Acquisition of the /r/-/w/ Contrast by Preschool Children

by

Hyuna Amelia Kim

A thesis prepared in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

(Communication Sciences and Disorders)

at the

UNIVERSITY OF WISCONSIN-MADISON 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This thesis would not have been possible without individuals who guided and encouraged me throughout the research project. I give the greatest thanks to my advisor, Dr. Jan Edwards for her support, guidance, and patience. I am thankful for my committee members Dr. Marios Fourakis and Dr. Gary Weismer, for their expertise and time. I am grateful for Dr. Mary E. Beckman and Dr. Patrick Reidy for developing stimulus preparation scripts and coding. I give special thanks to the /r/-/w/ analysis team, Dr. Benjamin Munson and Alisha Blackman at the University of Minnesota, for the great teamwork they have shown. I am grateful to all the Learning to Talk members for truly investing time and effort in data collection and analysis. I am thankful for the families and children for their generosity and time to participate in the production study, and adult participants in the speech perception study. Lastly, I am thankful to my family who never cease to love and encourage me in all conditions. Words cannot express my sincere gratitude to all of you.

This research was funded by the National Institute for Deafness and Other Communicative Disorders (NIDCD 02932) to Dr. Jan Edwards, Dr. Mary E. Beckman, and Dr. Benjamin Munson.

ABSTRACT

The purpose of this study was to 1) explore a robustness of contrast measure of preschoolers' acquisition of the /r/ vs. /w/ contrast based on the ratings of naïve adult listeners using a visual analogue scale (VAS); and 2) identify which factors at age 2 ½ to 3 predict children's production of /r/-/w/ contrast at age 3 ½ to 4 years. There were 105 child participants, with ages from 28 to 40 months at the time of their initial visit and from 39 to 52 months at their second visit, one year later. All children produced /r/ and /w/ in word-initial position in familiar words in a picture-prompted auditory word recognition task at their second visit. Their productions were recorded and then transcribed by a trained phonetician. Consonant-vowel syllables containing these productions were rated by 76 adults (approximately 20 adults per production) using a VAS scale. VAS ratings were significantly correlated with transcribed accuracy, suggesting that the VAS ratings are a valid measure of children's acquisition of the contrast. A robustness of contrast measure was developed from the VAS ratings and it was significantly predicted by children's articulatory ability and children's receptive vocabulary size at the initial visit, one year prior. Future research is needed to determine whether VAS listener ratings can be clinically useful and to understand the nature of the relationship between vocabulary size and speech production ability.

SPECIFIC AIMS

Children gradually learn to produce almost all of the speech sounds in their native language in the first 5 to 6 years of life. Some of the earlier sound contrasts (such as the contrast between /b/ and /p/) are acquired by age 3, while later-acquired sound contrasts (such as the contrast between /r/ and /w/) are mastered by ages 7 to 9 (e.g., Smit, Hand, Freilinger, Bernthal, & Bird, 1990). While many children have difficulty with these later-acquired contrasts, there has been little research on what risk factors might predict whether an individual child will have difficulty with these more challenging contrasts. The purpose of this study was to determine whether individual differences in speech perception or speech production at 2 ½ to 3 years predict individual differences in the acquisition of the /r/ vs. /w/ contrast at 3 ½ to 4 years. Acquisition of the /r/ vs. /w/ contrast will be measured by using native adult perceptual judgments to develop a robustness of contrast measure to differentiate correct [r] from correct [w] (see Bernstein, 2015 for a comparable measure for /t/ vs. /k/).

I am proposing to examine which factors predict the acquisition of the /r/ vs. /w/ contrast with the following specific aims:

1) Can native adult judgments using a visual analogue scale (VAS) be used to develop a robustness of contrast (ROC) measure that will reliably differentiate correct [r] from correct [w]?
2) Which measures at age 2 ½ to 3 years predict children's acquisition of the /r/-/w/ contrast at 3 ½ to 4 years? The measures that will be examined include:
    a) Speech perception ability, as measured by a minimal pairs task;

b) Receptive and expressive vocabulary skills, as measured by *Peabody Picture Vocabulary Test, 4th Edition* (PPVT-4, Dunn & Dunn, 2007) and *Expressive Vocabulary Test, 2nd Edition* (EVT-2, Williams, 2007);

c) Perceptual robustness of contrast for /t/-/k/ as measured by native adult perceptual judgments on a visual analog scaling (VAS) task (Bernstein, 2015);

d) Acoustic robustness of contrast for /t/-/k/ as measured by centroid (Johnson, 2016);

e) Psycho-acoustic robustness of contrast for /s/-/ʃ/ as measured by peak ERB (Reidy, 2015); and

f) Articulatory ability, as measured by the *Goldman Fristoe Test of Articulation, 2nd Edition* (GFTA-2, Goldman Fristoe, 2000).

Developing methods to sensitively assess speech-sound production will provide a means to advance our knowledge of the factors contributing to or hindering early speech sound development. This information will not only increase the potential for preventative measures, but will also promote early identification and development of more focused treatment approaches for children with speech sound disorders (SSD).

CHAPTER ONE

Introduction

Phonological development in typically developing children is a gradual process. The majority of speech sounds are mastered by about six years and some sounds are mastered considerably earlier (e.g., Smit et al., 1990; Kent, 1992). Some of the earliest sound contrasts, such as /p/ and /b/, are acquired by about 2 to 3 years of age. These early-acquired contrasts have relatively low motor demands. For example, many early-acquired contrasts involve rapid ballistic movements, such as the contrast between labial and alveolar stops (Kent, 1992). Other early-acquired sounds involve slow "ramp" movements; for example, /w/ is produced by a controlled speed over a relatively long time period. However, there are other speech sounds that are not mastered until about 5 to 6 years of age. These later-acquired sound contrasts are much more demanding in terms of motor control. For example, the contrast between the two sibilant fricatives of English (/s/ vs. /ʃ/) requires very precise tongue placement, while the contrast between /r/ and /w/ requires the simultaneous adjustment of both the tongue tip and tongue dorsum (Kent, 1992).

While most children develop a phonological system at a normal rate and with the typical order of acquisition, about 5% of children have difficulty with speech sound acquisition in spite of having normal hearing, normal motor control, normal cognitive ability, and normal socio-emotional development (National Institute on Deafness and Other Communication Disorders, 2010). These children are described as having a functional speech sound disorder (SSD). A functional SSD is "any combination of difficulties with perception, motor production, and/or the phonological representation of speech sounds and segments that impact speech intelligibility" (American Speech-Language-Hearing Association, 2001). Children with SSD are at risk in a number of areas beyond speech sound development, including reduced vocabulary size, decreased literacy skill, and ultimately poor academic performance and occupational outcomes (e.g., Rvachew, 2006; Rvachew & Grawburg, 2006; McCormack, McLeod, McAllister & Harrison, 2009).

A number of studies have investigated the risk factors associated with speech and language impairment (Snowling, Bishop, & Stothard, 1999; Campbell et al., 2003; Harrison & McLeod, 2010). For instance, Harrison and McLeod (2010) examined risk factors of speech and language impairment combined for 4- to 5-year-old children as part of a longitudinal study. They found that consistent risk factors included both biological and psychosocial factors (e.g., male, consistent hearing difficulties, reactive temperament), and family factors (e.g., lower household income, paternal and maternal education, maternal well-being). Campbell et al. (2003) examined 3-year-old children with speech delay of unknown origin. Out of seven variables, low maternal education, male sex, and family history of developmental communication disorder were found to be a significant risk factor for speech delay. Furthermore, a child with all of these listed factors was 7.71 times more likely to have a speech delay than a child without any risk factors.

Difficulty with speech perception is also associated with SSD. Speech perception ability has been shown to have a direct effect on speech production (Sherman & Gieth, 1967; Locke, 1980; Rvachew, 2006; Rvachew & Grawburg, 2006). Some research has found that speech perception training facilitates speech production abilities in children with phonological impairment of specific target sounds (Jamieson & Rvachew, 1992; Rvachew, 1994; Rvachew, Nowak, & Cloutier, 2004). For example, Jamieson and Rvachew (1992) conducted a study with four children with functional articulation disorders and examined whether speech perception training would yield positive effect in remediating expressive phonological errors. They found that perception training significantly facilitated speech production. Rvachew et al. (2004) examined similar benefits of perceptual training for children with SSD. Children with speech perception intervention had a significant gain in both perception and articulatory skills over the control group.

Understanding the risk factors and other speech-associated abilities that are specific to SSD will provide a means to advance our knowledge of the factors contributing to or hindering early speech sound development. This information will support to prevent potential risk factors, promote early identification, and develop more focused treatment approaches for children with speech sound disorders (SSD).

This paper will focus on the acquisition of a single phonetic contrast, the contrast between /r/ and /w/ in young English-speaking children. This contrast is particularly important, because /r/ is a high-frequency sound in English and therefore is important for intelligibility. Furthermore, the North American rhotic /r/ is extremely challenging, both with respect to speech sound development and therapeutic intervention (Shriberg, 2009; McAllister Byun, Hitchcock, & Swartz, 2014). The North American rhotic /r/ can be produced in two ways, the “retroflex /r/” and the “bunched /r/.” The “retroflex /r/” is produced with a raised tongue tip and a lowered tongue dorsum, and the “bunched /r/” is produced with a lowered tongue tip and a raised tongue dorsum (Espy-Wilson, 1992; Zhou, Espy-Wilson, Boyce, Tiede, Holland, & Choe, 2008). Other studies found that the tongue placement for the production or /r/ is not categorical, but rather continuous across the speakers and contexts (Delattre & Freeman, 1968; Westbury, Hashi, & Lindstrom, 1998). Whether or not the production of /r/ is categorical or continuous, methods of /r/ production require fine-grained motor control, because different parts of the tongue must be controlled simultaneously. Due to the complexity of the production of /r/, children often substitute [w] for a target /r/ (Smit et al., 1990; Slawinski & Fitzgerald, 1998). The [w] for /r/ substitution is common in children for several reasons. Both /w/ and /r/ are accompanied by lip rounding, /w/ is acquired earlier than /r/, and the motor control to produce /w/ is less demanding than the motor control to produce /r/.

