusion Criteria**

Participant data was excluded if they

* Incorrectly responded to the survey question of which talker they were attending to
* Report using improper equipment
* Percentage of incorrect Word/Nonword responses (above 50% 🡪 above chance)
* Duration of experiment
  + Trial level exclusions = trial duration?

**Results**

**[INCLUDE WHICH LISTS WERE RUN? APPENDIX??]**

**Discussion**

***Implications***

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.

**Acknowledgements**

Thank you to Dr. Tanya Kraljic and Dr. Arthur Samuel for their permission to use the stimuli they developed (Kraljic & Samuel, 2005), 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.

**References**

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

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

Cummings, Karboga, Yang, & Jaeger (2022). Causal Inference in Speech Perception. [Manuscript in preparation]

Cummings, S. N., & Theodore, R. M. (2022). Perceptual learning of multiple talkers: Determinants, characteristics, and limitations. *Attention, perception & psychophysics*, 84(7), 23352359.https://doi.org/10.3758/s13414-022-02556-6

Hirst, D., Qi, Z., & Daidone, D. (2018). Praat script Insert silence. <https://www.ddaidone.com/uploads/1/0/5/2/105292729/insert_silence_at_start_of_all_files_in_folder.txt>

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. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, 121(3), 459-465.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1-15.

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.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2), 204-238.

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>

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27(7-8), 979-1001.

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.

Woods, K.J.P., Siegel, M. H., Traer, J. & McDermott, J. H. (2017) Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics*.

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. <https://doi.org/10.1037/a0033182>

APPENDIX A

Words/Nonwords used in stimuli

**Critical S Words**

1. Ambition
2. Machinery
3. Brochure
4. Official
5. Crucial
6. Pediatrician
7. Flourishing
8. Reassure
9. Graduation
10. Vacation
11. Initial
12. Beneficial
13. Negotiate
14. Commercial
15. Parachute
16. Efficient
17. Publisher
18. Glacier
19. Refreshing
20. Impatient

**Critical Sh Words**

1. Parasite
2. Obscene
3. Medicine
4. Tennessee
5. Peninsula
6. Hallucinate
7. Arkansas
8. Compensate
9. Dinosaur
10. Rehearsal
11. Pregnancy
12. Democracy
13. Embassy
14. Legacy
15. Reconcile
16. Personal
17. Eraser
18. Episode
19. Literacy
20. Coliseum

**Filler Words**

1. America
2. Bakery
3. Ballerina
4. Blueberry
5. Bullying
6. Burglary
7. Camera
8. Continually
9. Directory
10. Document
11. Domineering
12. Dynamite
13. Eighty
14. Embody
15. Gardenia
16. Grammatical
17. Gullible
18. Hamburger
19. Honeymoon
20. Hurdle
21. Identical
22. Illuminate
23. Inhabit
24. Interior
25. Ironic
26. Keyboard
27. Knowingly
28. Laminate
29. Lengthen
30. Lethal
31. Liability
32. Lingering
33. Lobbying
34. Marina
35. Melancholy
36. Membrane
37. Metrical
38. Napkin
39. Negate
40. Nightmare
41. Ornament
42. Outnumber
43. Panicking
44. Parakeet
45. Pilgrim
46. Pineapple
47. Platonic
48. Predict
49. Purgatory
50. Tactical
51. Terminal
52. Therapeutic
53. Titanium
54. Turbulent
55. Tutorial
56. Umbrella
57. Undermine
58. Wealthy
59. Worldly
60. Wrinkle

**Filler Nonwords**

1. Youmgel
2. Wominid
3. Wojalto
4. Ungelnin
5. Tymolape
6. Tounamplem
7. Tilegkalo
8. Tamical
9. Ryligal
10. Rimkeluwar
11. Rengime
12. Rawamtee
13. Rakil
14. Pourilar
15. Ploupelai
16. Perkum
17. Omperoge
18. Nomerae
19. Niritaly
20. Nempring
21. Namuary
22. Mibgem
23. Meidnow
24. Makid
25. Logelai
26. Lilgrai
27. Lenediaw
28. Kloumidiger
29. Kermimer
30. Kelabidel
31. Kaldemia
32. Itempider
33. Inpaki
34. Imdalier
35. Ibirak
36. Hintarber
37. Halomimoc
38. Halken
39. Gerbualo
40. Gardimuallay
41. Ganla
42. Galliwinou
43. Gairelom
44. Emhoutic
45. Dilkuaund
46. Dadigal
47. Bowidai
48. Bikanian
49. Bamtell
50. Baliber
51. Anolipa
52. Anemer
53. Amalar
54. Alnadiro
55. Almikquary
56. Aknid
57. Ailounam
58. Aigi
59. Admunker
60. Acomining

