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? The goal of this experiment is to explore the automaticity of speech perception and adaptation when participants’ available attentional resources are limited. Specifically, how directing attention to one talker’s speech competes with adaption to a second talker speaking simultaneously in the environment.

The first challenge faced in designing this experiment was how to simulate two talkers with comparable atypical productions without the listener’s adaptation to one talker influencing their adaptation to the other. To answer the question of the level of automaticity behind speech perception, we will ask listeners to listen to *one* of the talkers, and then measure adaptation to *both* talkers and compare those results. Theoretically, we expect adaptation to the attended talker’s speech –which would serve as a means of comparison– and less-to-no adaptation for the unattended talker. However, listener’s judgement of many sound categories can be influenced by and applied to multiple talkers. Our experiment would require us to prevent a listener’s adaptation to one talker from influencing their adaptation to the other talker.

Our solution is to simulate both talker’s atypical production on a continuum that is *talker specific*. A speech sound continuum being talker-specific means that listeners adjust their perceived boundary between the two sounds independently for each talker. The talker specific continuum we will be using is S-ʃ (“s” as in “Sock” to “sh” as in “Shock”), where pronouncing the initial “s” sound in significant like “**sh**ignificant” would be an example of an atypical production. In their 2005 paper, Kraljic & Samuel created a set of stimuli for S-ʃ that was readily available. These audio recordings were al produced by the same female talker

The stimuli included:

* **Critical Words**
  + 20 “s” words
  + 20 “sh” words
  + An ambiguous version of each of the 20 “s” words
  + An ambiguous version of each of the 20 “sh” words
* **64 Filler Words** (real words with no “s” or “sh” sounds)
* **98 Filler Nonwords** (fake words that were created to sound as if they could be real words)
* **6 Item Asi-Ashi Test Series** (varying the percentage of “s” energy)
  + TASHI 25 *(S)*
  + TASHI 30
  + TASHI 40
  + TASHI 45
  + TASHI 50
  + TASHI 70 *(Sh)*

The second challenge we faced was using this set of stimuli to simulate two distinctly different voices that can simulate simultaneous speech that is *distinguishable*. We approached this hurdle in 3 ways:

1. We decided to manipulate the simulated voices so one would mimic a traditionally female-sounding voice, and the other would be traditionally male-sounding.

To accomplish this, all of the audios that will be used in this experiment were run through Praat. The male voice was created by changing the format shift ratio to 0.8, and the new pitch median to 100. The female voice was created by preserving the format shift ratio of 1.0 and making the new pitch median 180 (luthra et al., 2021).

1. Creating stereo files from the individual talker stimuli so one talker is heard more one ear, and the other talker in the opposite ear.

Outside of a laboratory environment, this would be akin to one talker is at the listener’s left side, while the other talker is present on the right. This will theoretically improve the listener’s ability to disentangle the talkers’ speech due to the simulated spatial differences.

1. The words within each stereo file were selected to sound relatively different from each other.

We wanted the two words within each stereo file to sound somewhat different from each other and wanted each talker to have a unique set of words that does not repeat. Our goal is to further separate the two voices so the listener will perceive them as two independent talkers.

Thus, we want to simulate an S-ʃ continuum for both talkers without repeating the same words *across talkers*. We decided that one talker would be assigned 10 of the “s” words and 10 of the “sh” words. The other talker would be assigned the remaining 10 “s” words, and the remaining 10 “sh” words. These words were listed alphabetically, and then every other word was assigned to one of the two talkers:

Table

Description automatically generatedTable

Description automatically generated

A picture containing text, scoreboard

Description automatically generatedTable

Description automatically generatedNext, each of Talker A’s “s” words were paired with one of Talker B’s “sh” words, and each of Talker B’s “s” words with Talker A’s “sh” words. Words with the same initial sound or with similar sounds either before or after the “s”/“sh” sound were not paired together. This resulted in two sets of materials: **Materials A** and **Materials B**.

Diagram

Description automatically generatedIntroducing different words for each talker and using two different genders for the talkers introduces factors that we would like to counterbalance. In half the participants, the male talker will produce the words assigned to Talker A and the female talker produce the words assigned to Talker B. This will be inverted for the other half of the participants. Additionally, we also need to counterbalance the atypical S-ʃ production. In half the experiments, the words in Materials A will be expressed using the ambiguous version of the recording and Materials B will be expressed using the unambiguous recordings. This will also be inverted for half the experiments.

Additionally, we also will counterbalance which talker and which materials are played in which ear (2x2x2). The result is 8 different list designs.

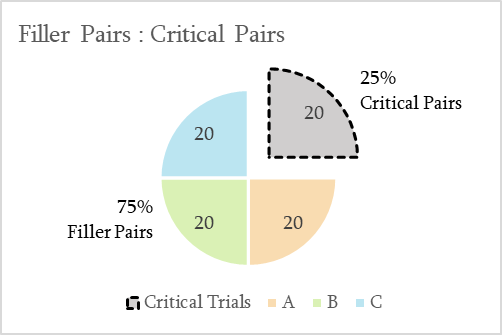
Each list will also contain *filler trials*, which will be words/nonword pairs where neither have an “s” or an “sh” sound. We will be including these trials to structure the task that participants will be asked to do for this experiment. For us to draw conclusions about the automaticity of speech perception from our results, we must first confirm that participants are able to hear and distinguish the talkers. We decided to use a word-nonword forced-choice lexical decision task so that we have a measure of the participant’s ability to distinguish the two voices.