A number of factors might be associated with the acquisition of a challenging speech contrast, such as the /r/-/w/ contrast. As noted above, speech perception skills have been found to be related to the acquisition of age-appropriate consonant articulation, at least in children with SSD's (e.g., Rvachew, 2006; Rvachew & Grawburg, 2006). There is also some evidence that vocabulary size plays a role, although the role of vocabulary is poorly understood. Nevertheless, a number of studies have observed that children with SSD's tend to have somewhat smaller vocabularies than children with typical phonological development (e.g., Munson, Edwards, & Beckman, 2005; Rvachew & Grawburg, 2006). However, the vocabulary scores of children with SSD's are generally within the normal range and there is little evidence for co-morbidity between functional speech and language disorders (Shriberg, Tomblin, & McSweeny, 1999). It was also hypothesized that articulatory ability more generally would be predictive of how readily a child would acquire a challenging phoneme contrast. This could be measured by a standardized articulation test, which elicits all consonants in all word positions and scores children relative to age expectations. Alternatively, it may be that a more fine-grained measure of the acquisition of a particular consonant contrast might be a better predictor. In this study, the robustness of contrast for two consonant contrasts (/t/ vs. /k/, which is acquired relatively early and /s/ vs. /ʃ/, which is acquired relatively late) were selected as potential predictors of the acquisition of the /r/ vs. /w/ contrast. It was hypothesized that the robustness of contrast of /s/ vs. /ʃ/ would be a better predictor than a similar measure for /t/ vs. /k/ because the acquisition of the sibilant place contrast places greater challenges on motor control than the acquisition of the lingual stop place contrast.

In this study, children's productions of /r/ will not simply be transcribed as correct or incorrect. This is because there is considerable evidence that children acquire contrasts gradually rather than categorically (e.g., Macken & Barton, 1980; Maxwell & Weismer, 1982; Lee, Potamianos, & Narayanan, 1999; Hewlett & Waters, 2004; Li, Edwards, Beckman, 2009; Munson, Edwards, Schellinger, Beckman, & Meyer, 2010). In many cases, children may go through a period of *covert contrast* before they produce distinct contrast between two consonants. *Covert contrast* is a subphonemic contrast that is distinguishable, but is not warranted being transcribed by a different phonetic symbol (Munson et al., 2010). Many studies have found evidence of covert contrast in the speech of children with and without phonological disorder. One of the earliest studies (Macken & Barton, 1980) investigated the acquisition of the voicing contrast in four 2-year-old children. Macken and Barton measured voice onset time (VOT) of word-initial stops and found that all four children went through a stage where the voiceless stops were perceived as voiced by adult listeners, although children produced the voiceless stops with longer VOTs than voiced stops. Li, Edwards, and Beckman (2009) also found evidence of covert contrast in the acquisition of voiceless sibilant fricatives (/s/-/ʃ/ in English-speaking children and /s/-/ɕ/ in Japanese-speaking children) using the acoustic measures of centroid frequency and onset F2 frequency (Li et al., 2009). Covert contrast also has clinical implications; children with phonological disorder who exhibited covert contrast made faster progress in therapy than children who made no contrast at all (Tyler, Figurski, & Langsdale, 1993).

A number of studies have examined the acoustic characteristics of the /r/-/w/ contrast (Dalston, 1975; Sharf & Ohde, 1983; Sharf & Ohde, 1984; Hoffman et al., 1985; Ohde & Sharf, 1987; Hagiwara, 1995; Slawinski & Fitzgerald, 1998). Dalston (1975) and others found that the distance between second formant (F2) and third formant (F3) onset frequencies distinguished /r/ from /w/. Stimuli with high F2 and low F3 onset frequencies were identified as /r/, and stimuli with low F2 and high F3 were identified as /w/. For /w/ the frequency of F2 was lower than F2 for /r/. Figure 1 presents a schematic spectrogram of F2 and F3 onset frequencies in /r/-/w/ contrast. Formant transition rates were also closely related to /r/ vs. /w/ contrasts. For instance, children's productions of /r/ were perceived more accurately when /r/ was produced with a shorter F2 transition rate, and F3 transition duration and rate of /r/ consistently differentiated from /w/ (Dalston, 1975, Ohde & Sharf, 1984).

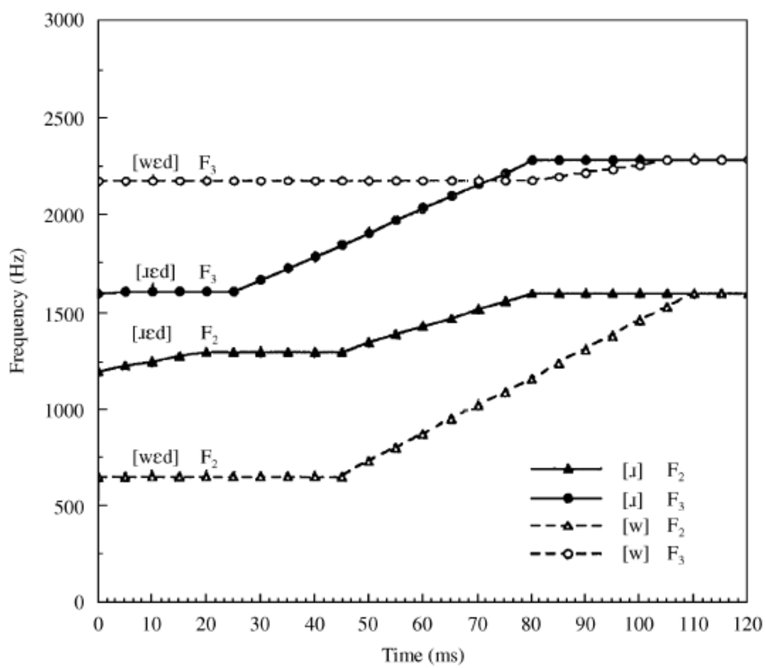

Figure 1. Schematic spectrogram of the F2 and F3 onset frequencies of word-initial /r/ and /w/. Solid lines represent F2 and F3 frequencies of /r/ "red". Dotted lines represent F2 and F3 frequencies of /w/ "wed" (taken from Slawinsky & Fitzgerald, 1998).

Several recent studies have used adult perception ratings rather than acoustic measures to identify covert contrast. As noted by Munson et al. (2010), covert contrast is “ubiquitous” when adult perceptual measures are used. Munson and colleagues found that adults rated substitutions of [s] for /ʃ/ as less /s/-like than correct productions of /s/ and also rated substitutions of [ʃ] for /s/ as less /ʃ/-like than correct productions of /ʃ/. These studies gave adults a range of options rather than simply “correct” or “incorrect”. Instead, adults rated sounds using a visual analogue scale (VAS) that had “s” on one endpoint and “sh” on the other endpoint. Listeners were asked to click anywhere on the line to indicate how “s”-like or how “sh”-like a particular consonant was. Stimuli were all consonant-vowel sequences and included productions of /s/ and /ʃ/ that were transcribed as one of the following: one, correct; two, substitutions of one sibilant fricative for the other; three, intermediate between /s/ and /ʃ/.

More generally, VAS has proven to be a simple and a reliable tool to measure the development of phonetic contrasts. Munson, Schellinger, and Carlson (2012) reviewed previous studies using the VAS measure in adult listener ratings and found that it was an effective way to examine gradual change in children’s speech production. They found that VAS yields high inter-rater reliability and high correlations with the acoustic characteristics of the sounds under consideration. Bernstein (2015) similarly used adult perceptual judgments with VAS to develop a robustness of contrast measure for the development of the /t/ vs. /k/ contrast. Adults listened to consonant-vowel sequences with initial /t/ and /k/ targets. The consonants used as stimuli had been transcribed as correct, intermediate between /t/ and /k/, or substitutions of /t/ and /k/ for each other. Listeners rated the productions on a VAS with endpoints labeled as “correct ‘t’” and “correct ‘k’”, and they were encouraged to use the entire line. The VAS judgments were used to develop a child-level robustness of contrast measure. To do this, a mixed-effects regression with random slopes for subjects was run across all participants. The dependent variable was the VAS judgments and the independent variable was consonant type. The child-level slopes were used as a perceptual robustness of contrast measure. Bernstein found that transcription accuracy (percent correct) and the robustness of contrast measure based on VAS judgments were highly correlated, but that the robustness of contrast measure was more sensitive to differences among children with high accuracy.

For this study, a similar perceptual robustness of contrast measure will be used to quantify children's acquisition of the /r/ vs. /w/ contrast. The purpose of this study was: 1) to examine whether native adult judgments using a visual analogue scale (VAS) can be used to develop a robustness of contrast (ROC) measure that will reliably differentiate correct [r] from correct [w]; and (2) to examine whether individual differences in speech perception or speech production abilities at an earlier age (2½ to 3 years) predicts individual differences in the acquisition of the /r/-/w/ contrast at a later age (3 ½ to 4 years). Measures that will be included are as follows:

a) speech perception ability, as measured by a minimal pairs task;
b) receptive and expressive vocabulary skills, as measured by *Peabody Picture Vocabulary Test, 4th Edition* (PPVT-4, Dunn & Dunn, 2007) and *Expressive Vocabulary Test, 2nd Edition* (EVT-2, Williams, 2007);
c) perceptual robustness of contrast for /t/-/k/ as measured by native adult perceptual judgments on a visual analog scaling (VAS) task (Bernstein, 2015);
d) acoustic robustness of contrast for /t/-/k/ as measured by centroid (Johnson, 2016);
e) psycho-acoustic robustness of contrast for /s/-/ʃ/ as measured by peak ERB (Reidy, 2015); and

f) articulatory ability, as measured by the *Goldman Fristoe Test of Articulation, 2nd Edition* (GFTA-2, Goldman Fristoe, 2000).