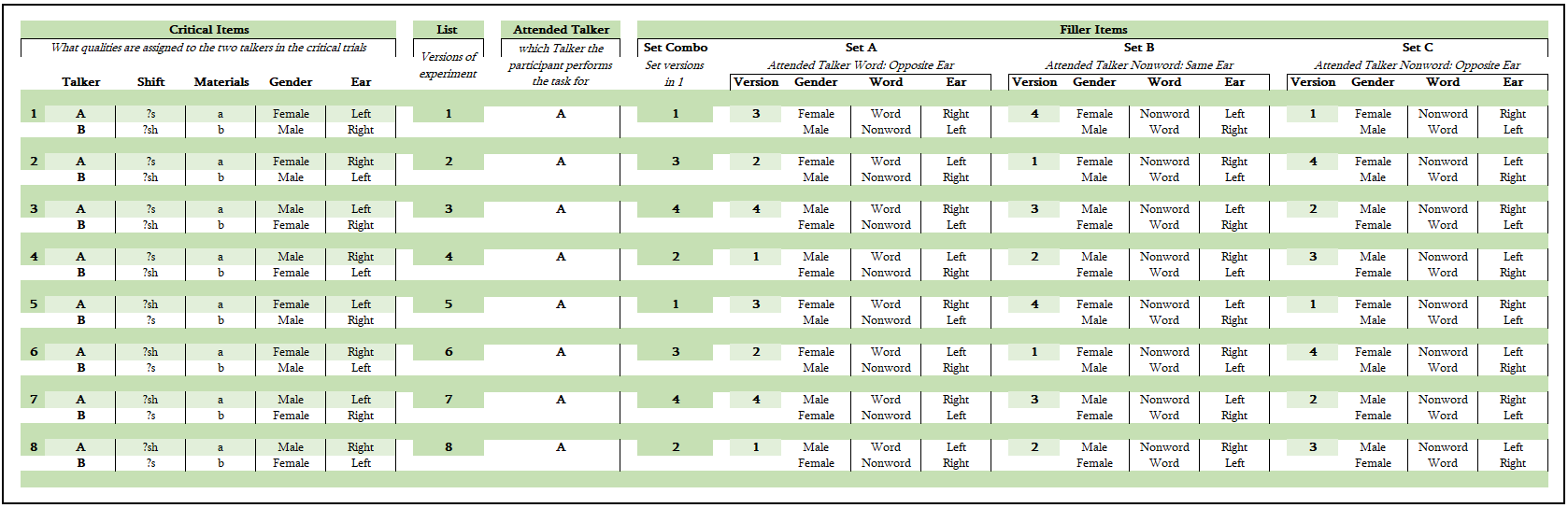
**Practice Words**

1. Corridor
2. Heroine
3. Parable
4. Younger

**Practice Nonwords**

1. Nawinow
2. Neramgory
3. Kerkrum
4. Rikmaral

*Appendix A:* All audio recordings were produced by the same female talker for Kraljic & Samuel (2005). Recording numbers reflect the item pairs: Critical S Words 1-20 is paired with Critical Sh Words 1-20; Filler Words 1-60 are paired with Filler Nonwords 1-60; Practice Words 1-4 are paired with Practice Nonwords 1-4. To synthesize a typical male-presenting voice, we applied a format shift ratio of 0.8 to the recordings and new pitch median of 100 Hz to the recordings. To synthesize a typical female-presenting voice, we applied a format shift ratio of 1 and a new pitch median of 180 Hz to the recordings (Luthra et al., 2021).



