The Effects of Stimulus Type on Interpersonal and Intrapersonal Speech Perception in Typical Adults

Melannee Wursten Ipsen

Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

BYU ScholarsArchive Citation

Ipsen, Melannee Wursten, "The Effects of Stimulus Type on Interpersonal and Intrapersonal Speech Perception in Typical Adults" (2022). Theses and Dissertations. 9545.

https://scholarsarchive.byu.edu/etd/9545

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
The Effects of Stimulus Type on Interpersonal and Intrapersonal Speech Perception in Typical Adults

Melannee Wursten Ipsen

A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science

Kathryn Cabbage, Chair
Christopher Dromey
Shawn Nissen

Department of Communication Disorders
Brigham Young University

Copyright © 2022 Melannee Wursten Ipsen
All Rights Reserved
ABSTRACT

The Effect of Stimulus Type on Interpersonal and Intrapersonal Speech Perception in Typical Adults

Melannee Wursten Ipsen
Department of Communication Disorders, BYU
Master of Science

Children with speech sound disorders (SSD) often have difficulties with speech perception. Speech perception is the ability to intake speech sounds and interpret them for meaning. Understanding children’s speech perception abilities is pertinent because children use perceptual skills to hone accurate production during SSD treatment. Different types of stimuli have been used in speech perception research. At present, it is unclear how different types of speech stimuli differentially impact speech perception in typical listeners or children with SSD. In this study, we investigated perceptual skills for different speech types in neurotypical adults to better understand how stimulus type impacts perception in individuals without SSD. Thus, we asked the following two research questions: 1) Is there a difference between synthetic speech (generated through a computer) and natural speech perception for adult listeners? 2) Is there a difference in interpersonal (listening to speech from another person) versus intrapersonal (listening to your own speech) natural speech perception for adult listeners? Twenty-five neurotypical adults participated in this study. Participants completed the Wide Range Acoustic Accuracy Scale (WRAAS) discrimination task for syllable pairs beginning with the phonemes /b/-/w/; /d/-/g/; and /r/-/w/ for synthetic speech, and rhyming words beginning with the same phonemes (‘bot’-‘watt’, ‘dot’-‘got’, ‘rot’-‘wot’) interpersonal synthetically altered natural speech (a standard speaker), and intrapersonal synthetically altered natural speech (each participant’s own voice recordings) for nine tasks total. Results show there was no statistical difference in discrimination ability between stimulus types for most phoneme contrasts, except for /d/-/g/ between synthetic and intrapersonal synthetically altered natural speech. There was no difference between interpersonal and intrapersonal perception of synthetically altered natural speech for any phoneme pair. Findings from this study will provide information for future similar studies conducted on children with and without SSD to determine how children perceive different types of speech. This future work will be used to help inform speech therapy decisions for children with SSD who may have speech perception difficulties.

Keywords: speech perception, synthetic, interpersonal perception, intrapersonal perception, speech sound disorders
ACKNOWLEDGMENTS

I would like to thank my committee members Dr. Cabbage, Dr. Dromey and Dr. Nissen for their guidance and support. I would especially like to thank Dr. Cabbage for being an excellent mentor. I have enjoyed learning from her during my time at BYU. I appreciate her sincere enthusiasm for this field, for answering my endless number of questions and for providing me with the tools to succeed. I would also like to thank my friends for lending a listening ear and for all the motivational snack runs! To my family, thank you for your loving encouragement and support throughout this process.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Speech Perception Abilities in Children With Speech Sound Disorders</td>
<td>3</td>
</tr>
<tr>
<td>Factors Affecting Speech Perception Performance</td>
<td>6</td>
</tr>
<tr>
<td>Type of Stimuli: Synthetic Versus Natural</td>
<td>6</td>
</tr>
<tr>
<td>Speech Errors vs Sounds With No Errors</td>
<td>7</td>
</tr>
<tr>
<td>Interpersonal vs. Intrapersonal</td>
<td>8</td>
</tr>
<tr>
<td>Purpose of Current Study</td>
<td>11</td>
</tr>
<tr>
<td>Method</td>
<td>12</td>
</tr>
<tr>
<td>Participants</td>
<td>12</td>
</tr>
<tr>
<td>Descriptive Measures</td>
<td>13</td>
</tr>
<tr>
<td>Hearing</td>
<td>13</td>
</tr>
<tr>
<td>Speech Production</td>
<td>13</td>
</tr>
<tr>
<td>Stimuli</td>
<td>13</td>
</tr>
<tr>
<td>Synthetic Speech</td>
<td>14</td>
</tr>
<tr>
<td>Synthetically Altered Natural Speech</td>
<td>15</td>
</tr>
</tbody>
</table>
Experimental Measures

Procedure(s)

Session 1

Session 2

Results

The /b/-/w/ Phonemic Contrast

The /d/-/g/ Phonemic Contrast

The /r/-/w/ Phonemic Contrast

Discussion

For Phoneme Contrasts in Most Listening Conditions, Listeners Exhibited Equivalent Speech Discrimination

No Difference Between Interpersonal and Intrapersonal Perception for Synthetically Altered Natural Speech

Limitations

Implications for Future Research

Conclusion

References

APPENDIX A: Review of the Literature

APPENDIX B: Participant Consent Form

APPENDIX C: Institutional Review Board Approval Letter
LIST OF TABLES

Table 1  Description of Word Pairs .................................................................16

Table 2  Summary Statistics Grouped by Phoneme Contrast Pair and Stimulus Type ........23
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td><em>Visual Representation of the Stimulus Continuum Endpoints</em></td>
<td>15</td>
</tr>
<tr>
<td>Figure 2</td>
<td><em>Discrimination Data From a Pilot Participant for /ba/-/wa/</em></td>
<td>20</td>
</tr>
<tr>
<td>Figure 3</td>
<td><em>/ba/-/wa/ Convergence Level for Each Stimulus Type</em></td>
<td>24</td>
</tr>
<tr>
<td>Figure 4</td>
<td><em>/da/-/ga/ Convergence Level for Each Stimulus Type</em></td>
<td>25</td>
</tr>
<tr>
<td>Figure 5</td>
<td><em>/ba/-/wa/ Convergence Level for Each Stimulus Type</em></td>
<td>26</td>
</tr>
</tbody>
</table>
DESCRIPTION OF THESIS STRUCTURE AND CONTENT

This thesis, *The Effects of Stimulus Type on Interpersonal and Intrapersonal Speech Perception in Typical Adults*, is written in a hybrid format. The hybrid format brings together traditional thesis requirements with journal publication formats. The preliminary thesis reflects requirements for submission to the university. The thesis report is presented as a journal article and conforms to length and style requirements for submitting research reports to communication disorders’ journals. Excerpts from this thesis may be used for publication with the thesis author being listed as a contributing coauthor. An annotated bibliography is included in Appendix A, participant consent form in Appendix B and the Institutional Review Board Approval Letter in Appendix C.
Introduction

Speech sound disorders (SSDs) are among the most common communication disorders that speech-language pathologists (SLPs) encounter in professional practice (American Speech-Language-Hearing Association [ASHA], 2020). Approximately 89% of school-based clinicians report having children with SSDs on their caseloads (ASHA, 2020). ASHA defines SSD as difficulty with one or more of the following areas: “perception, motor production, or phonological representation of speech sounds and speech segments” (ASHA, n.d.). More specifically, a child with SSD produces speech sound errors beyond an age that is typical. Children with SSDs range in severity from children who produce single sound errors (e.g., difficulty only producing /r/) to children producing errors for many phonemes potentially affecting entire phoneme classes (e.g., a child substitutes plosive phonemes for all fricative phonemes). Children with SSD are at risk for having additional difficulties in the areas of phonological awareness, decoding, spelling, and/or social-emotional wellbeing (Farquharson, 2019).

Surveys have revealed that most SLPs use a traditional articulation approach to help children with SSD remediate their sound errors (Brumbaugh & Smit, 2013; McLeod & Baker, 2014). This approach is grounded in a sensory-motor framework and principles of motor learning and practice to first, ensure a child can distinguish the sensory and articulatory properties of the targeted phoneme, and second, provide ample practice to reinforce motor learning of the phoneme (Cabbage & DeVeney, 2020). Thus, a traditional articulation approach first helps the child to learn what sound is in error through auditory discrimination practice, also known as ear training (Van Riper & Irwin, 1959). The next stage of treatment is for the SLP to teach the proper placement of articulators to produce the correct sound. Throughout this process, the child
needs to listen to, understand and recognize the SLP’s modeled production of the sound and try to match it in their own productions. Finally, to generalize the production, the child uses their own auditory feedback to self-monitor when they are producing the correct sound and when they are not, so they can correct the error in real time. A subset of children with SSD have marked difficulty acquiring accurate production and continue to exhibit residual speech errors despite the provision of adequate therapy (Flipsen, 2015). Speech perception, a child’s ability to process auditory signals through “discrimination, identification, recognition, and judgment” of speech sounds (Hearnshaw et al., 2019) may play a critical role for individuals who do not respond to a traditional articulation approach. It is possible that if a child has difficulty with speech perception, they may have difficulty in multiple stages of treatment for a traditional articulation approach due to the need to correctly perceive the SLP’s modeled production and the reliance on auditory feedback to self-monitor his/her own productions.

At present, it is unclear why some children with SSD respond well to treatment and other children continue to have difficulty despite the provision of adequate therapy. A recent study has demonstrated a relationship between perception skills and therapy outcomes (Preston et al., 2020). Preston and colleagues (2020) used ultrasound visual feedback (UVF) in conjunction with a traditional articulation approach to treat older children with residual speech errors in /r/ production. The researchers specifically investigated whether treatment that included UVF alone versus UVF + auditory perceptual training would yield different gains in therapy. Thirty-eight children ages 8-16 with /r/ errors participated in a pretreatment speech perception assessment task followed by treatment for /r/ errors. Half of the children received just UVF treatment, and the other half received UVF + auditory perceptual training. Prior to receiving treatment, the researchers assessed speech perception with an identification task comprised of a 10-step
continuum from /r/ to /w/ using the stimuli “rake” to “wake.” Findings showed both the UVF and UVF + auditory perceptual training made equal gains in improving /r/ production (Preston et al., 2020). However, findings also showed that children who performed better on the pretreatment speech perception task made more overall gains in /r/ productions than those with poorer auditory perception, regardless of the type of treatment they received. Thus, in this study children with relatively poorer speech perception abilities made less progress in treatment than their peers with relatively stronger speech perception. This study suggests that not all children with SSD exhibit comparable speech perception skills, and speech perception ability may be linked to treatment outcomes.