This study differs from previous research in several respects. First, gradient measures of articulatory development (robustness of contrast) rather than simply accuracy (percent correct) will be used. Second, the children in this study are all typically developing, rather than children with SSD, although it is expected that not all children will produce /r/ accurately at ages 3 ½ to 4 years. The study is designed to determine whether their acquisition of the /r/ vs. /w/ contrast can be predicted by measures of speech perception and speech production at age 2 ½ to 3 years.

CHAPTER TWO

Methods

There are three parts to this section. First, the data collection for the production and other data for the child participants are described. Second, the perception experiment with adults that was used to collect the VAS data is described. Finally, the analyses used to address the specific aim of this thesis are described.

CHILD SPEECH PRODUCTION

Participants:

Participants consisted of 105 children with descriptive information shown in Table 1. Children were between ages of 28 and 40 months when they entered the study. This experiment is part of a larger longitudinal study and so the participants were also tested again one year later. All participants were recruited from the Madison, Wisconsin and Minneapolis, Minnesota metropolitan areas. All subjects met the following criteria: 1) no diagnosis of speech, language, or hearing disorder/delay, or developmental disorder via parent report; 2) normal hearing assessed via hearing screening; and 3) monolingual native speaker of English. Maternal education level was determined from a demographic survey that primary caregivers filled out when they brought their child for the study.

Table 1. Descriptive information on child participants (standard deviation [SD] in parentheses).

Descriptor	Value
Boys/Girls	52/53
Mean age in months (SD) and range at time point 2	46.4 (3.42), range = 39-52
Number of AAE speakers	8
Mean PPVT-4 standard score (SD) at time point 1	113 (17)
Mean GFTA-2 standard score (SD) at time point 1	93 (13)
Maternal education level[1]	Low = 10 Middle = 22 High = 73

[1]Low maternal education level = less than high school, high school diploma, GED); middle maternal education level = trade school, technical/associate's degree, some college; and high maternal education level = college or graduate degree.

***Procedure*:**

<u>*Stimuli*</u>. The stimuli were words with word-initial /r/ or /w/, presented in a picture-prompted auditory word repetition task. The pictures were color photographs of the target words, based on age of acquisition norms (Edwards & Beckman, 2008). Examples of these pictures are shown in Figure 2. To control for anticipatory coarticulation, the vowel after the target consonant was balanced across both target consonants. Four words were selected to fill in each of the vowel quadrant (i.e., front and high, front and non-high, back and high, back and non-high). There were no familiar words for /w/ and only a very limited number of familiar words for /r/ in the back-high vowel quadrant, so no words from this quadrant were elicited. If there were not four familiar words for a target sound for a particular vowel quadrant, then one or more words were repeated twice. Table 2 shows the words used for the /r/ and /w/ contrast. The stimuli were recorded by two young adult female speakers, one spoke Mainstream American English (MAE)

and one spoke African American English (AAE). If the caregivers of the participants spoke African American English to their child at the time of the initial visit, then the stimuli were presented in AAE. Otherwise, the MAE stimuli were used. This decision was made by one of two examiners, who were both native speakers of AAE and fluent dialect-shifters between AAE and MAE. Stimuli were presented in children's native dialect to minimize the processing demands that might be involved if stimuli were presented in an unfamiliar dialect (e.g., Edwards, Gross, Chen, MacDonald, Kaplan, Brown, & Seidenberg, 2014) and to so that children would be repeating words in a familiar dialect. The recordings were segmented and then normalized for amplitude across all tokens.

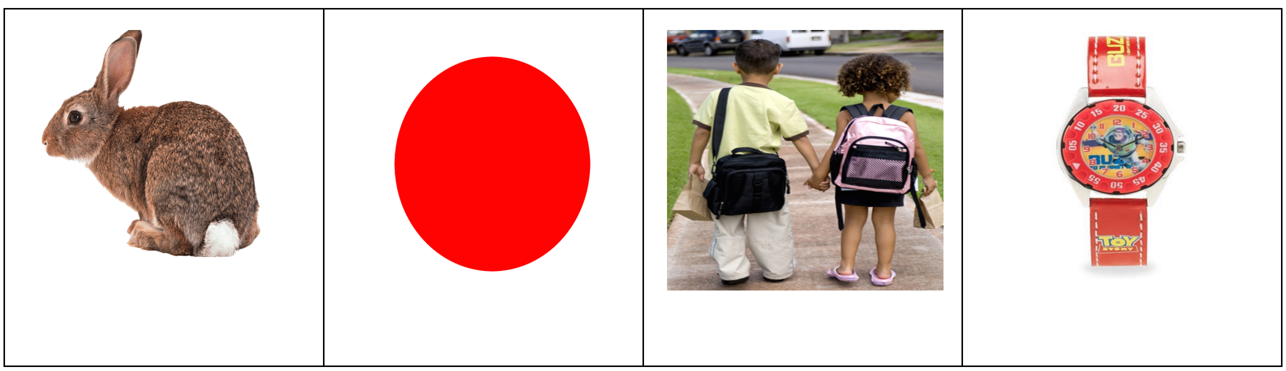

Figure 2. Examples of color photographs and stimuli used for the Real Word Repetition Test: (from left to right) *rabbit, red*, *walk*, and *watch*.

Procedure. Each participant was recorded individually in a quiet room. For the picture-prompted auditory word repetition task, participants were presented with an image on the computer screen paired with an auditory presentation of the target word. Children were asked to repeat each word as accurately as possible. Children's productions were recorded with a Shure SM 81 cardioid condenser and a Marantz PMD671 solid-state recorder. The /r/- and /w/-initial words were a subset of a larger word list that included 90 words. The words were presented in a pseudo-random order so that two repetitions of a word did not occur in a row. If a word was repeated, a different picture and sound file were used in each repetition. There were four practice trials prior to the experimental trials. If there was a problem with a child's response on a trial, the stimulus was replayed only once and/or the child was prompted verbally (e.g., *What did the computer say?*). Children were given a second chance to respond under the following circumstances: the child or the experimenter talked over stimulus, the child's response overlapped with background noise, the child was distracted and did not respond, the child's response was too quiet or loud (clipping), the child produced an incorrect word, and the child did not respond at all. Children received tangible, visual, and verbal reinforcements to motivate and keep them on task.

Table 2. Target /r/- and /w/-initial words. Asterisk (*) indicates the stimulus was repeated two times

Vowel Quadrants	**/r/**	**/w/**
Front – High	*reading* *ring* *rain* *raisins*	*wheel* *window* *wind* *waiting*
Front – Non-high	*red** *rabbit**	*wet** *web**
Back – High		
Back – Non-high	*running* *rock** *rocking*	*watch** *walk* *washer* *water*

<u>*Analysis*</u>. Children's productions of /r/ and /w/ were analyzed using Praat (Boersma & Weenink, 1992). There were three steps to the analysis: (1) word segmentation, (2) stimulus extraction and tagging, (3) transcription, and (4) stimulus development. For word segmentation,

the onset and offset of each target word was marked. For stimulus extraction, the tagger listened to the initial consonant and vowel in isolation from the rest of the word and decided whether the target response was transcribable. The tagger was instructed to tag the first usable production out of the child's one or two possible productions. A production was "useable" if there was no noise obstruction in the initial consonant-vowel sequence, and if the onset of voicing was visible in the spectrogram. If no useable productions were available for a trial, it was marked as missing data. After selecting a production, the tagger marked consonant onset and syllable offset. Two criteria were used for consonant onset: either the onset of voicing or the first upward waveform of the periodicity. There were two ways to mark the offset. The vowel offset was marked either 1) at the end of the glottal pulse of the vowel (see Figure 3) or 2) at the end of the syllable for consonants ending with nasals and liquids (see Figure 4). The two-way offset tagging was used, because the distinction of the vowel offset and certain consonant types (i.e., nasal, liquid) were not always clear-cut.

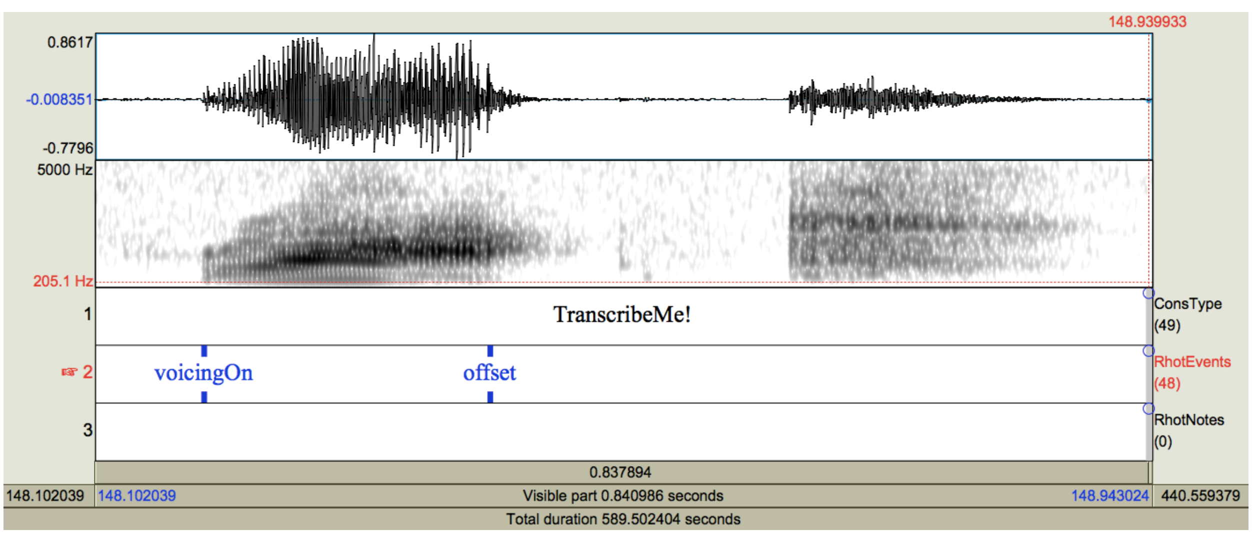

Figure 3. Production of *rock* by a child participant. Markers of a consonant onset ("voicingOn") and vowel offset ("offset") in the stimulus tagging analysis. Duration of the boundary to be 228 milliseconds (ms).