*Appendix B:* A table depicting the possible Critical Item and Filler Item combinations. The factors we anticipate potentially confounding our results are the Attended Talker *Shift, Gender,* and *Ear.* For this experiment, the Attended Talker is always the talker assigned Materials A. The *Shift* column describes what ambiguous sound each talker produces (either ?s or ?sh), the *Gender* column describes the simulated voice assigned to the talker (male of female), and the *Ear* column describes which ear of the headset that the talker’s voice is produced in. Every combination of *Shift* X *Gender* X *Ear* produces a total of 8 *List* combinations. Based on these factors, a *Set Combo* of Filler Items is selected. Each Set includes 20 Filler Items, with 20 Filler Words and 20 Filler Nonwords. The words within a given set share the same *Gender* X *Word X* *Ear* configuration. The *Gender* X *Word X* *Ear* factors applied to the Set is described by the *Version* column*.* The *Gender* column describes the simulated voice assigned to the talker (male or female), the *Word* column describes whether the talker produces Words or Nonwords within those 20 Filler Items, and the *Ear* column describes which ear the talker is heard in. The *Gender* of the talkers is consistent across all Filler Sets. The 20 Filler Items in *Set A* have the Attended Talker producing a Word in the opposite Ear assigned to the Critical Items. The Filler Items in *Set B* have the Attended Talker producing a Nonword in the same Ear as assigned to the Critical Items. The Filler Items in *Set C* have the Attended Talker producing a Nonword in the opposite Ear assigned to the Critical Items. As a result, each participant hears the Attended Talker produce a Word for 50% of the trials, hears the attended talker in the Right Ear for 50% of the trials, and hears the Attended Talker produce a Word in the Right Ear for 25% of the trials.

APPENDIX B

Counterbalancing of non-nuisance factors

APPENDIX C

Participant Instructions

**Attentional effects on listening**

Thank you for your interest in our study! This is a psychology experiment about how people understand speech. You will listen to recorded speech, and press a button on the keyboard to tell us what you heard.

**Please read through each of the following requirements. If you do not meet all requirements, please do not take this experiment.** You can click the names below to expand or close each section.

### **Experiment length**

The experiment takes 15-20 minutes to complete and you will be paid $3.20

### **Language requirements (grew up speaking American English)**

You must be a native speaker of American English. **If you have not spent almost all of your time until the age of 10 speaking English and living in the United States, you are not eligible to participate.**

### **Environment requirements (quiet room)**

Please complete this experiment in one sitting and in a quiet room, away from other noise. Please do NOT look at other web pages or other programs while completing this experiment. It is important that you give this experiment your full attention.

### **Headphone check**

Please complete the following headphone test to make sure your audio setup is compatible with this experiment, and that your headphones are set to a comfortable volume.

First, set your computer volume to about 25% of maximum. Press the button, then **turn up the volume on your computer until the calibration noise is at a loud but comfortable level**. Play the calibration sound as many times as you like:

**Trial 1/6**

Which sound was the softest?

 **1st** sound  **2nd** sound  **3rd** sound

### **Additional requirements**

**Please do NOT take this experiment multiple times, and do NOT reload this page.** If you share an MTurk/Prolific account with others who have taken this experiment, please make sure that they have not yet taken this experiment. We cannot use data from reloaded or repeated experiments, and won't be able to approve your work.

We use cookies and MTurk/Prolific qualifications to make it easy for you to recognize whether you have taken this experiment previously. If you accept our cookies and do not delete them, this should prevent you from accidentally taking the experiment more than once.

### **Reasons work can be rejected**

If you pay attention to the instructions and **do not respond randomly**your work will be approved. **Please do NOT reload this page, even if you think you made a mistake.** We will not be able to use your data for scientific purposes, and you will not be able to finish the experiment. We anticipate some mistakes will be made, but those will NOT affect the approval of your work.

We will only reject work if you a) **clearly** do not pay attention to the instructions, b) reload the page, or c) repeat the experiment. We reject far less than 1% of all completed experiments.

### **Experiment instructions**

The purpose of this experiment is to investigate listeners’ ability to pay attention to a specific talker when there are multiple talkers speaking at once.  
  
The experiment has two parts. In the first part, you will hear recordings of a female and a male talker speaking simultaneously. Your task is to **focus only on the female talker**. For each recording, you have to determine whether the female talker produced a word or a non-word.

In the second part, you will hear recordings from the same two talkers. This time, each recording will only contain speech from one talker at a time.

### **Informed consent**

By accepting this experiment, you confirm that you have read and understand the [consent form](https://www.hlp.rochester.edu/experiments/consent/RSRB45955_Consent_2024-01-12.pdf), that you are willing to participate in this experiment, and that you agree that the data you provide by participating can be used in scientific publications (no identifying information will be published). Sometimes we share non-identifying data elicited from you — including sound files — with other researchers for scientific purposes (your MTurk/Prolific ID will be replaced with an arbitrary alphanumeric code).