**Speech Perception Abilities in Children With Speech Sound Disorders**

Several studies have investigated speech perception skills in children with SSD and the results are varied. Lof and Synan (1997) conducted a systematic review to investigate the relationship between speech perception and speech-sound production errors in children with SSD. This review included 34 studies from 1931-1994. Twenty-one studies found that children with SSD performed more poorly on speech perception than typically developing children, whereas 13 studies found that children with SSD did not have poor speech-sound discrimination. The authors found several factors that could account for the variety in results. First, the articles had a wide range of subject ages ranging from preschool to 24 years old. Research evidence suggests that speech perception continues to develop into adolescence (Hazan & Barrett, 2000), so it is unsurprising that preschoolers may perform differently than adolescents. Secondly, the different studies had many different speech perception testing methods (e.g., discrimination versus identification), subjects of different ages (e.g., preschool to early adulthood) and speech abilities (e.g., one to two speech errors versus multiple errors). Third, some studies only assessed
perception of a child’s errored sounds and others did not assess the children’s errored sounds. It is possible that different tasks and different stimulus types may yield disparate results between studies (Locke, 1980). Finally, many studies did not distinguish the difference in speech perception when listening to another speaker’s speech and listening to the subject’s own speech productions.

More recently, Hearnshaw et al. (2019) conducted a systematic review and meta-analysis to describe speech perception skills in young children (ages 3;0 - 6;11) with SSD. This comprehensive review initially investigated 15,423 articles published between 1913 and 2016. Ultimately, 71 articles (including 73 studies) fit the criteria of the systematic review. Criteria for inclusion involved studies researching children with SSD and speech perception abilities, as well as the participants being 3;0 to 6;11 years old. Results revealed that 60 out of the 73 studies concluded that the majority of young children with SSD have difficulty with speech perception. Two articles concluded that children with SSD only had speech perception difficulties if they had a concomitant language impairment. One study found that young children with SSD only have speech perception deficits if the child also has phonological processing difficulties. Four of the studies found that children with SSD have difficulty with speech perception when listening to their own productions but not others with typical speech. Twelve articles, however, found that children with SSD do not have difficulty with speech perception.

While the previous two studies reflected the variety of speech perception research, one factor that neither discussed in depth was the impact of the type of stimuli. Cabbage and Hitchcock (in press) conducted a systematic review investigating speech perception skills in children with residual speech sound disorders (RSSDs; errors persisting beyond 8 years of age) from the years 1990-2020. In total, 11 studies met the study criteria and findings showed that the
majority of papers concluded that children with RSSD tend to have poorer speech perception abilities than their typically developing peers. However, similar to Hearnshaw et al. (2019) and Lof and Synan (1997), Cabbage and Hitchcock (in press) found that the nature of perceptual skills was not consistent across children, some studies showing that children with SSD did not appear to have deficits while others did. Most studies researched the child’s errored sound and found their perception to be poorer for these sounds. A specific variable that Cabbage and Hitchcock (in press) investigated that was not discussed in the other two meta-analyses (Hearnshaw et al., 2019; Lof & Synan, 1997) was the type of stimuli used. The 11 studies used three types of stimuli: synthetic speech, natural speech, and/or synthetically altered natural speech. The studies that found that children with RSSD had speech perception difficulties consistently used synthetic speech or synthetically altered natural speech whereas studies using natural speech were less likely to find perceptual deficits in children with SSD. These findings suggest the importance of the stimuli type used to investigate speech perception in children with SSD.

The findings of these meta-analyses are similar in that they all found methodological variability that may contribute to disparate speech perception results across studies. The varied results could be due to the wide variety of testing methodologies. In the current study, we investigate methodological differences related to stimulus type (synthetic speech versus natural speech) and in a subcategory of natural speech, whether the participant is listening to their own voice versus another speaker. Specifically, we examine the effect of stimulus type in a sample of neurotypical adults with no history of speech sound disorder to establish a baseline understanding of how typical listeners perceive varied stimulus types.
Factors Affecting Speech Perception Performance

Type of Stimuli: Synthetic Versus Natural

The use of synthetic speech stimuli, speech that is artificially created, in speech perception research has been historically common. Researchers often use synthetic speech because different aspects of the speech signal, including individual acoustic variables, can be precisely controlled. While different aspects of the speech can be controlled, the resulting stimulus token often sounds unnatural, thus limiting the potential ecological validity of its use (Pisoni, 1997). Natural speech may be more ecologically valid because listeners encounter natural speech in everyday situations and are thus more used to hearing and reacting to natural speech. However, acoustic variables within a natural speech signal are much more difficult to precisely control. Another type of speech stimulus is synthetically altered natural speech, which is created by editing one or more acoustic parameters of a natural speech token using digital software tools (e.g., PRAAT, Adobe Audition). Some of these acoustic parameters include timing, fundamental frequency, spectral shape, amplitude pattern, formants, phonetic transplantation (taking a recording of one phoneme and replacing it with another), etc. (Kawahara & Morise, 2011; Shuster, 1998; Strömbergsson et al., 2014). Researchers benefit from this type of stimulus because it sounds like natural speech but can be controlled for experiments. Previous research has shown that children with SSD have more difficulty perceiving error sounds when using synthetic speech (Hitchcock et al., 2020; Rvachew & Jamieson, 1989) but may perform better when listening to natural speech (Preston et al., 2015). Due to synthetic speech allowing the researcher to precisely control specific acoustic variables, the researchers are able to identify what specific variables are most important for perception (Cabbage & Hitchcock, in press) but listeners rarely encounter synthetic speech in everyday life.
Moreover, few studies have investigated differences in perception of synthetic versus natural speech in children with SSD. In this study, we aim to present listeners with both synthetic speech and natural speech to better understand perceptual differences that may exist between different types of stimuli.

**Speech Errors vs Sounds With No Errors**

To be successful in therapy, children with SSD need to be able to distinguish their errored sounds from non-errors. The ability to do this will potentially help the child be able to understand and monitor his/her productions in real time and then adapt the production if needed. Several studies have found that children with SSD have more speech perception difficulty with tasks that involve their error sounds versus non-error sounds (Berti et al., 2020; Hitchcock et al., 2020; Lapko & Bankson, 1975; Roepke & Brosseau-Lapré, 2019; Rvachew & Jamieson, 1989; Shuster, 1998). Specific phonemes that have proven difficult for children with SSD to perceive include /s/ (Lapko & Bankson, 1975; Roepke & Brosseau-Lapré, 2019; Rvachew & Jamieson, 1989), /ʃ/ (Roepke & Brosseau-Lapré, 2019; Rvachew & Jamieson, 1989), and /r/ (Shuster, 1998). Several studies have demonstrated that children with sound errors have poor perception when tested with natural speech (Lapko & Bankson, 1975), synthetic speech (Hitchcock et al., 2020; Rvachew & Jamieson, 1989), as well as natural speech from children who also have a SSD (Roepke & Brosseau-Lapré, 2019). Berti et al. (2020) found a correlation between a child’s speech production errors and perception of those same sounds, but only for fricative errors. Other researchers found a correlation between a child’s speech production errors and poor perception when tested with natural speech, but only when listening to their own productions (Hoffman et al., 1983; Shuster, 1998).
Many children with SSD show speech perception deficits for the sound(s) that they are unable to produce. This is relevant for treatment, due to the high need for feedback and self-monitoring that is involved in traditional articulation therapy. If the child has poor speech perception for their errored sound(s), therapy may progress more slowly than for those children without perceptual deficits (Preston et al., 2020). In this study, we aim to present listeners with a commonly errored phoneme for children with SSD speaking Standard American English: /r/. Although listeners in the current study produce /r/ correctly, in the future this work will be extended to children with SSD who specifically do not produce /r/ correctly.

*Interpersonal vs. Intrapersonal*

Relevant to the current project, we investigated whether a listener’s perception abilities differ when they listen to recordings of other people’s speech versus their own speech. These different types of perception will further be referred to as interpersonal perception, when a person is listening to speech tokens produced by another speaker, and intrapersonal perception, when a person is listening to their own speech tokens. This has clinical relevance as during therapy children listen to the clinician’s correct production (interpersonal perception) produce the modeled production, and ultimately monitor whether their own production is correct or incorrect (intrapersonal perception). Previous research has shown that children with SSD have more difficulty perceiving inaccuracies in their own speech as compared to others (Lapko & Bankson, 1975; Roepke & Brosseau-Lapré, 2019; Wolfe & Irwin, 1973). For some children, this is only a problem when detecting their own errors in rhotics (Hoffman et al., 1983; Shuster, 1998) or fricatives (Berti et al., 2020).

Investigation of how children with SSD perceive others’ speech versus their own has been of interest to researchers for many decades. Aungst and Frick (1964) investigated whether
children’s consistency of misarticulated /r/ was related to being able to determine if their own productions were correct or incorrect. The children, ages 8;0-10;3, judged their own speech production accuracy when compared to another speaker, when compared to a recording of their own voice, and while speaking. Findings showed that (a) judging another person’s speech as correct or incorrect was not correlated with being able to judge one’s own speech accuracy, (b) judging another person’s speech was not correlated with consistency of articulation, and (c) being able to judge one’s own speech accuracy was highly related to the accuracy of articulation.