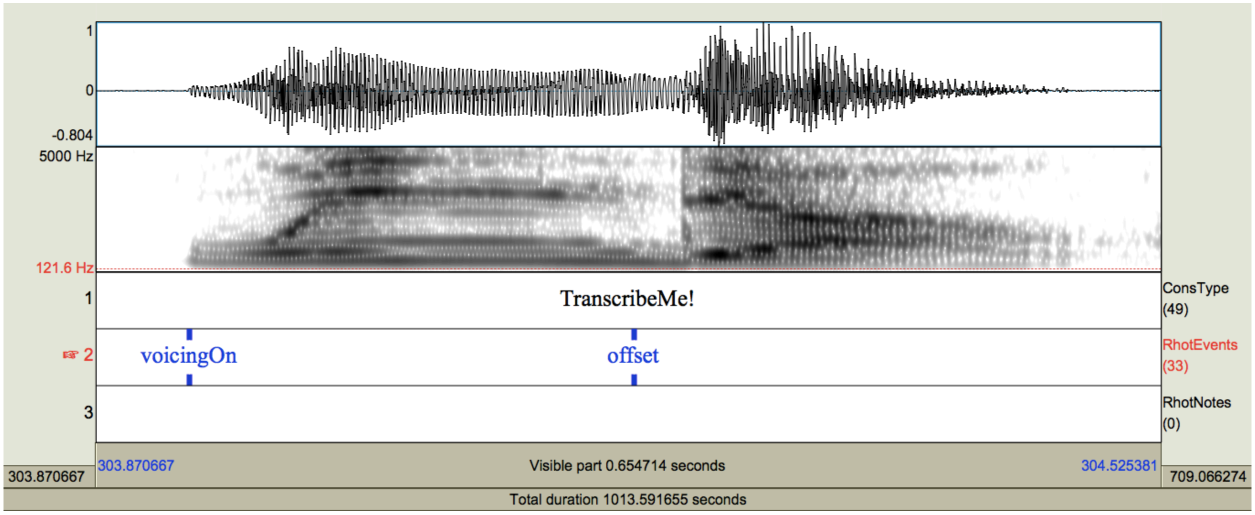

Figure 4. Production of *window* by a child participant. Markers of a consonant onset and vowel offset ending with nasals in the stimulus tagging analysis. Duration of the boundary to be 300 ms.

The third analysis phase was transcription. A native-speaker phonetician transcribed the initial consonant using the different transcription choices shown in Table 3.

Table 3. Transcription category descriptions for /r/ - /w/.

Transcription	Description
[r]	Perceived as correct [r] for target /r/; counted as correct.
[$r:w]	Perceived as intermediate [r] or [w], but closer to [r] for target /w/; counted as incorrect.
[w:$r]	Perceived as intermediate [r] or [w], but closer to [w] for target /w/; counted as correct.
[$w:r]	Perceived as intermediate [w] or [r], but closer to [w] for target /r/; counted as incorrect.
[r:$w]	Perceived as intermediate [w] or [r], but closer to [r] for target /r/; counted as correct.
[w]	Perceived as correct [w] for target /w/; counted as correct.
Other	Perceived as outside of [r] and [w]; counted as incorrect.

The last analysis phase was stimulus development for the adult perception task. In previous studies (i.e., Bernstein, 2015), individual stimuli comprised the initial consonant plus

150 ms of the following vowel. Because we could not reliably identify the end of the initial consonant, we instead developed a systematic procedure to insure that all stimulus tokens were similar in length, and included the initial consonant and enough of the following vowel that listeners could take into account any anticipatory coarticulation effects that might influence their ratings. This protocol is described in the following section. The duration between the “consonant onset” and “syllable offset” for all target productions was measured. Following this, the durations of these intervals were measured. Based on this distribution, it was determined that selecting a fixed stimulus length of 175 ms would require the least additional editing. Approximately 5% of the potential stimuli had intervals between "consonant onset" and "syllable offset" that were shorter than 175 ms. A native speaker of English listened to these stimuli. For n=109 whose durations were greater than 100 ms, it was determined that they were appropriate to include as stimuli, as the initial consonant and vowel were judged to be identifiable. The remainder, which were less than 100 ms long, were judged to be inappropriately brief. Approximately 95% of all tokens had durations that were greater than 175 ms. The majority of these were truncated to 175 ms, a duration at which it was judged that they contained enough of the following vowel for it to be perceptible to listeners. For the remaining that were longer than 175 ms, the stimuli were truncated at "syllable offset" if the resulting stimulus was 175 ms - 185 ms long. There were some cases where it was judged that the child prolonged the initial /r/ or /w/. When these were truncated using the 175 ms criterion, the resulting stimulus did not include enough of the vowel for it to be used as a stimulus in this experiment. The "consonant onset" mark in those productions was adjusted on a case-by-case basis so that the resulting stimulus was between 175 ms and 185 ms in length, and was judged by a native speaker of English to contain enough of the initial consonant and the vowel to be appropriate as a stimulus in this experiment.

There were a small number of productions with protracted initial consonants where it was not possible to readjust the markers to get a useable stimulus between 175 ms and 185 ms long; those productions were discarded. The finally selected and adjusted CV boundaries were later used in the VAS adult perceptual experiment.

Other Measures:

The child participants also completed additional measures of language, speech production, and speech perception at their initial visit. These are described in this section.

Speech perception. A minimal-pair discrimination task was used to measure speech perception. For this task, children were asked to discriminate between two words that differed by a single phoneme (e.g. *peas* vs. *keys*). Table 4 shows the 15 word pairs included on the minimal pairs task. All items were words that would be familiar to young children. The stimuli were recorded by two young adult female speakers, one in Mainstream American English (MAE) and the other in African American English (AAE). Children received stimuli in their native dialect. For each word pair, color photographs of two words were presented one at a time along with their object name presented auditorally. When the two images were presented together and one of the images was named, the child was asked to point to the picture that was named on a touch screen. Percent correct was computed automatically.

Table 4. Pairs of words used in the minimal pair tasks.

	Minimal Pair Sounds	Minimal Pair Words
1	/b/ - /k/	bee – key
2	/b/ - /p/	big – pig
3	/k/ - /dʒ/	car – jar
4	/tʃ/ - /k/	cheese – keys
5	/k/ - /h/	cold – hold
6	/g/ - /dʒ/	goose – juice
7	/h/ - /p/	hen – pen
8	/ɔr/ - /aʊ/	horse – house
9	/dʒ/ - /m/	juice – moose
10	/k/ - /p/	keys – peas
11	/u/ - /aʊ/	moose – mouse
12	/s/ - /θ/	mouse – mouth
13	/k/ - /t/	sick – sit
14	/sl/ - /sw/	sleep – sweep
15	/ɑr/ - /ɔr/	star – store

Acoustic robustness of contrast for /t/ vs. /k/ and psycho-acoustic robustness of contrast for /s/ vs. /ʃ/. At the initial data collection point, children participated in a picture-prompted auditory word repetition task to elicit productions of /t/, /k/, /s/, and /ʃ/ initial words. The procedure was identical to the procedure to elicit /r/ and /w/ initial words one year later. Tables 5 and 6 provide the stimuli for these two contrasts. Children's productions of these word-initial consonants were transcribed and analyzed acoustically.

Table 5. Target /t/- and /k/-initial words. All words were repeated twice.

Vowel Quadrants	/t/	/k/
Front – High	*tickle* *table* *tape*	*kitty* *kitchen* *cake*
Front – Non-high	*teddy bear*	*candy* *cat*
Back – High	*tooth* *toast*	*cookie* *coat*
Back – Non-high	*tummy* *tongue*	*cup* *car*

Table 6. Target /s/- and /ʃ/-initial words. (*) indicates words were repeated twice. (+) indicates words were repeated four times.

Vowel Quadrants	/s/	/ʃ/
Front – High	*sick** *scissors**	*sheep+*
Front – Non-high	*sandwich** *sad**	*share+*
Back – High	*soup** *soap**	*shoe+*
Back – Non-high	*sun** *sock**	*shovel** *shower**

For the /t/ and /k/ contrast, only tokens that were produced as alveolar or velar stops (or at an intermediate place of articulation) were included in the analysis. As noted above, Bernstein (2015) used adult perceptual judgments with VAS to develop a robustness of contrast measure for the development of the /t/ vs. /k/ contrast. Adults listened to consonant-vowel sequences with initial /t/ and /k/ targets and rated the productions on a VAS with endpoints labeled as "correct 't'" and "correct 'k'". A mixed-effects regression with random slopes for subjects was run across all participants. The dependent variable was the VAS judgments and the independent variable was consonant type. The child-level random slopes were the robustness of contrast measure. This perceptual robustness of contrast measure was available for 59 of the 105 participants.