### **Begin the experiment**

### Once you press the green button, these instructions will disappear, so make sure you understand them fully before you click.

I confirm that I meet the requirements for this experiment, that I have read and understood the instructions and the consent form, and that I want to start the experiment.

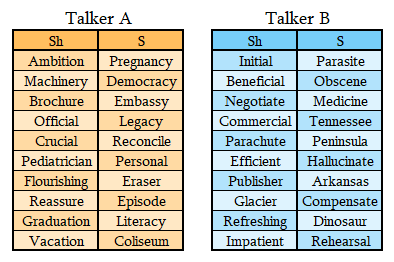
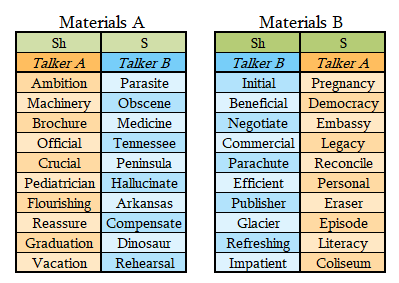
### **Further (optional) information**

Sometimes it can happen that technical difficulties cause experimental scripts to freeze so that you will not be able to submit a experiment. We are trying our best to avoid these problems. Should they nevertheless occur, we urge you to (1) take a screen shot of your browswer window, (2) if you know how to also take a screen shot of your Javascript console, and (3) [email us](mailto:hlplab@gmail.com) this information along with the HIT/Experiment ID and your worker/Prolific ID.

If you are interested in hearing how the experiments you are participating in help us to understand the human brain, feel free to subscribe to our [lab blog](http://hlplab.wordpress.com/) where we announce new findings. Note that typically about 1-2 years pass before an experiment is published.

**Figure 2**

**Figure 1**

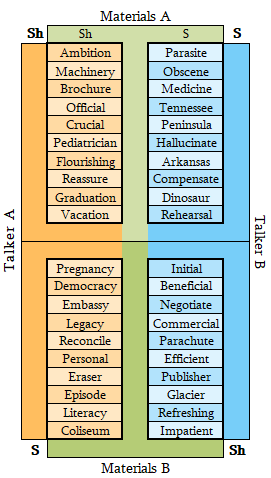
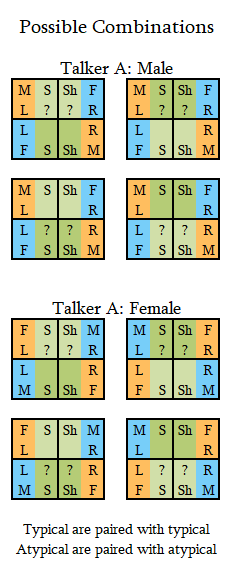


*Figure 1: The lists of s and ʃ words that will be produced in Talker A’s voice (left) and Talker B’s voice (right).*

*Figure 2: How Talker words are paired into Critical Items and grouped into Materials A (left) and Materials B (right) to assign an ambiguous sound to half the Critical Items.*

**Figure 3b**

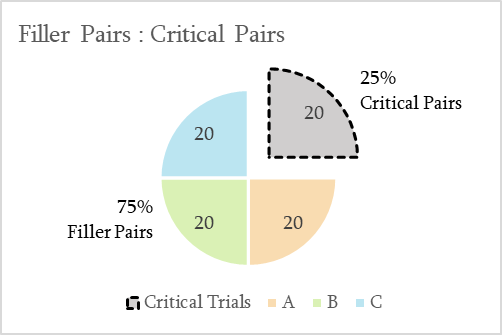
**Figure 3a**



*Figure 3a: A visual illustrating how the Critical Words spoken by Talker A and Talker B are divided into Materials A and Materials B. In this experiment, the words in Materials A are produced with an ambiguous sound and the words in Materials B are produced with an unambiguous sound. During the Critical Trials, each Talker is always heard in the same ear (e.g., Talker A in the Left Ear and Talker B in the Right Ear, as Shown above).*

*Figure 3b: Potential Configurations of factors. The “??” symbols represent which materials are assigned the ambiguous sound.*