While it is understood that children have better perception when listening to natural speech when compared to synthetic speech (Hitchcock et al., 2020; Rvachew & Jamieson, 1989), recent methodologies allow the synthetic modification of natural speech (Shuster, 1998; Strömbergsson et al., 2014). Using synthetically altered natural speech allows the researcher to control certain parameters while maintaining the naturalness of the speech signal. Shuster (1998) was the first to present children with SSD with synthetically altered natural speech. She investigated whether children ages 7;1-13;11 with SSD, who were unable to produce /r/ and had only one other sound error at most, could distinguish between errors in their own speech and another person's speech. In addition, she investigated each child’s ability to judge the accuracy of each production. She utilized an innovative method that involved digitally correcting a child’s error through a linear predictive coding parameter manipulation/synthesis to determine whether children could detect productions that were incorrect or “corrected.” The children listened to 200 words (50 corrected words produced by the child, 50 incorrect words produced by the child, 50 corrected words produced by another child and 50 incorrect words produced by another child) and were asked to tell the examiner if the word was his/her own speech, as well as if the /r/ was
correct or incorrect. Findings revealed that most of the subjects could accurately judge their own corrected utterances and subjects performed at chance level for judging their own /r/ incorrect productions. The children were better able to judge the identity of the speaker (self vs other) when it was an incorrect production. It was also easier for the children to identify the speaker when it was another child rather than their own speech.

Strömbergsson et al. (2014) investigated how children with and without SSD react to hearing their own synthetically altered natural speech. The children, ages 4-6 years old, were recorded producing selected target words containing errored phonemes of velar fronting (e.g., producing /t/ for /k/, /d/ for /g/ and /n/ for /ŋ/). Immediately after the child was recorded saying the word, they were asked if the pronunciation was correct or incorrect. Some of the words were then synthetically altered through phonetic transplantation, where the researchers replaced a recording of one phoneme with another. After a delay, the children listened to a recording of a word (the child’s original production or the synthetically altered) and were again asked if the pronunciation was correct or incorrect. Strömbergsson et al. (2014) found that children with SSD identified their own incorrect utterances as correct when immediately played back, but when the playback was delayed, the children with SSD perceived their incorrect utterances as incorrect. These findings show that children with SSD could have difficulty with self-monitoring during treatment sessions, but with a delay may accurately identify their errored productions.

To understand if there is a correlation between perception of children’s own productions compared to perception of adult productions, Berti et al. (2021) studied ten children, ages 4-6, diagnosed with SSD who had no prior speech therapy. Children were shown two figures from a minimal pair and then played a recording from either a typical adult speaker, or the child’s own productions of one of the words shown. After hearing the recording, the children touched which
figure matched the spoken word. The researchers concluded that the children made more errors on the perception task when they listened to their own productions (66% errors) than when they listened to the typical adult productions (30% errors). Therefore, it appears that children with SSD have more difficulty correctly perceiving their own speech productions than adult productions.

Taken together, these studies highlight the importance of considering intrapersonal (using synthetically altered natural speech) and interpersonal speech perception in children. Additional research is needed to determine whether there are meaningful differences in how children perceive other speakers (interpersonal) vs their own speech productions (intrapersonal) using synthetically altered natural speech.

**Purpose of Current Study**

Hitchcock et al. (2020) researched interpersonal speech perception in children with SSD, typically developing children, and adults by presenting listeners with synthetic speech in a forced-choice discrimination task using a computer program, the *Wide Range Acoustic Accuracy Scale* (WRAAS). The WRAAS has thus far been used to examine discrimination of synthetic speech syllables contrasted by phonemes. The current study will investigate the effect of three different speech types: synthetic speech, synthetically altered natural speech of a standard adult speaker (interpersonal perception), and synthetically altered natural speech of each participant’s own voice (intrapersonal perception) using the WRAAS task in adult listeners. This study will determine the effects of stimulus type on speech perception for listeners without speech production deficits. Previous studies (Garner, 2021; Hitchcock et al., 2020) performed the same synthetic WRAAS task with adults and children with and without SSD. Both studies concluded that the typically developing children performed similarly to adults for all phoneme contrasts.
Due to this study experimenting with interpersonal and intrapersonal synthetically altered natural speech, conducting this investigation in adults will provide important knowledge regarding typical performance in children prior to comparison to children with SSD. The results from this study will help establish baseline expectations for how typical listeners perceive speech of different types. Thus, this study will extend to a larger body of work that will investigate the perception of these different speech stimuli in children with and without SSD. We ask the following research questions:

1. Is there a difference between synthetic speech and natural speech perception in a discrimination task for adult listeners?

2. Is there a difference in interpersonal versus intrapersonal perception using synthetically altered natural speech in a discrimination task for adult listeners?

Method

The Institutional Review Board at Brigham Young University granted approval for the recruitment of human subjects and the execution of this study. The researcher reviewed procedures, risks, and benefits of participating in this study and participants provided their consent.

Participants

A total of 25 females aged 18 years and older participated in this study. We recruited female participants to resemble the acoustic characteristics of child speakers more closely (Robb & Smith, 2002). Moreover, we selected a female standard speaker because a large majority (~95%) of SLPs are female (ASHA, 2020). Thus, children are more likely to encounter female SLPs that will serve as the verbal model for targeted speech sounds in therapy. We conducted a power analysis using previously reported effect sizes (Hitchcock et al., 2020) to determine that a
sample size of 25 participants would be sufficient to detect any differences between stimulus
types in our study. All adults were native English speakers and had no history of any
neurological or cognitive impairment (e.g., autism, Down syndrome) as well as no history of
speech and language concerns, per self-reported case history. Participants were recruited through
personal contact, social media, and flyers/announcements.

Descriptive Measures

Hearing

All participants underwent a hearing screening to ensure hearing within normal limits.
Participants passed a hearing screening threshold at 20dB at 1000, 2000, and 4000 Hz, consistent
with current screening practices.

Speech Production

To determine that the participant did not have an SSD, all participants read The
Grandfather Passage and The Rainbow Passage. These passages are phonetically balanced to
represent the phonemes spoken in English. The participants also recorded the speech tokens that
were used in the intrapersonal speech perception task. Participants were seated in front of a
RODE NT1 desktop microphone where they read the target words (e.g., bot, watt, dot, got, rot,
watt) in random order from a PowerPoint presentation. The participants completed the task three
times to ensure an adequate quality recording was obtained.

Stimuli

The participants listened to three different types of stimuli: synthetic speech, synthetically
altered natural speech from a standard adult speaker, and synthetically altered natural speech of
each participant’s own voice recordings.
**Synthetic Speech**

The synthetic speech stimuli were created using a Klatt synthesizer (Klatt, 1980) and have been used previously (Cabbage, 2013; Garner, 2021; Hitchcock et al., 2020). Spectral representations for each syllable pair are shown in Figure 1. The first presented stimulus pair was /ba/-/wa/. This pair contains 81 steps along its continuum, each sound differing in the transition duration from 25 msec to 105 msec in one msec steps. The second presented pair was /da/-/ga/. This pair differed in the F3 onset frequency, ranging from 1800 Hz to 2700 Hz, differing in 20 Hz steps along the continuum. This pair contained 46 differing stimuli. The final stimulus pair that was presented was /ra/-/wa/. This pair differed in the F3-F2 distance, ranging from 1500 Hz to 2500 Hz in 25 Hz steps. There were 41 steps along the continuum for this pair. Aside from the acoustic parameters mentioned above, all other acoustic measurements remained the same between each syllable pair. Each continuum was pilot tested in order to avoid floor and ceiling effects.
Synthetically Altered Natural Speech

Synthetically altered natural speech stimuli were created from recordings of a standard adult female speaker and each participant’s own voice repeating each of the words in the word pairs “bot”-“watt”, “dot”-“got”, and “rot”-“watt” to mirror the phoneme contrasts used in the
synthetic speech task (/b/ - /w/, /d/ - /g/, /r/ - /w/). See Table 1 for a description of the word pairs.

**Table 1**

*Description of Word Pairs*

<table>
<thead>
<tr>
<th>Stimuli word-pair</th>
<th>Transcription</th>
<th>Phoneme contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>“rot” – “watt”</td>
<td>/rat/ – /wat/</td>
<td>rhotic-glide</td>
</tr>
<tr>
<td>“dot” – “got”</td>
<td>/dat/ – /gat/</td>
<td>alveolar stop-velar stop</td>
</tr>
<tr>
<td>“bot” – “watt”</td>
<td>/bat/ – /wat/</td>
<td>bilabial stop-bilabial glide</td>
</tr>
</tbody>
</table>

All recorded words were selected and preprocessed using acoustic software (e.g., Audacity) to ensure words in each word pair were normalized in amplitude and duration. They were further analyzed using acoustic software (e.g., Wavesurfer, PRAAT) and visual inspection to trace the formant frequencies for the words in each word pair. Trained research assistants identified three to five timepoints associated with formant transitions in each word pair. Thus, each word pair had its own unique time points designated. The actual synthetic alteration was accomplished by the TANDEM-STRAIGHT vocoding Matlab script (Kawahara & Morise, 2011) which systematically interpolates speech parameters between the word pairs to create the continua. Each word pair was individually processed for every speaker. For the creation of the synthetically altered speech continua, trained research assistants imported both words in a word pair (e.g., ‘bot’ and ‘watt’) into Matlab for the synthetic alteration. For the TANDEM-STRAIGHT program to create the continuum, the research assistants manually marked the key timepoints identified in the visual inspection. TANDEM-STRAIGHT then automatically interpolated the iterations of the formant frequencies that alter each stimulus in the continuum.
from one endpoint in the word pair (e.g., ‘bot’) to the other (e.g., ‘watt’). TANDEM-STRAIGHT allows the user to designate the number of steps in the continuum. Based on pilot data, we standardized the continua length across all word pairs to 50 stimulus steps between the endpoints of each word pair (i.e., ‘bot’-‘watt’, ‘dot’-‘got’, ‘rot’-‘watt’). The speech parameters that were allowed to vary naturally within TANDEM-STRAIGHT were timing, amplitude, fundamental frequency, source spectra, and formant trajectories. The resulting stimuli retained the naturalness, identity, and other suprasegmental characteristics of the original speaker (Matsui & Kawahara, 2003).