An acoustic robustness of contrast measure for the acquisition of the /t/ vs. /k/ contrast was available for 94 of the 105 participants. For this measure, the centroid of the fricative spectrum was used as the summary acoustic measure. The acoustic centroid is higher for /t/ than for /k/ because /t/ has a more front place of articulation and therefore a smaller front cavity. Only /t/ and /k/ before back vowels were included in the analysis, because centroid is a more reliable predictor of the place of articulation contrast in a back vowel context, and because /k/ has a more front position (more similar to /t/) in a front vowel context. The process for computing the centroid is described in detail in Johnson (2016). After the centroid had been computed for all productions, a mixed-effects regression with random slopes for subjects was run across all participants. The dependent variable was centroid and the independent variable was consonant type. The robustness of contrast measures was the child-level random slopes.

For the /s/ - /ʃ/ contrast, only tokens that were produced as sibilant fricatives were included in the analysis. The fricative noise spectrum was isolated for analysis. Peak ERB, a psychoacoustic spectral peak measure that represents the point of highest amplitude frequency (Moore, Glasberg, & Baer, 1997), was used as the psycho-acoustic measure to differentiate /s/ from /ʃ/. A mixed-effects regression with random slopes for subjects was run across all participants. The dependent variable was the peak ERB value and the independent variable was consonant type. The child-level slopes were the psycho-acoustic robustness of contrast measure for the acquisition of the /s/ vs. /ʃ/ contrast and were available for 76 of the 105 participants in this study. See Reidy (2015) for more details of this analysis and Figure 5 for an example of the /s/ - /ʃ/ measure.

Table 7 summarizes the robustness of contrast measures for each contrast along with the number of participants that were analyzed. There were two reasons for the discrepancies among the number of participant for various measures. For the perceptual robustness of contrast measures, the /t/ vs. /k/ VAS study was run with productions from a subset of the child participants from the /r/ vs. /w/ study. For the acoustic and psycho-acoustic robustness of contrast measures, the analyses could only be done if children produced voiceless lingual stops (for the /t/ vs. /k/ measure) or sibilant fricatives or affricates (for the /s/ vs. /ʃ/ measure). The productions of children who did not produce these sounds correctly or who produced other substitutions could not be included in these analyses.

Table 7. Summary of robustness of contrast measures.

Contrast	Perceptual robustness of contrast (child-level random slopes model with VAS ratings as dependent variable)	Acoustic robustness of contrast (child-level random slopes from model with centroid (stops) or peak ERB (fricatives) as dependent variable)
/r/ vs. /w/ contrast (at time point 2)	N = 105	Not available
/t/ vs. /k/ contrast (at time point 1)	N = 59	N = 94
/s/ vs. /ʃ/ contrast (at time point 1)	Not available	N = 76

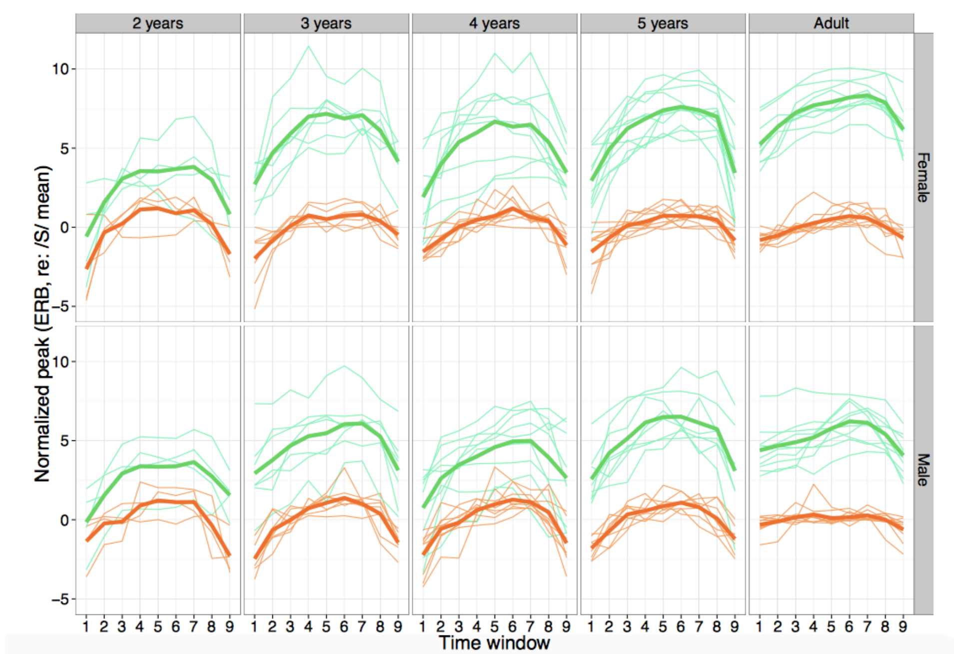

Figure 5. Peak ERB measure for /s/ - /ʃ/ (Reidy, 2014). The mean peak ERB trajectories for /s/ (green) and /ʃ/ (orange) is shown as thick, dark path. Each participant's mean peak ERB is shown as light, thin path.

<u>*Articulation*.</u> Children were administered the *Goldman Fristoe Test of Articulation* (GFTA-2, Goldman & Fristoe, 2000). The GFTA-2 is a norm-referenced test designed to measure children's ability to produce consonants in initial, medial, and final position and some initial clusters at the word level. Responses were recorded and transcribed by trained phoneticians. The standard score was calculated based on the number of errors.

<u>*Vocabulary size*</u>. The *Peabody Picture Vocabulary Test, 4th Edition* (PPVT-4, Dunn & Dunn, 2007) and the *Expressive Vocabulary Test, 2nd Edition* (EVT-2, Williams, 2007) were administered to all children to assess receptive and expressive vocabulary size, respectively.

ADULT SPEECH PERCEPTION

***Participants*:**

Participants were 31 males and 45 females with descriptive information shown in Table 8. They were recruited from Madison, Wisconsin and Minneapolis, Minnesota through fliers on campus, in-class announcement to undergraduate students in the Department of Communication Sciences and Disorders, and word of mouth. Participants met the following criteria: 1) 18 years old or older; 2) native speakers of North American English, 3) monolingual or elementary proficiency in another language; and 4) no history of speech, language, or cognitive impairment according to self-report.

***Procedure*:**

<u>*Stimuli*.</u> The stimuli were a total of 2,220 word-initial /r/ and /w/ spoken by children in the earlier study from the picture-prompted auditory real word repetition task. There were roughly 20 listeners in each experiment version (A, B, C, D). The total number of stimuli was evenly divided to 550 stimuli in each version and the distribution of the four transcription

categories for the two consonants was roughly equivalent across the four versions of the experiment.

Table 8. Information of adult participants.

Gender	Number of listeners	Mean Age in years (SD, range)
Male	31	22.57 (5.06, 18 – 45)
Female	45	21.26 (3.83, 18 – 32)

Procedure. Each listener was tested in a quiet room. The stimuli were played through Sennheiser HD 280 Pro circumaural headphones while “Listen” appeared on the computer screen. After each stimulus was played, a visual analogue scale (VAS) appeared on the computer screen. Dell LATITUDE E7240 laptops were used across Wisconsin and Minnesota to ensure the procedure was exactly the same for all participants, including the screen size and the length of the visual analogue scale. The length of the visual analogue scale was set to 200 mm to fit suitably within the computer screen. Moreover, a VAS-200 mm scale has been found to be sensitive enough to detect response differences (Svensson, 2000; Cowley & Youngblood, 2009). Two labels appeared on the screen each at the end of the VAS, “the /w/ sound” on the left side and “the /r/ sound” on the right (see Figure 6). Listeners were instructed to click on the place on the line that represented how close the sound was to an “r” or a “w”. The listeners were encouraged to use the entire line. Participants completed five familiarization trials prior to the experiment. The stimuli were presented in randomized order to each listener. Although the responses were not timed, participants were informally directed not to spend too much time on any single stimulus item.

Figure 6. VAS presented in the perception task.

***Analysis*:**

The adult perceptual judgments were used to develop a perceptual robustness of contrast measure for the /r/-/w/ contrast for each child. The robustness of contrast measure was originally adopted by Holliday et al. (2015), where the original target contrast was /s/-/ʃ/ and the predictor was psycho-acoustic robustness of contrast as measured by the peak ERB. This measure was extended to VAS listener ratings of /t/ vs. /k/ in Bernstein (2015), and this prompted the use of adult VAS listener ratings of /r/ vs. /w/ as a continuous predictor in this study. A perceptual measure rather than an acoustic measure was chosen because it was not possible to accurately measure formant frequencies of the second and third formants in the children’s productions. An example of a child’s production of /r/ is shown in Figure 7. It can be observed that the third formant frequency is poorly defined, probably because it is in word-initial position.