Textbox

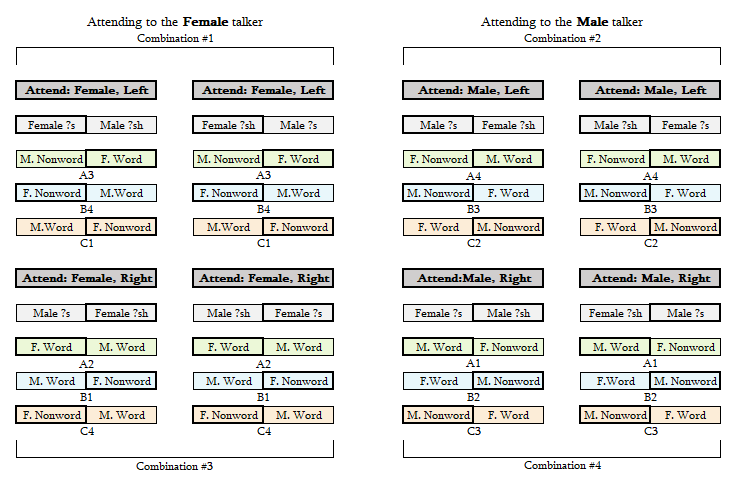


**Figure 4c**

Diagram

Description automatically generated

**Figure 4a**

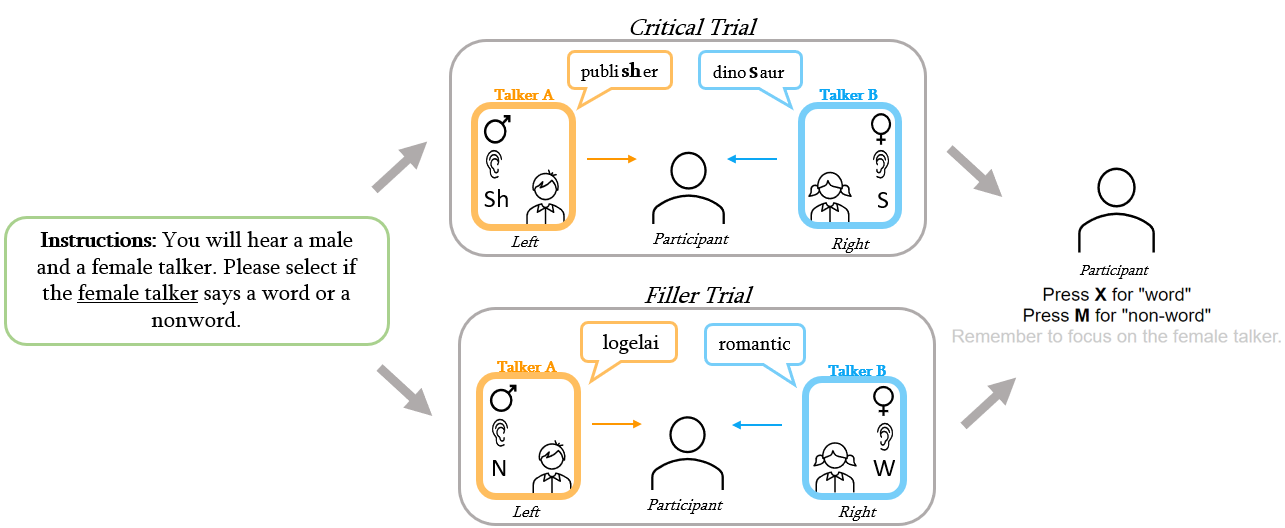


**Figure 4b**

Textbox

Textbox

**Figure 5**



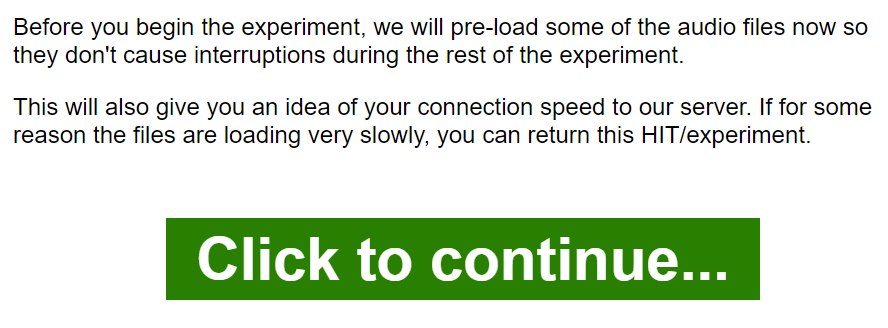


Figure 5: A visual diagram of a single trial. Participants will be instructed to attend to either the male or female talker at the beginning of the experiment (left, green box). Each subsequent exposure trial will feature two talkers, either in a critical trial (top) where both talkers produce a word 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).