**Experimental Measures**

Speech perception was measured using the *Wide Range Acoustic Accuracy Scale* (WRAAS) program (Hitchcock et al., 2020). WRAAS is a computer program used to measure a listener’s ability to discriminate sounds that differ. Participants completed three categories of speech perception tasks: (a) synthetic speech, (b) synthetically altered natural speech of a standard adult speaker (interpersonal); and (c) each participant's own synthetically altered natural speech (intrapersonal). The syllables used for the synthetic speech stimuli (/bɑ/-/wɑ/, /dɑ/-/gɑ/, /rɑ/-/wɑ/) contain the same initial phonemes as the real word pairs utilized for the synthetically altered natural speech stimuli (‘bot’-‘watt’, ‘dot’-‘got’, ‘rot’-‘watt’). The /b/ - /w/ contrast was selected as the control phoneme contrast as this is a very early developing phoneme contrast in both speech perception and speech production (Nittroer et al., 2013). The /d/ - /g/ contrast was selected due to extensive previous work which has shown discrimination difficulty for several clinical populations such as developmental language disorder (Tallal, 1980), dyslexia (Cabbage et al., 2016), and learning disabilities (Kraus et al., 1996). Lastly, the /r/ - /w/ contrast was selected because future work for which this study provides pilot data involves children who do
not produce the /r/ sound correctly and thus, it is predicted children with SSD will have a specific
deficit for this comparison relative to their typically developing peers.

Participants were seated at the computer and fitted with Sennheiser HD280 Pro closed
ear, circumaural headphones for audio delivery. The volume was set to the same standard
computer volume level for all participants and then adjusted to the listener's comfortable
listening level based on their input. The participants responded via either a handheld 2-button
response box in his/her hands (akin to a video game controller) or two designated keys on the
computer to respond to each trial (partway through the data collection process, the 2-button
response box failed, and all subsequent participants used the keyboard). Participants first
completed a practice WRAAS task using syllables /ba/ and /pa/. The /ba-pa/ practice syllable
pair was created with only 8 steps in the continuum to facilitate ease of training of the WRAAS
task. Participants completed the WRAAS task 3 times for each stimulus type (synthetic,
synthetically altered standard adult, synthetically altered own speech), once for each phoneme
contrast (/b/ - /w/, /d/ - /g/, /r/ - /w/). All participants completed the synthetic speech task during
session 1. Presentation of the synthetically altered standard adult speaker and their own
synthetically altered voice was counterbalanced across participants. During the WRAAS task,
participants heard two pairs of stimuli from a selected continuum at a time, presented in a 4IAX
paradigm for each trial. One pair in the trial contained a standard stimulus item presented twice
(e.g., ‘bot’-‘bot’); the other pair was the standard stimulus item paired with a comparison
stimulus item from somewhere along the 50-step continuum of the designated phoneme contrast
pair. The participant was asked to select the pair that was different by selecting the
 corresponding button. The WRAAS algorithm randomly selected which pair, same or different,
was presented first. When deciding which comparison stimulus item to present, the WRAAS
algorithm is designed to first select the most distinct stimulus item (e.g., stimulus 1 is the
standard [e.g., ‘bot’] and stimulus 50 represents the opposite end of the continuum [e.g., ‘watt’]).
If the participant accurately selected the pair containing the comparison stimulus item, the
WRAAS algorithm halved the distance along the continuum (e.g., the next trial contained
stimulus item 1 and comparison stimulus item 25). Subsequent trials were based on the
participants previous correct/incorrect response.

Through an iterative process, presentation of stimuli with the WRAAS algorithm
continued until the participant reached a 71% accuracy response rate. This is known as the just-
oticeable difference (JND) between stimuli, consistent with signal detection theory (Green &
Swets, 1966). The program stopped at this point yielding a convergence level (CL) which is
defined as the distance between the standard stimulus (e.g., stimulus 1) and the comparison
stimulus along the continuum that occurs when the participant is 71% accurate. See Figure 2 for
a demonstration of trial responses for a pilot participant.
Procedure(s)

The study took place in laboratory rooms with sound booths at the BYU John Taylor Building at a time convenient for the participant. Participants completed two research sessions. Session 1 took approximately 30 minutes. Session 2 took approximately 30 minutes.

Session 1

Session 1 began with a hearing screening to ensure hearing was within normal limits. Participants read *The Grandfather Passage* and *The Rainbow Passage* to ensure normal speech production skills. At this point, participants viewed a PowerPoint slideshow with target words (refer to Table 1). Participants were seated at a table with a microphone and recordings were made at a sampling rate of 44.1 kHz and an amplitude resolution of 16 bits, using a desktop
microphone (RODE NT1) and a Zoom H4n digital recorder. Participants produced each target word at least 3 times to ensure an adequate recording was obtained.

Next, participants completed a practice speech perception task. They were fitted with headphones attached to a computer for the speech perception task. Participants held a 2-button response box or used the keyboard to indicate responses for all speech perception tasks. The participant first completed the practice WRAAS task with a synthetic /ba/ - /pa/ practice syllable pair. All participants successfully completed the practice task. Participants then completed the synthetic speech perception task. They listened to syllable pairs /ba/ - /wa/, /da/ - /ga/, and /ra/ - /wa/ as described above.

Between research sessions, a research team prepared and processed the recorded speech tokens for play back in session 2. All recordings were digitally transferred to a computer and segmented into individual words using Audacity (Audacity Team, 2021). All words were screened for mispronunciations, peak clipping, and background noise and normalized at -.5 dB (re: 16 bits = 96 dB peak). The remaining words were ranked by at least 3 listeners in order of best quality. Following this process, the top-ranked token was selected for inclusion in the study. The paired recordings (‘bot’ – ‘watt’, ‘dot’ – ‘got’, ‘rot’ – ‘watt’) were digitally processed to be the same duration. The word pairs were then processed using a customized MATLAB script (TANDEM STRAIGHT; Kawahara et al., 2009) that digitally interpolates a continuum of 50 stimulus items that systematically change from one of the words in the pair to the other, as described above. Time between session 1 and 2 varied from 10-18 weeks.

**Session 2**

At the beginning of session 2, participants completed another hearing screening to ensure they had not developed a hearing loss in between sessions (e.g., developed allergies or a cold).
After the hearing screening, the participants completed the same practice speech perception task described above to refamiliarize the participant with the WRAAS program.

Next, the participants completed the synthetically altered natural speech perception tasks. They were presented with 6 WRAAS tasks using synthetically altered natural speech for each phoneme contrast (‘bott’ – ‘watt’, ‘dot’ – ‘got’, and ‘rot’ – ‘watt’) from a standard speaker and the participant’s own recorded speech. The order of presentation of the syllable pairs was counterbalanced across all participants. Participants were again fitted with headphones attached to a computer for the speech perception tasks. At the end of session 2, participants received $10 for participating.

Results

In this study we compared speech perception in a series of discrimination tasks across three stimulus types (synthetic speech, interpersonal synthetically altered natural speech, intrapersonal synthetically altered natural speech) and three phoneme contrasts (/b/ - /w/, /d/ - /g/, /t/ - /w/). Descriptive statistics for the different phoneme contrasts across each stimulus condition is presented in Table 2.
Table 2

Summary Statistics Grouped by Phoneme Contrast Pair and Stimulus Type

<table>
<thead>
<tr>
<th></th>
<th>/b/-/w/</th>
<th>/d/-/g/</th>
<th>/r/-/w/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synth</td>
<td>Inter</td>
<td>Intra</td>
</tr>
<tr>
<td>Mean</td>
<td>6.28</td>
<td>8.16</td>
<td>6.48</td>
</tr>
<tr>
<td>SD</td>
<td>3.47</td>
<td>4.53</td>
<td>5.31</td>
</tr>
<tr>
<td>Median</td>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Range</td>
<td>13</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>15</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>Mode</td>
<td>5*</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

Note. * = multiple modes exist. The smallest value is shown. Mean = convergence level.

The /b/-/w/ Phonemic Contrast

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in discrimination of the /b/-/w/ phoneme contrast for different types of speech stimuli: synthetic speech, interpersonal synthetically altered natural speech, and intrapersonal synthetically altered natural speech. There were no outliers, and the data were normally distributed for each speech stimulus condition, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(2) = .290, p = .865$. The listeners did not differ in perception between speech stimulus type $F(2, 24) = 1.765, p = .182$, partial $\eta^2 = .069$, with average convergence levels for perception of synthetic speech ($M = 6.28, SD = 3.470$), interpersonal synthetically altered natural speech ($M = 8.16, SD = 4.534$), and intrapersonal synthetically altered natural speech ($M = 6.48, SD = 5.308$), as seen in Figure 3.
Figure 3

/b/-/w/ Convergence Level for Each Stimulus Type

Note. The middle line of the box represents the median, and the x in the box represents the mean.

The /d/-/g/ Phonemic Contrast

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in discrimination of the /d/ and /g/ phoneme contrast across the different types of speech stimuli. There were no outliers, and the data were normally distributed for each speech stimulus condition, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(2) = 3.288, p = .193$. As seen in Figure 4, listeners differed in their speech perception according to speech stimulus type, $F(2,24) = 5.124, p = .010$, partial $\eta^2 = 0.176$. Post hoc analysis with a Bonferroni adjustment revealed that listeners had statistically poorer discrimination of the /d/-/g/ phoneme contrast for synthetic speech ($M = 13.360, 95\% \text{ CI} [8.737, 17.983]$) as compared to intrapersonal synthetically altered natural speech ($M = 5.560, 95\% \text{ CI} [2.421, 8.699], p = 0.027$), but discrimination of this contrast in synthetic speech did not differ from
interpersonal synthetically altered natural speech ($M = 8.520$, 95% CI [6.151, 10.889], $p = 0.222$). Post hoc analysis also revealed no differences between /d/ - /g/ discrimination for interpersonal and intrapersonal synthetically altered natural speech ($p > .05$).