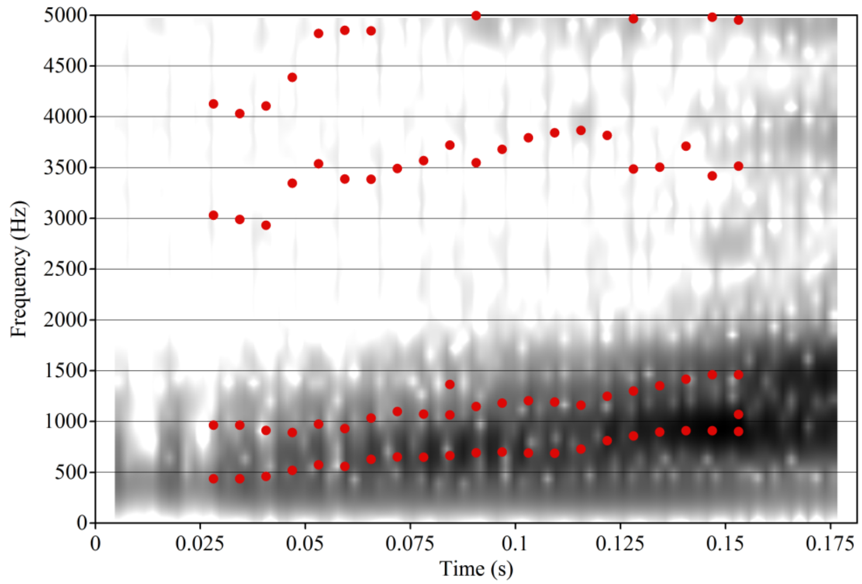

Figure 7. Formant frequencies of a word-initial /r/ produced by a child participant.

The pixel locations of the listeners' responses on the VAS task were converted into values ranging from 0 to 1.0. The responses for each item for each child were across the twenty listeners. As in Bernstein (2015), a mixed-effects regression with random slopes for subjects was run across all participants. The dependent variable was the VAS judgments and the independent variable was consonant type. The child-level slopes were then used to determine how many productions could be correctly classified for each child. The robustness of contrast measure was defined as the percent of productions that were correctly classified for each child.

A steeper slope indicates a more robust contrast between /r/ and /w/, and a shallower slope indicates the opposite.

CHAPTER THREE

Results

Question 1: Is the perceptual robustness of contrast measure for /r/ vs. /w/ valid and informative?

Transcription accuracy. As noted above, children's speech production of /r/-/w/ were transcribed. Productions were transcribed as correct or incorrect. Incorrect productions were transcribed as clear substitutions (e.g. [w] for /r/) or as intermediate ([r:w] if the production was in between /r/ and /w/ but more like /r/, or [w:r] if the production was in between /r/ and /w/ but more like /w/). Table 9 shows the means for the different transcription categories for /r/ vs. /w/ across the 105 child participants. It can be observed that /w/ was produced much more accurately than /r/; /r/ is produced with about 26% accuracy while /w/ is produced with about 83 percent accuracy. Only 25 percent of the children produced /r/ correctly, although this number rises to almost 50 percent if the productions that are intermediate but more like /r/ are included. These results are consistent with previous research on the acquisition of the /r/ vs. /w/ contrast (Hoffman et al., 1983; Sharf & Ohde, 1984; Ohde & Sharf, 1988).

Table 9. Number of tokens for the four transcription categories for target "r" and "w" across the four versions of the experiment.

Transcribed:	Target: r	Target: w
[r]	284 (25.56%)	3 (0.27%)
[r:w]	233 (20.97%)	24 (2.16%)
[w:r]	426 (38.34%)	166 (14.97%)
[w]	168 (15.13%)	916 (82.60%)
Total	**1,111 (100%)**	**1,109 (100%)**

<u>*VAS ratings vs. transcription accuracy*</u>**.** Mean VAS ratings by transcription category are shown in Figure 8. A mixed effects model ANOVA was used to examine the relationship between the VAS listener ratings and transcribed accuracy. The within-subjects factor was the six transcription categories, and the between-subjects factor was the four experimental versions (A, B, C, D) of the VAS task. The dependent variable was the VAS listening ratings (as described above, the pixel location of each response was converted into values ranging from 0 for /w/ to 1.0 for /r/. There was a significant effect of transcription category on VAS rating, $F[5, 444] = 202.5$, $p < 0.001$. The effect of experiment version was not significant. Bonferroni-corrected post-hoc paired comparisons were made to compare each transcription category to each other category. All pairwise differences were significant at the $p < 0.001$ level, except for comparisons between correct productions and clear substitutions ([r] for /r/ vs. [r] for /w/, and [w] for /w/ vs. [w] for /r/). There were only 3 tokens that were [r] for /w/ substitutions, so it is not surprising that there were no differences between correct /r/ productions and [r] for /w/ substitutions. In previous research (Munson, Edwards, Schellinger, Beckman, & Meyer, 2010), a difference between correct productions and clear substitutions was observed for correct [ʃ] productions vs. [ʃ] for /s/ substitutions. It is unclear why this was not observed for correct [w] productions vs. [w] for /r/ substitutions, but it might have been because the transcriber used the intermediate categories extensively, as observed in Table 9.

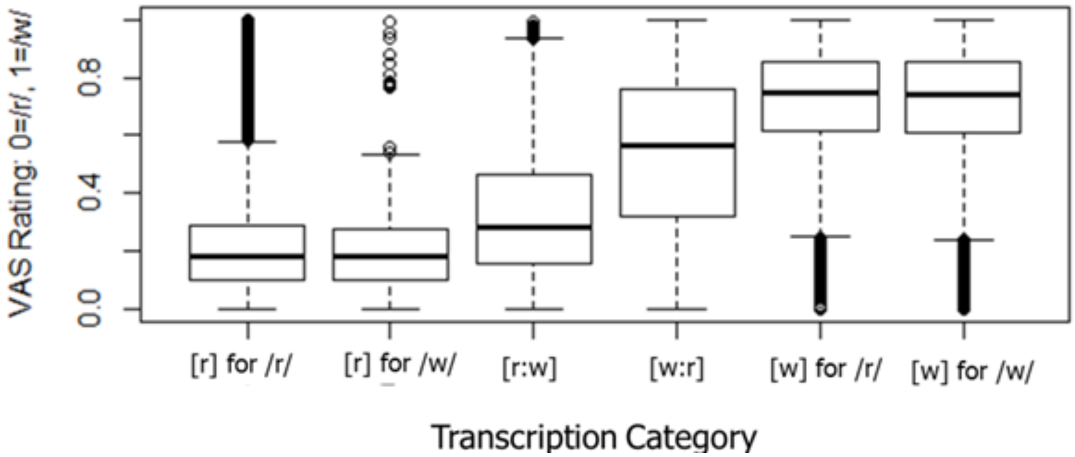

Figure 8. Boxplot of the relationship between transcription category and VAS rating.

<u>*Perceptual robustness of contrast*</u>. As noted above, VAS ratings were used in a mixed-effects logistic regression and child-level random slopes were extracted. Robustness of contrast was quantified as the child-level random slopes. Figures 9 and 10 show the slope from the logistic regression analysis and box plots of the VAS ratings for two participants, one (Figure 9) with high transcription accuracy (100%) and a relatively steep slope, and another (Figure 10) with lower transcription accuracy (74%) and a relatively shallow slope.

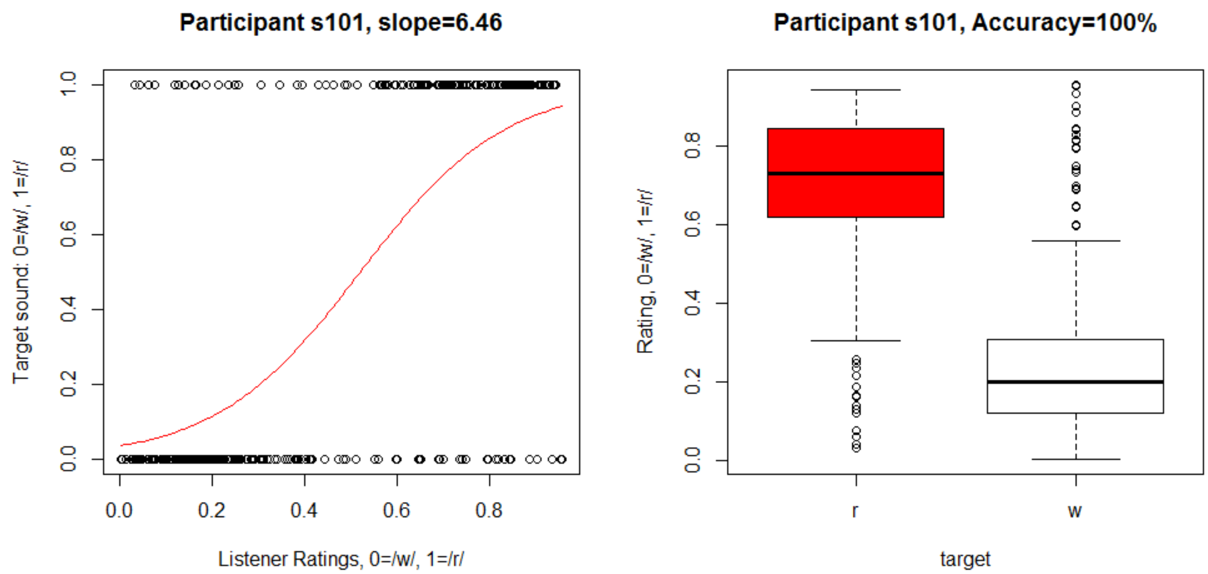

Figure 9. Subject-level results of logistic regression (left plot) and box plot of VAS ratings (right plot) for a child with high transcription accuracy.