At the beginning of the experiment, participants will be instructed to attend to either the male or the female talker, and then to respond to the lexical decision task for the attended talker’s speech. For our analyses, we would want to include filler trials that give the participant a 50% chance of responding correctly on all trials if they were to respond randomly. Additionally, we would want to avoid the critical trials being the only trials where the attended talker is saying a word, as both talkers will be saying a real word during those trials. Finally, we should also counterbalance which ear the talker is speaking in to avoid participants removing one side of their headset to make the task easier.

For there to be a 50% chance that the attended talker is in one ear, plus a 50% chance that the attended talker is also saying a word, the critical trials would have to make up 25% of the trials in the experiment. This would result in 80 total trials, where 60 trials are fillers.

Like with the critical trials, we also do **not** want to repeat audios (words/nonwords) *within or across voices*. Rather than creating 60 filler pairs that each use unique words/nonwords for each of the 8 lists, we attempted to take a more systematic approach:

First, we divided the total 60 unique word/nonword pairs into 3 equal sets of 20 to create *Sets A, B, and C*.



Next, we made 4 versions of each set:

*Version 1*

* The **Female Talker** is in the **Left Ear** and produces a **Nonword**
  + The **Male Talker** is in the **Right Ear** and produces a **Word**

*Version 2*

* The **Female Talker** is in the **Left Ear** and produces a **Word**
  + The **Male Talker** is in the **Right Ear** and produces a **Nonword**

*Version 3*

* The **Male Talker** is in the **Left Ear** and produces a **Nonword**
  + The **Female Talker** is in the **Right Ear** and produces a **Word**

*Version 4*

* The **Male Talker** is in the **Left Ear** and produces a **Word**
  + The **Female Talker** is in the **Right Ear** and produces a **Nonword**

Diagram

Description automatically generated

Then we chose the appropriate version of each set to create each of the 8 lists. *Which version was selected is dependent on the Ear and Gender of the Attended Talker in the* ***Critical Trials.***

* For **Set A**, the attended talker spoke in the **Opposite Ear** as in the critical trials and produced a **Nonword**.
* For **Set B**, the attended talker spoke in the **Opposite Ear** as in the critical trials and produced a **Word**.
* A screenshot of a computer

  Description automatically generated with medium confidenceFor **Set C**, the attended talker spoke in the **Same Ear** as in the critical trials and produced a **Nonword**.

This approach guarantees that no words or nonwords are repeated between or across talkers within a given trial and works within the confines of the original stimuli from Kraljic & Samuel, 2005.

*How can we interpret results based on this design:*

* If the participant clicks **randomly**, they will be ~50% correct for either talker.
* If the participant **always listens to the unattended talker and responds correctly** for that talker, they will be correct 25% of the time for the attended talker\*

*\*This is because the correct response for both the attended talker and the unattended talker is Word for all of the critical trials*

* If the participant **always listens to the attended talker and responds correctly** for that talker, they will be correct 100% of the time for the attended talker.
* If the participant **responds Word for 100% of the trials**, they will be correct for the attended talker 50% of the time\*

*\*And correct 75% of the time for the unattended talker*

* If the participant **responds Nonword for 100% of the trials**, they will be correct for the attended talker 50% of the time\*

*\*And correct 25% of the time for the unattended talker*

* If the participant **always listens to Ear A** (the ear that the Attended Talker produces critical trials in), they will be correct 50% of the time for the attended talker.
* If the participant **always listens to Ear B** (the ear that the Unattended Talker produces critical trials in), they will be 75% correct of the time for the attended talker.

**If significantly more than 50%** of the participant’s responses are correct, then we can rule out the participant’s responses being correct due to chance.

**If significantly more than 75%** of the participant’s responses are correct, then we can rule out the participant attending to the wrong ear.

For the test continuum, there are six items on one full test phase. Each talker will speak the full continuum six times, for a total of twelve test blocks, where the six trials are heard in both ears, spoken by a single talker, before continuing to the next block. The blocks will be counterbalanced so that the half the participants hear the female talker first, and the other half the male. To possibly reduce the cognitive load of this task, the talker will swap every two blocks.

At the end of the experiment, we will then measure where the perceptual boundary is between “s” and “sh” for both talkers. Participants will listen to the Asi-Ashi test continuum in both talkers’ voices and judge whether they heard “Asi” or “Ashi” for each trial. These responses will allow us to estimate the categorization boundary between “s” and “sh” for each talker by developing psychometric curves (examples below) of the probability of participants perceiving the sound as “s” (0) or “sh” (1), given the percentage of “s” energy used to create that version of the Asi/Ashi.

Graphical user interface, application, table, Excel

Description automatically generated

*Predicted results if adaptation* ***is not*** *automatic:*

*Adaptation is found* ***only*** *for the attended talker.*

*Predicted results if adaptation* ***is*** *automatic:*

*Adaptation is found for* ***both*** *talkers*

*Predicted results if there is* ***no*** *adaptation to either talker:*

*Simultaneous exposure to two talkers could disrupt the necessary attention to adapt to* ***either*** *talker*