**Figure 4**

/d/-/g/ Convergence Level for Each Stimulus Type

Note. The middle line of the box represents the median, and the x in the box represents the mean.

**The /r/-/w/ Phonemic Contrast**

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in discrimination for the /r/ and /w/ phoneme contrast for the different types of speech stimuli. There were no outliers, and the data were normally distributed for each speech stimulus condition, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(2) = 1.239, p = .538$. The listeners did not differ in discrimination for this contrast across speech stimulus type $F(2, 24) = 1.168, p = .320$, partial $\eta^2 = .046$, with average convergence levels for perception of synthetic speech ($M = 11.76$, $SD = 8.536$), interpersonal synthetically
altered natural speech ($M = 9.88, SD = 6.287$), and intrapersonal synthetically altered natural speech ($M = 9.04, SD = 6.127$), as seen in Figure 5.

**Figure 5**

/r/-/w/ Convergence Level for Each Stimulus Type

![/r/ vs /w/](image)

Note. The middle line of the box represents the median, and the x in the box represents the mean.

Overall, there was no statistical difference between stimulus type (synthetic versus synthetically altered natural speech) or interpersonal versus intrapersonal speech perception, except for discrimination of the /d/-/g/ phoneme contrast in synthetic speech versus intrapersonal synthetically altered natural speech.

**Discussion**

The purpose of this paper was to investigate the potential differences in discrimination of three phoneme contrasts (/b/-/w/, /d/-/g/, /r/-/w/) across three stimulus types (synthetic, interpersonal synthetically altered natural speech, intrapersonal synthetically altered natural speech). Understanding how adult listeners differ in discrimination of these speech stimulus types will help establish a baseline understanding of phoneme discrimination across varied stimulus types for comparison to children with and without SSD. We aimed to determine: (a) if
there was a difference between synthetic speech and natural speech perception in a discrimination task for adult listeners, and (b) if there was a difference in interpersonal versus intrapersonal perception of synthetically altered natural speech in a discrimination task for adult listeners.

**For Phoneme Contrasts in Most Listening Conditions, Listeners Exhibited Equivalent Speech Discrimination**

On average, discrimination did not differ across stimulus type for any phoneme contrast for the adult listeners in this study. We did, however, find a difference in speech discrimination for the phoneme pair /d/ vs /g/ between synthetic speech and the intrapersonal synthetically altered natural speech condition. Adult participants had on average, poorer perception for synthetic /da/ - /ga/ than they did when discriminating their own productions of ‘dot’ – ‘got.’

Even though the adults in the current study differed between synthetic and intrapersonal synthetically altered natural speech for the /d/-/g/ phoneme contrast, our findings for synthetic speech discrimination align with the findings of other studies. First, the /d/-/g/ contrast appears to be particularly difficult for listeners when presented in synthetic speech. Kraus et al. (1996) researched the perception abilities for synthetic phoneme pairs /ba/-/wa/ and /da/-/ga/ between typical children and children with learning disabilities and/or ADD. The researchers concluded that (a) the typically developing children performed better than the children with learning disabilities and/or ADD for both phoneme contrasts, and (b) the /da/-/ga/ phoneme pair was more difficult for all groups to discriminate than the /ba/-/wa/ pair, regardless of disability status. Hitchcock et al. (2020) investigated the difference in speech perception in neurotypical adults and children (ages 7-13) with and without SSD using synthetic phoneme pairs. Participants completed a syllable (/ba/-/wa/, /da/-/ga/, /ra/-/wa/) discrimination task through the
WRAAS computer program. The adult group mean convergence level for /da/-/ga/ was 11.96, which is numerically similar to the mean convergence level for /da/-/ga/ in the current study which was 13.36. Additionally, Garner (2021) also investigated the difference in speech perception between neurotypical adults and children with and without SSD when listening to the same synthetic syllable pairs (/ba/-/wa/, /da/-/ga/, /ra/-/wa/). Garner (2021) found the adult group mean convergence level when listening to /da/-/ga/ to be 11.68, as compared to 13.36 in the current study. While it appears that /d/-/g/ synthetic speech is more difficult than intrapersonal synthetically altered natural speech, this is not the case for other phoneme contrasts. Further research is needed to better understand this discrepancy.

**No Difference Between Interpersonal and Intrapersonal Perception for Synthetically Altered Natural Speech**

We found no statistical difference between interpersonal and intrapersonal synthetically altered natural speech for any phoneme contrast. It is possible that due to the delayed presentation of the participant’s own voice recording, the participants may not have perceived the recordings as intrapersonal perception. Importantly, we note there was an unexpected extended time delay between sessions while the research team processed the participant recordings. This delay occurred as the result of a break between academic semesters as well as troubleshooting the TANDEM STRAIGHT Matlab script which was new to the research team. This may have further contributed to the participants failing to process the stimuli as truly intrapersonal. Indeed, at the end of the second session we anecdotally asked participants whether they recognized any of the speakers in the synthetically altered natural speech conditions and the majority (n = 19) failed to recognize that they were listening to their own speech in any of the tasks.
The distinction between real time perception and delayed perception is important. Wolfe and Irwin (1973) compared children’s (grades 1-6) perceptions of their own productions with the correct or incorrect productions of another speaker. They sought to determine whether children who consistently misarticulate /r/ could accurately compare their own productions in real-time to another speaker versus accurately comparing their own productions to another speaker from a recording. The children said a word, listened to another speaker say the word and were asked to tell if the sounds were the same or not. Two weeks later, they listened to the recordings of their own productions as well as another speaker and were asked to tell if the sounds were the same or not. Findings showed that children were able to distinguish their misarticulations from the correct productions of another speaker more accurately during the task that was recorded than the task that was conducted in real-time.

Strömborgsson et al. (2014) investigated children, ages 4-6, with and without SSD to understand how children perceive their own voice. They asked the children to say the stimulus word and were immediately asked if their production was correct or incorrect. After completing several other tasks, the children relistened to their productions through a recording and were again asked if the production was correct or incorrect. Strömborgsson et al. concluded that when immediately following their real-time production, children reported their productions to be correct, but after the delay the children reported their productions as incorrect.

**Limitations**

We note several limitations of the current study. First, all participating subjects were female. While this is a limitation, this decision was intentional due to future studies only testing young children since, on average, children have similar fundamental frequencies as adult females (Robb & Smith, 2002). Moreover, we selected a female speaker as the standard speaker as a
large majority (~95%) of SLPs are female (ASHA, 2020). Thus, children are more likely to encounter female SLPs that will serve as the verbal model for targeted speech sounds in therapy. Future work should, however, include males as both standard speakers and participants to better represent the general population.

Another limitation of this study was the time delay between session 1 and session 2. It was expected to be about 2-4 weeks and ended up ranging from 10-18 weeks. This was due to unforeseen complications resulting from the project team learning how to use and troubleshoot the TANDEM-STRAGHT program in Matlab. Additionally, this time frame included a substantial break between academic semesters (~4 weeks) when the research team did not process any stimuli. Future studies should not expect as long of a time delay now that these issues have been resolved. This delay may have impacted the participant experience that may be different for participants with a shorter time delay between sessions. This will be explored in future work.

A final limitation concerns the sample of participants in the current study. Several participants (n = 9) were graduate students in speech-language pathology, thus they had extensive training in speech and hearing, whereas the remaining subjects (n = 16) were undergraduate college students with little or no background knowledge in speech and hearing. To address this limitation, we ran the analyses by group and found the results did not differ between groups.

Implications for Future Research

In the future, we will look at differences in perception according to stimulus type and interpersonal versus intrapersonal perception of synthetically altered natural speech in children with and without SSD. We will also look at the difference in real-time perception versus delayed
perception of a child’s own voice. In the current study, it is possible that participants did not register their own recordings as their own voice because of the significant delay between recording and the perceptual task. When asked after the second session whether they recognized any of the speakers in the recordings, the majority reported no familiarity with the voices. The purpose of this study was to determine differences between interpersonal and intrapersonal perception but it is possible that we did not successfully tap into intrapersonal perceptual mechanisms because so many listeners were unaware that they were hearing their own voice recordings.

Conclusion

The primary aim of this paper was to explore the difference in perception for different types of stimuli (synthetic speech vs interpersonal synthetically altered natural speech vs intrapersonal synthetically altered natural speech) for adult listeners across three different phoneme contrasts (/b/-/w/, /d/-/g/, /r/-/w/). There was no statistically significant difference between synthetic speech, interpersonal synthetically altered natural speech and intrapersonal synthetically altered natural speech for any phoneme pair, except for the /d/-/g/ contrast between synthetic speech and intrapersonal synthetically altered natural speech. There was no difference between interpersonal perception and intrapersonal perception for any phoneme contrast in the synthetically altered natural speech conditions; however, this may be due to the unforeseen significant time delay between sessions and the subsequent impact on the listeners’ lack of awareness that they were hearing their own voice. Moving forward we will research speech perception in children with and without SSD to compare synthetic speech, interpersonal synthetically altered natural speech and intrapersonal synthetically altered natural speech. The findings from the current study and the future study with children may help SLPs tailor
intervention to individual needs of children with SSD. Such information may guide SLPs to
determine whether the child has perception difficulties which may guide therapy in a more
effective direction to target speech perception deficits in addition to articulation.
References


https://www.doi.org/10.1126/science.273.5277.971


https://doi.org/10.1044/cicsd_24_S_57


https://doi.org/10.3109/02699206.2014.926994


APPENDIX A

Review of the Literature


Objectives: The purpose of this study was to determine if children’s consistency of misarticulated /r/ is related to being able to determine if one’s own productions are correct or incorrect. The children judged their own speech productions (a) when compared to another speaker, (b) when compared to a recording of one's own voice, and (c) while speaking.