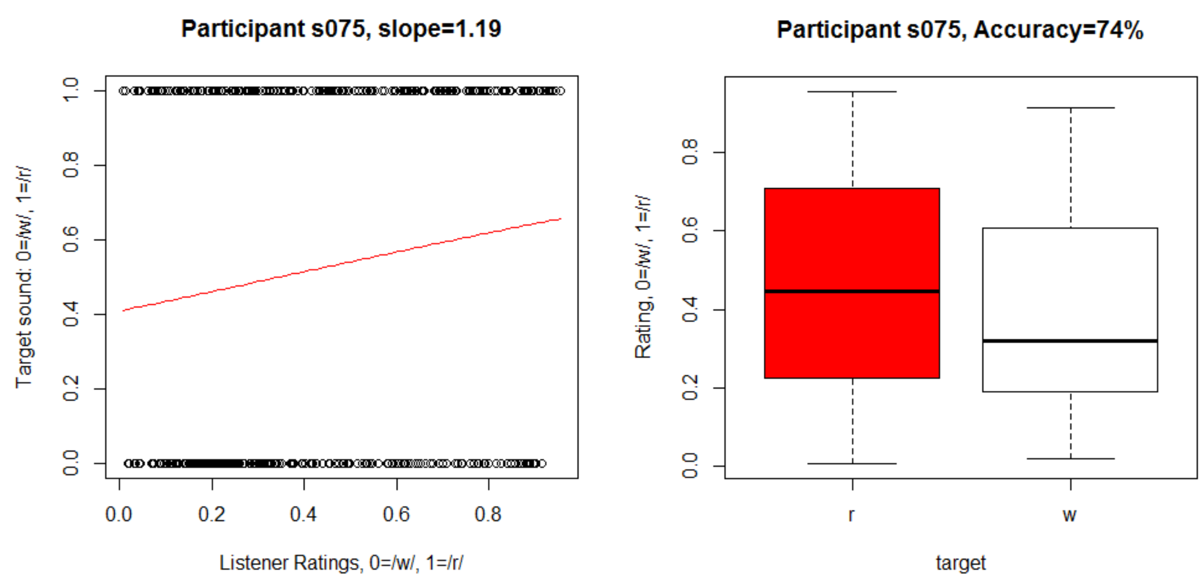

Figure 10. Subject-level results of logistic regression (left plot) and box plot of VAS ratings (right plot) for a child with low transcription accuracy.

For each child, an average accuracy score and an average robustness of contrast score were computed. The correlation between these two measures was .92 ($p < .001$). Figure 11 shows the relation between these two measures. A strong relationship can be observed between the two measures, but there is a greater spread for VAS ratings for participants with 100% transcription accuracy, suggesting that VAS ratings may be a more fine-grained measure than transcription accuracy.

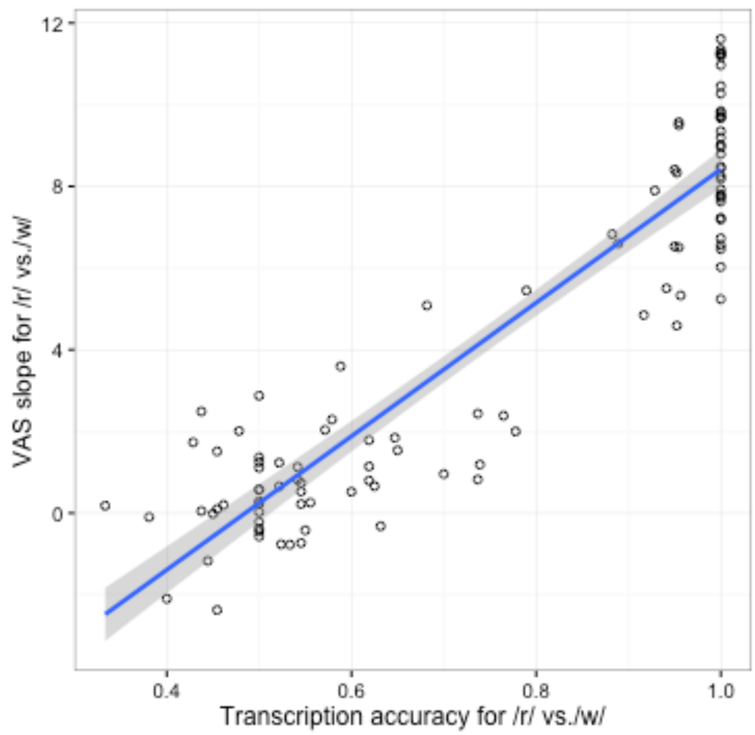

Figure 11. Correlation between transcribed accuracy and individual-subject VAS slopes.

Question 2: What measures at 2 ½ to 3 years of age predict the acquisition of the contrast between /r/ and /w/ at 3 ½ to 4 years of age?

The second question addressed in this study was which measures at age 2 ½ to 3 years predicted children's acquisition of the /r/ vs. /w/ contrast at 3 ½ to 4 years. Two linear regression models were run to address this question. In model 1, the dependent variable was the perceptual robustness of contrast measure for /r/ vs. /w/. The independent variables were: (a) speech perception ability, as quantified by performance on the minimal pairs task (percent correct); (b) acquisition of an early-acquired contrast (/t/ vs. /k/), as quantified by both a perceptual robustness of contrast measure and an acoustic robustness of contrast measure for /t/ vs. /k/; (c) acquisition of a later-acquired contrast (/s/ vs. /ʃ/), as quantified by the psycho-acoustic robustness of contrast measure for /s/ vs. /ʃ/; (d) general articulatory ability, as quantified by the standard score on the GFTA-2; (e) receptive and expressive vocabulary size, as quantified by standard scores on the PPVT-4 and the EVT-2, respectively. In model 2, the dependent variable was transcription accuracy for the /r/ vs. /w/ contrast. The independent variables were: (a) speech perception ability, as quantified by performance on the minimal pairs task (percent correct); (b) acquisition of an early-acquired contrast (/t/ vs. /k/), as quantified by transcription accuracy for the /t/ vs. /k/ contrast; (c) acquisition of a later-acquired contrast (/s/-/ʃ), as quantified by transcription accuracy for the /s/-/ʃ/ contrast; (d) general articulatory ability, as quantified by the standard score on the GFTA-2; (e) receptive and expressive vocabulary size, as quantified by standard scores on the PPVT-4 and the EVT-2. For both models, the only significant predictors were GFTA standard score (*Estimate* = .12, *standard error* = .03, $t = 3.68$, $p < .001$ for model 1; and *Estimate* = .007, *standard error* = .002, $t = 3.61$, $p < .001$ for model 2) and PPVT-2 standard score (*Estimate* = .08, *standard error* = .02, $t = 3.47$, $p < .001$ for model 1; and *Estimate* = .004, *standard error* = .001, $t = 2.56$, $p = .012$ for model 2). Together they predicted 40% (model 1) and 34% (model 2) of the variance in the dependent variable. No other measures were significant predictors. Figures 12 and 13 show the relation between GFTA-2 and PPVT-4 standard scores at age 2 ½ to 3 years and the two measures of the acquisition of the /r/ vs. /w/ contrast at age 3 ½ to 4 years for the two models.

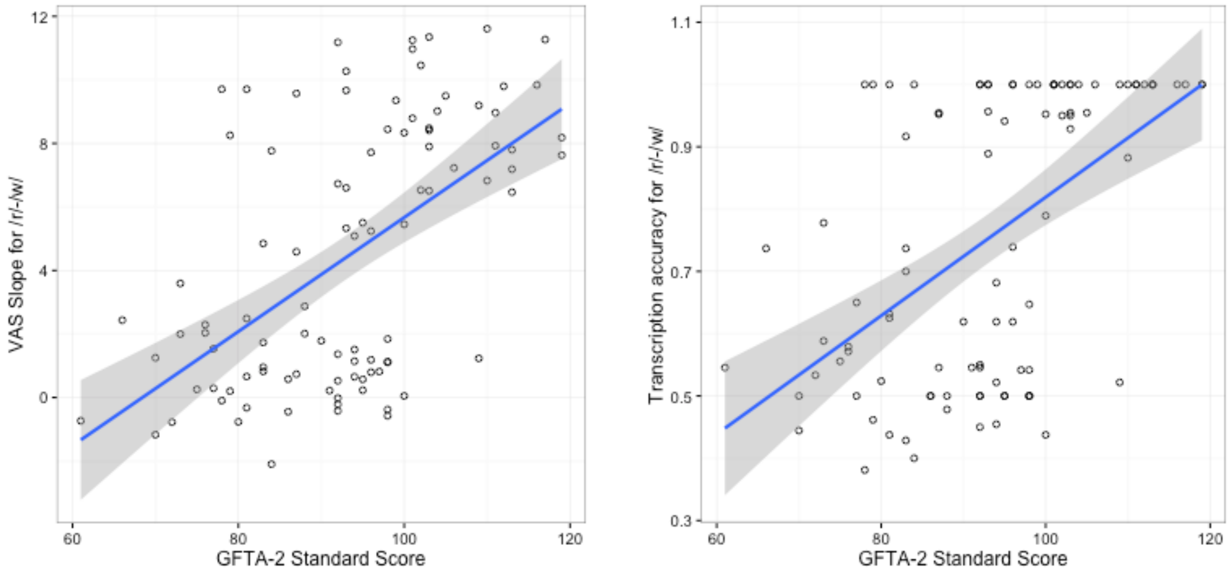

Figure 12. Relation between articulatory accuracy and two measures of the acquisition of the /r/ vs. /w/ contrast (perceptual robustness of contrast [left plot] and transcription accuracy [right plot]).