Methods: 27 children (16 boys and 11 girls) ages 8;0-10;3 who misarticulate /r/ participated in four tests. 30 test words were drawn on index cards and placed in front of the children. The children participated in different tasks comparing their own productions of these words to other speakers, their recordings, and their real time speech.

Results: This study came to three conclusions: a) judging another person’s speech as correct or incorrect is not related to being able to judge one’s own speech, b) judging another person’s speech is not related to consistency of articulation, c) being able to judge one’s own speech is highly related to the consistency of articulation.

Conclusions: The ability to determine one’s own speech as correct or incorrect is highly correlated to consistency or articulation, therefore tests looking at self-speech perception would be useful in diagnosis, therapy and research.

Relevance to the current study: This article is relevant to the current study because both studies researched the subjects’ ability to judge their own speech
productions as well as another speaker. Both studies had subjects listen to recordings of their own voice.


**Objectives:** The authors researched the relationship between speech perception and speech production. They wanted to find out if there would be a positive correlation between production and perception, a difference between production and perception, and if there is a correlation between production and perception.

**Methods:** Thirty-five children, ages 4;8-6;8, diagnosed with SSD participated in this project. The children were shown pictures of items and asked to name them, so the researchers could know what production errors the children made. For the perception task, the children listened to typical adult recordings of minimal pair words. They first completed the word recognition and training phase to familiarize the participants with the stimuli words. For the testing phase, the children heard a word from the adult recordings, and were asked to choose, between two pictures, which one correlated to the recording they heard.

**Results:** The researchers concluded that production and perception are correlated in that if a child has more production errors, they will have more perception errors as well. When the researchers looked at specific types of phonemes, they found a correlation between production and perception, but only for fricative sounds (not stops or sonorants). They also concluded that there was a difference between production and perception. The children’s production was lower than their perception abilities.
Conclusions: Production and perception are correlated, especially for fricatives. Production is worse than perception abilities.

Relevance to the current study: This study provides important information for speech sound productions and perception of different classifications of speech sounds.


Objectives: The purpose of the study was to find the relationship between speech perception and production through three hypotheses: A) They believed there would be a difference between the perception of typical speech and the child’s own atypical speech productions. B) The relationship between perception and production depends on what stimuli was given (typical speech vs their own atypical productions). C) The children would not have the same errors in their atypical productions as in the perception of typical and their own atypical speech.

Methods: Ten children, ages 4-6, diagnosed with SSD but had no prior speech therapy treatment, participated in this study. The researcher presented minimal pair figures to the children for the naming task to record the stimuli. For the perception task, the children were shown two figures from a minimal pair and then played a recording from either a typical adult speaker, or the child’s own productions of the one of the words shown. After hearing the recording, the children touched which figure went with the spoken word. When a child heard their own atypical production of a word, the researchers only marked their response correct if they pointed to the figure that the child’s recording was from, even if it sounds like the production for the other figure in
the minimal pair. For example, if a child had a production error replacing /t/ for /k/, in the recordings of the minimal pairs /ti/ and /ki/ their recordings would both sound like /ti/. When the errored recording was played back for the child, it was only marked correct if they pointed to the picture of the “key”.

Results: The researchers concluded that the children made more errors on the perception task when they listened to their own productions (66% errors) than when they listened to the typical adult productions (30% errors).

Conclusions: Children with SSD have more difficulty perceiving their own speech productions than adult typical productions.

Relevance to the current study: This study researched the difference between interpersonal and intrapersonal perception, similarly to the current study. The current study is also using adult typical productions and the comparison stimuli.


Objectives: The purpose of this study was to determine if children with dyslexia perceived sine-wave speech and amplitude-comodulated sine-wave speech better than children with speech delay. This study also questioned if children with persistent speech delay would have difficulty perceiving phonemes they couldn’t not produce compared to phonemes they could produce.

Methods: 36 children, ages 7;6 - 9;6, were categorized into three groups: typically developing, dyslexia and persistent speech delay. Children listened to single words of
both amplitude-comodulated sine-wave speech and sine-wave speech and repeated what they heard.

**Results:** All children, despite having dyslexia or persistent speech delay, performed better with the amplitude-comodulated speech rather than the sine-wave speech. Children with persistent speech delay had difficulty perceiving words beginning with their errored sound (/r/) with both the amplitude-comodulated speech and the sine-wave speech.

**Conclusions:** Children with specific speech delay have difficulty perceiving temporal fine structure acoustic characteristics. This study confirms that children with a speech sound disorder have lower speech perception for their errored sound regardless of the type of speech (amplitude-comodulated speech or sine-wave speech).

**Relevance to the current study:** Both studies researched speech perception with different types of speech/stimuli.


**Objectives:** This study wanted to find out the relationship between children who have SSD and having difficulties with speech perception. As a result, they created a systematic review and meta-analysis.

**Methods:** The authors searched 8 different databases and came up with a total of 15,423 articles, which was narrowed down to 71 articles, published between 1931 and 2016, that fit the criteria for the study. Studies needed to investigate children with SSD
and speech perception skills and have children between the ages of 3;0 and 6;11 with SSD. The authors created their own rating scale to report methodology of the studies. They used Comprehensive Meta-Analysis to create the meta-analysis for this study.

**Results:** The majority of studies reported that children with SSD had problems with speech perception as well. The meta-analysis reported that there is a significant difference between typically developing children and children with SSD in relation to lexical and/or phonetic judgment tasks.

**Conclusions:** Although research regarding speech perception has changed over the years, it can be concluded that most children with SSD will have difficulties with speech perception.

**Relevance to current study:** This study shows the variety of speech perception research and important factors that need to be taken into consideration. The variety of testing methodologies reflects the wide variety of results regarding speech perception research. This meta-analysis provides important information for this study about different testing parameters and overall results that most children with SSD also have difficulty with speech perception.


https://doi.org/10.1044/2020_PERSP-20-00037

**Objectives:** This study investigated the difference in speech perception in adults, typically developing children and children with SSD. They asked three questions: “1) How does perception of various syllable contrasts compare between TD children and adults? 2) Do
children with SSD differ in their perception of syllable contrasts when compared to TD children and adults? 3) Do children with SSD exhibit poorer perception relative to TD children and/or adults for syllable contrasts that include their speech production errors?”

Methods: The study consisted of 3 subgroups: 24 adults, 15 TD children, and 15 children with SSD. All of the children were 7-13 years old. All participants participated in a syllable discrimination task through the computer program WRAAS. The consonant-vowel syllable contrasts (/ba/-/wa/, /da/-/ga/, /ra/-/wa/) only differed by one acoustic parameter. Subjects were asked to distinguish which pair of syllables were different (ex: /ba/-/wa/ or /ba/-/ba/).

Results: Adults and TD children showed no statistically significant difference in discriminating syllable contrast. Adults and children with SSD significantly differed. TD children and children with SSD only differed on the /ra/-/wa/ contrast.

Conclusions: Children with an SSD are less accurate at distinguishing /ba/-/wa/, /da/-/ga/, /ra/-/wa/ when compared to adults and TD children and therefore have more difficulty with speech perception.

Relevance to the current study: This study relates to the current study because both research speech perception in typically developing adults. This study provided evidence for a contrast in performance of speech perception in children with a SSD, TD children and adults. Like this study, the current study had participants discriminate syllable contrasts that differed by one acoustic parameter. Another feature that is the same is that participants were asked to tell which syllable pair was different.

**Objectives:** The purpose of this study was to research three questions about the relationship between children’s perception and production of /r/. A) Do children who do not pronounce /r/ misperceive the /r/-/w/ contrast when produced by children who do not misarticulate? B) Can children who misarticulate /r/ accurately judge their own productions of the /r/-/w/ contrast to be correct or incorrect? C) Is there a relationship between perception of /r/-/w/ contrast and the child’s production of /r/?

**Methods:** Twelve children, ages 6;6 to 8;4, participated in an activity where adult’s recordings of stimulus words were presented, and the child was required to repeat the stimulus word while the examiner recorded what the child said. Two weeks later the children returned and responded to recordings of all subjects to answer the three research questions about self-perception, and perception of correct and incorrect /r/ and /w/.

**Results:** All children were able to tell the difference between a correctly pronounced /r/ and /w/. Children with no articulation problems were able to accurately distinguish an /r/ error as /w/, whereas children who misarticulate /r/ guessed for the /r/ errors and therefore responded that half were /r/ and half were /w/ productions. Some children who misarticulate /r/ are able to tell the difference between their /r/ and /w/ productions through manipulation of F₂ and F₃.

**Conclusions:** Children who misarticulate /r/ are unable to accurately determine their own errored productions as correct or incorrect. All children were able to distinguish /r/ and /w/ in other speakers.
Relevance to the current study: This article relates to the current study because both are researching interpersonal and intrapersonal perception.


Objectives: This study wanted to investigate the relationship between auditory discrimination (external and internal), consistency of articulation and articulation stimulability. They did so through asking three questions: a) Is there a correlation between auditory discrimination (external and internal) and misarticulation of /s/? b) Does the misarticulation of /s/ correlate to stimulability of /s/? And c) Is there a correlation between auditory discrimination and stimulability of /s/?

Methods: The participants in this study were 25 kindergartners and first graders who had no previous speech therapy and had at most only one other sound error other than /s/. The children participated in the McDonald Screening Test of Articulation to measure consistency of misarticulation of /s/. They also participated in the Carter-Buck Nonsense-Syllable Imitation Test for stimulability of /s/ and the Farquhar-Bankson In-depth Test of Auditory Discrimination to measure internal and external auditory discrimination of /s/. The auditory discrimination test included seven sections. Sections 1-4 assessed the subject’s ability to discriminate /s/ when produced by another speaker. Sections 1-3 required the subject to indicate awareness of the /s/ phoneme whereas section 4 required the subject to make a right/wrong judgment. Sections 5-7 assessed the subject’s ability to discriminate /s/ in his/her own productions. Section 5 required the subject to produce a given word and then acknowledge if the word has an /s/ phoneme in
it. For section 6, the subject was required to use a same/different judgment between the
examiners production and the subject’s production of the same word. In section 7 the
subject is required to produce a word with the /s/ phoneme and then judge his/her own
production as right or wrong.

Results: Researchers found that there is a strong correlation between the
children’s consistent production of /s/ and their ability to discriminate his/her own
productions (internal). Children that produced a greater number of correct /s/ in the
articulation test performed better on the discrimination test. They also found that children
who produced more correct /s/ in the articulation test produced more correct /s/ in the
nonsense syllable stimulability test. There was statistically significant correlation
between the stimulability task and the discrimination task. Table 2 shows the percentage
of correct items on sections of the Farquhar-Bankson In-depth Test of Auditory
Discrimination. Table 3 shows the correlation between the different sections of the
auditory discrimination test.

Conclusions: Children who consistently misarticulate sounds will have more
difficulty discriminating their own speech sounds than other speakers. Children with
more speech sound errors are less stimulable than those with fewer errors.

Relevance to current study: This study observed the correlation between children
with a SSD’s external and internal speech perception. They did this through several
different types of judgment tasks. The current study requires the participants to respond
in a similar way to section 6 of this study. In the current study participants will select a
button corresponding to the syllable/word pair that was different.

https://doi.org/10.1044/cicsd_24_S_57

**Summary:** This systematic review looked at different studies to determine the relationship between speech perception and speech-sound production errors. The different studies had many different testing methods, subjects of different ages and speech abilities. This could account for the wide variety of results from the different studies. Table 1 includes all of the studies that found a connection between speech-sound discrimination and sound production, whereas Table 2 includes all of the studies that did not find a connection. Table 4 includes all of the studies that found auditory discrimination training to be beneficial. As a result, from this systematic review, the authors concluded that to determine if a child has speech perception difficulties, a valid assessment must be used. A valid assessment means that it only tests the speech sounds that the child misarticulates, and it must evaluate the child's internal, perceptual representation. They also concluded that if a child has difficulties with speech perception, then some type of auditory-input therapy may be beneficial.

**Relevance to current study:** This study is beneficial to the current study because it discussed the criteria for a valid assessment for speech perception. The current study assessed speech sounds that the child participants could not produce, and it measured their internal perception.


**Objectives:** The researchers wanted to know if treatment that included ultrasound visual feedback (UVF) would be enhanced by also using auditory perceptual training in conjunction.

**Methods:** 38 children ages 8-16 with /r/ errors participated in pretreatment trials followed by treatment for /r/ errors and distortions using UVF. The subjects were split up into two groups and half did treatment with just UVF, and the other half did treatment with UVF and auditory perceptual training. The pretreatment trials were used to assess auditory perception through a 10-step continuum from /ɹ/ to /w/ using the stimuli “rake” to “wake”. Children were randomly played a stimulus on the continuum and asked to click a response telling if they heard “rake” or “wake”. For the treatment part of this study, children completed 14 treatment sessions from ASHA certified SLPs. Treatment started at the syllable level then progressed to words, phrases, and sentences. The children that received auditory perceptual training, in addition to UVF, by having the children listen to a production and then judge if it was correct or incorrect.

**Results:** There was no difference between the two groups. Both the UVF and UVF+auditory perceptual training made equal gains in improving /r/ productions. They also concluded that children who performed better on the pretreatment trials for auditory perception made more gains in /r/ productions than those with poorer auditory perception despite the type of treatment they received.

**Conclusions:** Children with poorer speech perception do not make as much progress with treatment as those with good speech perception abilities.
Relevance to current study: This study relates to the current study because it shows that there is a problem regarding children with poor speech perception making progress in their treatment. Knowing this justifies the purpose of the current study.


**Objectives:** The purpose of this study was to examine the /s/~/ʃ/ contrast between three different subject groups by asking three questions. “1) Do TD and SSD children who produce the /s/~/ʃ/ contrast accurately discriminate these sounds when listening to recorded productions of children with TD, children with SSD, and themselves? 2) Do children with /s/~/ʃ/ collapse have poorer perception of the /s/~/ʃ/ contrast than children who produce a contrast between these sounds? 3) Do children with covert /s/~/ʃ/ contrast perceive the covert contrast in their own recorded utterances?” (Roepke & Brosseau-Lapré, p. 3764)

**Methods:** Participants included children ages 4-5;11 (14 girls and 7 boys). The three subject groups that participated in this study included 1) children with SSD who collapse (meaning they produce the target sounds the same) /s/~/ʃ/, 2) children with SSD who contrast /s/~/ʃ/ (can produce the target sounds but have incorrect productions of other phonemes), and 3) typically developing children. The stimuli for the tasks included seven monosyllabic /s/~/ʃ/ minimal pairs spoken by a preschool-aged boy and girl with SSD who collapse /s/ and /ʃ/, a school-aged boy and girl with typical speech and the subjects’ own productions. During the task the participant would see two images on a
touch screen tablet, hear a production of a word and was asked to touch the picture that corresponds to the word they heard.

Results: Children who produced the /s/~/[ʃ]/ contrast were able to distinguish the contrast well from the typical developing speakers. Children with an SSD contrast had more difficulty perceiving the contrasts when the stimuli were produced by the SSD speakers than their typically developing peers. Typically developing children and children with an SSD contrast were able to perceive their own speech. Children with an SSD collapse perceived the /s/~/[ʃ]/ contrast more accurately for stimuli produced by typically developing speakers than their own productions. When children with covert contrasts listened to their own productions, they performed at chance for discriminating the target sound.

Conclusions: Typically developing children and children with SSD contrast perceive their own speech well. Children with SSD have less phonological knowledge than typically developing children, even for sounds they can produce.

Relevance to current study: This study is beneficial to the current study because they discuss the difference of internal speech perception for children with a SSD specifically for the sounds they can and cannot produce. They discovered that children who could not produce the speech sound performed at chance for internal perception whereas those who could produce the speech sound performed similar to typically developing children. This is important because the current study is only looking at speech sounds that the children with SSD most often cannot produce.
Objectives: This paper discussed two different experiments that research the relationship between speech production and speech perception. Experiment 1 compared /s/ and /ʃ/ in a minimal pair contrast in both a continuum and independently. Experiment 2 compared /s/ and /θ/ in a minimal pair contrast both in a continuum and independently. They wanted to know if a person’s perception abilities correlated to their ability to produce the speech sounds they were producing.

Methods: Experiment 1 involved 3 groups: typical adults (ages 20-50), typically developing children (ages 4;8-5;11, 6 female and 6 male), and children with SSD, who could not produce the sounds that were tested (ages 8;8-6;0, 2 female and 10 male). There were 2 stimulus sets for this experiment. Stimulus set 1 included a 7-step continuum between synthetic minimal pairs with /s/ and /ʃ/. Stimulus set 2 did not have a continuum and just presented the synthetic minimal pairs. The subjects were played a recording of the stimuli and asked to point to the corresponding picture. Experiment 2 included 3 groups: typical adults (ages 17-29, 5 female and 5 male), typical developing children (ages 6;10-7;11, 8 female and 5 male), and children with SSD, who could not produce the sounds that were tested (ages 6;7-8;2, 2 female and 7 male). The stimuli and procedure were the same as in Experiment 1 except they used the phonemes /s/ and /θ/.

Results: The results of Experiment 1 were different for stimulus set 1 and 2. For stimulus set 1, the adults, typical children and children with SSD all performed at statistically different levels. The adults performed with the highest accuracy, followed by
typical children, and then children with SSD. Within the children with SSD group, 5 children completed the task similarly to the adults and typically developing children, whereas the remaining 7 children were unable to correctly identify stimuli as seen in Figure 4. However, for stimulus set 2, the adults and typical children performed the same and the children with SSD performed significantly worse than the other two groups. The researcher pointed out that the difference between stimulus sets is most-likely due to including a full range of stimuli or not. The results for Experiment 2 were consistent with the results from Experiment 1. Overall, the researchers concluded that typically developing children had better perception abilities than the children with SSD. The results from the experiments also support the notion that children with SSD don't have poor perception overall, but they have poor perception on the specific sound they are unable to produce.

**Conclusions:** Children with SSD had more difficulty with perception than typically developing children and adults. Typically developing children only differed from adults when the Stimulus set included a continuum. Children with SSD have more difficulty perceiving sounds that they are unable to produce.

**Relevance to current study:** This study relates to the current study because both are looking at speech perception in relation to a particular sound.


**Objectives:** The purpose of this study was to determine if children who have an errored production of /r/ can distinguish between their own speech and another person's speech as well as if each production was incorrect or “corrected”.
Methods: 26 children, ages 7;1-13;11 were split into two groups. Group 1 consisted of children who had just started treatment for /r/ in the public schools. Group 2 consisted of children that had been receiving treatment for /r/ for at least two years and were still unable to produce /r/ consistently. All children listened to a recording of 200 words (50 corrected words produced by the child, 50 incorrect words produced by the child, 50 corrected words produced by another child and 50 incorrected words produced by another child). The children were asked to tell the examiner if the word was his/her own speech, as well as if the /r/ was correct or incorrect.

Results: Most subjects could accurately judge their own corrected utterances. Subjects performed at chance for judging their own /r/ incorrect productions. There was no statistical significance between subjects judging their own incorrect utterances versus other children’s incorrect utterances. There was no difference between Group 1 and Group 2 on judging their own incorrect utterances.

Conclusions: These results show a connection between speech perception and production in individuals with a phonological disorder.

Relevance to the current study: This article is relevant to the current study because both looked at how subjects perceived internal and external speech.


Objectives: This paper discusses two studies in which they questioned children’s ability to recognize their own voice on a recording, and if this was dependent on age of the
child, the child’s phonological competence, or the time delay between making the recording and listening to it.

Methods: Study 1: 48 children (25 were 4-5 years old, 23 were 7-8 years old) listened to a recording of an adult say a word as a picture was shown to them. The children then repeated the word and their response was recorded. Immediately after their response was given, the child listened to 3 other children’s recordings of the same word along with their own recording in a randomized order and was asked to tell the instructor which was their own voice.

Study 2: 21 children (4-7 years old) diagnosed with phonological impairment participated in the same procedure as those in Study 1.

Results: The studies concluded that children could recognize their own voice from a recording independent of age or having a phonological impairment. However, recognizing one's own voice was dependent on time delay and older children were less accurate at determining their own voice. Children with a phonological impairment do not rely on their speech errors as a cue to distinguish their own voice from others.

Conclusions: Children are able to distinguish their own recorded voice from others independent of age and phonological competency. Recordings of the child’s own speech might be useful in intervention.

Relevance to the current study: This article researched whether or not children are able to recognize their own voice from a recording and if their age or phonological competency changed their accuracy. The current study used recordings of the subjects’ own voice in the experiment.
Objectives: The purpose of this study was to determine how children with and without a phonological disorder react to hearing synthetically modified versions of their own speech. The study discusses if the children were able to notice a difference and if they accurately judged the utterances as correct or incorrect productions. This study also looked at how the children’s perception of their utterances changed after a delayed playback.

Methods: A total of 31 children (11 diagnosed with phonological disorder and 20 typically developing), ages 4-6 years old, were recorded saying target words. Some of the words were synthetically modified through phonetic transplantation. The children listened to a recording of a word (the child’s original production or the synthetically modified) and were asked if the pronunciation was correct or incorrect. This occurred immediately after the child was recorded saying the word and after a delay.

Results: Children with a SSD identified their own incorrect utterances as correct when immediately played back. When the playback was delayed, the children with SDD perceived their incorrect utterances as incorrect.

Conclusions: A child’s ability to accurately perceive their speech is dependent on the amount of time since the production of the utterance.

Relevance to the current study: This article is relevant to the current study because both looked at the subject's speech perception and had the subjects choose between same/different parameters.

**Objectives:** The purpose of this study was to determine if children who consistently misarticulate /r/ could accurately compare his/her own productions in time to another speaker (interoceptive task) versus if the children could accurately compare his/her own productions to another speaker from a recording (exteroceptive task). This study also looked at children’s abilities at different ages as well as test items being presented in children’s voices versus adult voices.

**Methods:** 40 children were split into two groups (Group 1: 20 children in grades 1-3. Group 2: 20 children in grades 4-6) and performed various tasks. The “Type 1 task” included the child comparing his/her own production with the correct production of another speaker. The “Type 2 task” included the child comparing his/her own production with the incorrect production of another speaker. The “interoceptive task” involved the child naming a picture, listening to another speaker name the picture from a recording, and determining if the sounds were the “same” or “not the same”. The “exteroceptive task” involved the child listening to his/her own productions from a recording followed by another speaker from a recording and determining if they were the “same” or “not the same.”

**Results:** Children were able to distinguish their misarticulations from the correct productions of another speaker more accurately during the exteroceptive task than the interoceptive task. There was no difference between the interoceptive task and the exteroceptive task for Task 2 (comparing incorrect productions).
Conclusions: Children show more difficulty monitoring their own productions in live time rather than when played back to them from a recording.

Relevance to the current study: This study is relevant to the current study because both studies observed internal speech perception. Additionally, the stimulus used in both studies was a recording of the subject’s own productions.
APPENDIX B

Participant Consent Form

Consent to be a Research Subject

**Title of the Research Study:** Investigating interpersonal versus intrapersonal speech perception in adult listeners

**Principal Investigator:** Katy Cabbage, Ph.D.

**Introduction**
This research study is being conducted by Katy Cabbage, Ph.D. at Brigham Young University to determine the difference of speech perception when listening to another speaker vs your own speech. You were invited to participate because you are 18 years old or older and have no history of any neurological or cognitive impairment (e.g., autism, Down syndrome) as well as no history of speech or language concerns, per self-reported case history.

**Procedures**
This study is about how people perceive speech differently when spoken by another person versus their own speech. To participate you must be a native English speaker. The study will take place at the BYU John Taylor Building in Room 109 at a time convenient for you. The study involves two sessions of activities.

During the *first session*, you will complete a hearing screening, a speech sample, and a listening task. This session will take about 30 minutes. The hearing screening will involve you listening to a series of tones and raising your hand to indicate which tones you heard. The speech sample will involve you reading two passages aloud while we audio record your voice. The listening task will involve you listening to different sets of prepared words. We will provide instructions for everything we ask you to do and allow you to ask questions to ensure you understand what to do. During this session you will also read a list of words that we will audio record for later use in the study. These recordings will be used in the research study. With your permission (see below), we may use these audio recordings for future research and training purposes. You may elect to allow us to only use these recordings in the current study. Your decision will not affect your ability to participate in this study or you relationship with the Principal Investigator or the university.

During the *second session*, you will complete additional listening tasks with different recorded words from those you listened to in the first session. This session will take about 60 minutes.

During both sessions, you will be allowed to take breaks as often as necessary. We expect both sessions to be completed within 1-3 weeks of each other, based upon your availability.

**Risks/Discomforts**
There is minimal risk associated with this study. It is possible that during participation, you may become bored with the tasks. We will provide you with breaks as often as is necessary. You may stop participation at any time.

**Benefits**
There will be no direct benefits to you. It is hoped, however, that through your participation researchers may learn about speech perception practices that may be able to assist speech therapists in improving their treatment methods.

IRB NUMBER: IRB2021-355
IRB APPROVAL DATE: 12/01/2021

BYU IRP 09/2021
Confidentiality
The research data will be kept in a secure location on password protected and encrypted computers accessible only to research staff. All forms will be stored in a locked filing cabinet accessible only to research staff. All identifying information will be removed. The data will be indefinitely archived on secure password protected computers and accessible only to research staff.

Data Sharing
We will keep the information we collect about you during this research study for analysis. Your name and other information that can directly identify you will be stored securely and separately from the rest of the research information we collect from you.

De-identified data from this study may be shared with the research community, with journals in which study results are published, and with databases and data repositories used for research. We will remove or code any personal information that could directly identify you before the study data are shared. Despite these measures, we cannot guarantee anonymity of your personal data.

The results of this study could be shared in articles and presentations but will not include any information that identifies you unless you give permission for use of information that identifies you in articles and presentations.

Compensation
You will receive $10 cash at completion of your participation. If you elect to withdraw at any time, you will still receive full compensation.

Participation
Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your class status, grade, or standing with the university.

Questions about the Research
If you have questions, concerns, or complaints, you can contact the Principal Investigator, Katy Cabbage at (801)422-0507 or kcabbage@byu.edu.

Questions about Your Rights as Research Participants
If you have questions regarding your rights as a research participant contact Human Research Protection Program by phone at (801) 422-1461; or by email: at BYU.HRPP@byu.edu.

Statement of Consent
I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Name (Printed): ____________________ Signature ____________________ Date: _______________

IRB NUMBER: IRB2021-355
IRB APPROVAL DATE: 12/01/2021
_____ Initial here to allow us to keep your information in a secure database to contact you for future studies.

AUDIO RELEASE

We will be audio recording you during participation in this study. Please indicate what uses of this audio you are willing to permit, by initialing next to the uses you agree to and signing below. This choice is completely up to you. We will only use the audio in the ways that you agree to. In any use of the audio, you will not be identified by name.

_____ I give permission for my recordings to be used for research lab training purposes.

_____ I give permission for my recordings to be used for research presentations at conferences.

_____ I give permission for my recordings to be used for teaching purposes in the PI’s university classes.

_____ I do not give permission for my recordings to be used for any purpose outside of this study.

I have read the above descriptions and give my express written consent for the use of the audiotapes as indicated by my initials above.

Name (Printed): __________________________ Signature __________________________ Date: __________
APPENDIX C

Institutional Review Board Approval Letter

Memorandum

To: Katy Cabbage
Department: BYU - EDUC - Communications Disorders
From: Sandee Aina, MPA, HRPP Associate Director
       Wayne Larsen, MAcc, IRB Administrator
       Bob Ridge, Ph.D., IRB Chair
Date: December 01, 2021
IRB#: IRB2021-355
Title: Investigating interpersonal versus intrapersonal speech perception in adult listeners

Brigham Young University’s IRB has approved the research study referenced in the subject heading as expedited level, categories 4, 6, and 7. This study does not require an annual continuing review. Each year near the anniversary of the approval date, you will receive an email reminding you of your obligations as a researcher. The email will also request the status of the study. You will receive this email each year until you close the study.

The IRB may re-evaluate its continuing review decision for this decision depending on the type of change(s) proposed in an amendment (e.g., protocol change that increases subject risk), or as an outcome of the IRB’s review of adverse events or problems.

The study is approved as of 12/01/2021. Please reference your assigned IRB identification number in any correspondence with the IRB.

Continued approval is conditional upon your compliance with the following requirements:

1. A copy of the approved informed consent statement and associated recruiting documents (if applicable) can be accessed in IRIS. No other consent statement should be used. Each research subject must be provided with a copy or a way to access the consent statement.
2. Any modifications to the approved protocol must be submitted, reviewed, and approved by the IRB before modifications are incorporated in the study.
3. All recruiting tools must be submitted and approved by the IRB prior to use.
4. In addition, serious adverse events must be reported to the IRB immediately, with a written report by the PI within 24 hours of the PI’s becoming aware of the event. Serious adverse events are (1) death of a research participant; or (2) serious injury to a research participant.
5. All other non-serious unanticipated problems should be reported to the IRB within 2 weeks of the first awareness of the problem by the PI. Prompt reporting is important, as unanticipated problems often require some modification of study procedures, protocols, and/or informed consent processes. Such modifications require the review and approval of the IRB.

Instructions to access approved documents, submit modifications, report complaints, and adverse events can be found on the IRB website under IRIS guidance: https://orca.byu.edu/IRB/Articulate/Study_Management/story.html.