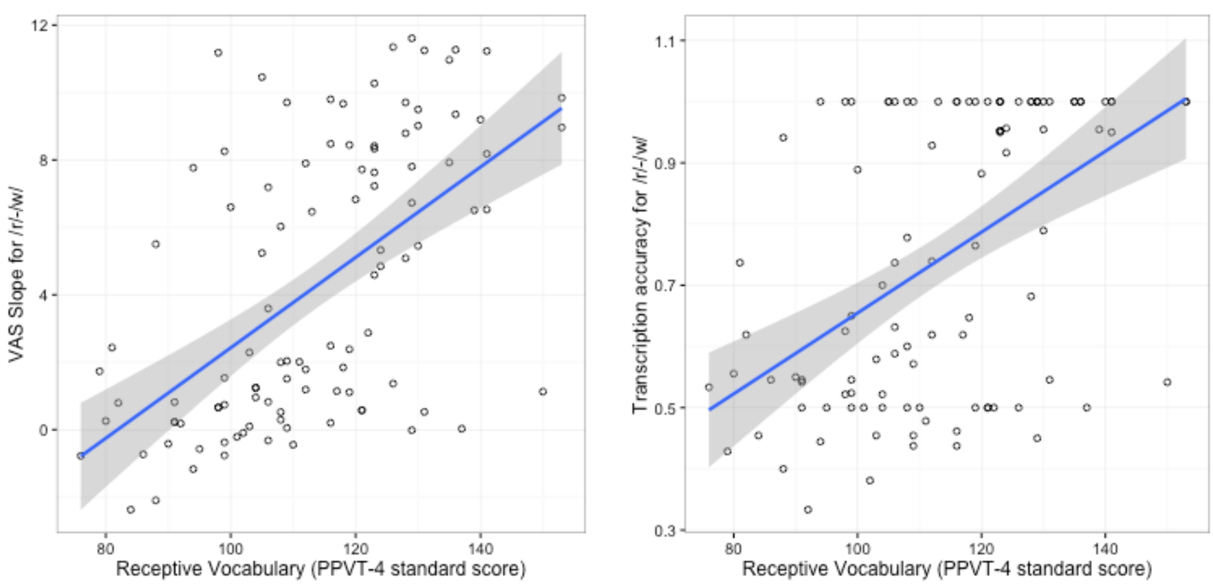

Figure 13. Relation between receptive vocabulary size and two measures of the acquisition of the /r/ vs. /w/ contrast (perceptual robustness of contrast [left plot] and transcription accuracy [right plot]).

Post hoc analysis: Relationships among accuracy measures and robustness of contrast measures across different contrasts.

Contrary to what had been predicted, the fine-grained measures of articulatory accuracy at the younger age were not significant predictors of the acquisition of the /r/-/w/ contrast. A series of correlations were run to further examine the relationships among these different measures. First the three transcription accuracy measures were correlated with each other and with the GFTA-2 standard score. The results are shown in Table 10. It can be observed that all pair-wise correlations are significant, although some of the correlations are relatively small. It should be noted that transcription accuracy for all three contrasts is more strongly correlated with GFTA-2 standard score, then these transcription accuracy measures are correlated with each other.

Table 10. Pair-wise correlations among all transcription accuracy measures and GFTA-2 standard scores (Time Point 1 (TP1) = at age 2 ½ to 3 years; Time Point 2 (TP2) = at age 3 ½ to 4 years). Significant correlations are in bold font.

	Transcription accuracy for /r/ vs. /w/ at time point 2	Transcription accuracy for /s/ vs. /ʃ/ at time point 1	GFTA-2 standard score at time point 1
Transcription accuracy for /r/ vs. /w/ at time point 2	—	—	**r = .53, p < .001 (n = 96)**
Transcription accuracy for /s/ vs. /ʃ/ at time point 1	**r = .27, p = .005 (n = 102)**	—	**r = .57, df = 92, p < .001 (n = 94)**
Transcription accuracy for /t/ vs. /k/ at time point 1	**r = .24, p = .014 (n = 101)**	**r = .44, p < .001 (n = 100)**	**r = .43, p < .001 (n = 92)**

The robustness of contrast measures were also correlated with each other. The measures included in this analysis were: perceptual robustness of contrast for /r/ vs. /w/, perceptual robustness of contrast for /t/ vs. /k/, acoustic robustness of contrast for /t/ vs. /k/, and psycho-acoustic robustness of contrast for /s/ vs. /ʃ/. GFTA-2 standard scores were also included in the series of correlations. The results are shown in Table 11. It can be observed that many of these correlations are non-significant. Even among the significant correlations, it can be observed that – with two exceptions (the correlation between the perceptual robustness of contrast measure for /r/ vs. /w/ and GFTA-2 standard score, and the correlations between the perceptual robustness of contrast measures for /r/ vs. /w/ and /t/ vs. /k/) – these correlations are generally lower than the comparable comparisons in Table 10 above.

Table 11. Pair-wise correlations among all perceptual robustness of contrast measures and GFTA-2 standard scores (Time Point 1 (TP1) = at age 2 ½ to 3 years; Time Point 2 (TP2) = at age 3 ½ to 4 years). Significant correlations are in bold font.

	Perceptual robustness of contrast for /r/ vs. /w/ at time point 2	Psycho-acoustic robustness of contrast for /s/ vs. /ʃ/ at time point 1	GFTA-2 standard score at time point 1
Perceptual robustness of contrast for /r/ vs. /w/ at time point 2	—	—	**r = .56, df = 94, p < .001 (n = 96)**
Psycho-acoustic robustness of contrast for /s/ vs. /ʃ/ at time point 1	r = .21, p = .07 (n = 76)	—	**r = .25, p = .04 (n = 73)**
Acoustic robustness of contrast for /t/ vs. /k/ at time point 1	r = .40, p = .70 (n = 92)	r = .15, p = .21 (n = 69)	r = .06, p = .56 (n = 85)
Perceptual robustness of contrast for /t/ vs. /k/ at time point 1	**r = .35, p = .006 (n = 59)**	—	**r = .38, p = .004 (n = 56)**

CHAPTER FOUR

Discussion

This study addresses two questions. The first was whether a gradient robustness of contrast measure for acquisition of the /r/ vs. /w/ contrast based on perceptual ratings in a VAS task could be developed. The answer to this is a qualified "yes." On the positive side, the VAS ratings reflected the transcription categories. Listeners were able to differentiate among four different categories: [r], [r:w], [w:r], and [w]. Furthermore, transcribed accuracy and the robustness of contrast measure (child-level slopes from the logistic regression analysis) were significantly correlated. The fact that there was a range of the VAS ratings for listeners who had been transcribed as 100% accurate suggests that VAS ratings may be more sensitive than transcription accuracy, at least among individuals with high accuracy levels. This finding is consistent with the results of Bernstein (2015) and Munson et al. (2010). However, the low correlations among the different robustness of contrast measures and the higher correlations among the different transcription accuracy measures suggests that further research is needed to understand and refine these robustness of contrast ratings. While VAS would help clinicians monitor the nature of gradual change in speech-sound acquisition and assess child's speech progress in detail, more research is needed before VAS ratings can be used as a clinical tool.

The study also examined whether measures of speech perception, speech production, and vocabulary size at age 2 ½ to 3 predict children's acquisition of the /r/-/w/ contrast at age 3 ½ to 4 years. The best predictor of the contrast was general articulatory ability, as measured by standard score on the GFTA-2. This result is not surprising. A number of studies have found that children with more speech sound errors at a young age are at risk for continuing to have speech sound errors at a later age (Preston & Edwards, 2010; Preston, Hull, & Edwards, 2013). The more fine-grained measures of children's speech production abilities, the robustness of contrast measures for /t/ vs. /k/ and /s/ vs. /ʃ/ were not significant predictors of the acquisition of the /r/ vs. /w/ contrast. At this point, there are a number of reasons for this finding. First, it could be due to a problem with how acquisition of these contrasts has been quantified. However, the relationship between acquisition of /s/ vs. /ʃ/ and /t/ vs. /k/ was also not observed in the model that only included measures of transcription accuracy. This finding could be related to the fact that the speech motor control involved in producing the /r/ vs. /w/ contrast is very different from the motor control involved in producing the /s/ vs. /ʃ/ and the /t/ vs. /k/ contrasts.

It was also the case that speech perception skills, as quantified by performance on the minimal pairs task was not a significant predictor of the acquisition of the /r/ vs. /w/ contrast. This finding differs from previous literature (Rvachew, 1994; Rvachew et al., 2004). The findings of Rvachew and colleagues focused primarily on acquisition of the /s/ vs. /ʃ/ contrast. It is possible that perception of the /r/ vs. /w/ contrast is not as important for speech production or it may be that the minimal pairs task was not a challenging enough task of children's speech perception abilities. Furthermore, Rvachew and colleague's findings were from children with phonological disorders, so another possibility is that speech perception training facilitates speech production only if children have particular difficulty with the perception of specific target sounds that they produce incorrectly.

Receptive vocabulary skills, as measured by PPVT-4 standard scores, were also a significantly predictor of the acquisition of the /r/ vs. /w/ contrast. It is unclear why receptive vocabulary is associated with articulatory proficiency, although this result has been observed by other research (Munson et al., 2005; Rvachew & Grawburg, 2006). Further research is needed to understand this relationship.

There are a number of limitations of this study. First, there was a relatively small sample size (16 to 20 adult participants) in each perception experiment. Second, the inter-rater reliability for the VAS listener ratings for the /r/ vs. /w/ contrast was lower than desirable. The mean inter-rater reliability was 0.69 (median=0.71) when ten percent of the stimuli were repeated across experiments. Intra-rater reliability for the VAS task should also have been examined. Third, there were unequal numbers of participants across different measures, so it was difficult to determine whether some of the differences in the correlation values were related to differences in power, to differences due to the fact that the same participants were not included in all of the analyses, or to differences in the measures themselves. Fourth, the number of child participants from families with different maternal education were unevenly distributed. The number of children from families with high maternal education (i.e., college or graduate degrees) was greater than the number of children who come from families with middle (i.e., trade school, technical/associate's degree) and low maternal education (i.e., less than high school, high school diploma, GED) combined. Maternal education is a strong correlate of children's speech and language development, so it is unknown whether similar results would have been observed if participants' speech production ability and other speech-language related measures more faithfully represented the true population mean.

CHAPTER FIVE

Conclusion

To conclude, this study found that VAS listener ratings of the /r/ vs. /w/ contrast were highly correlated with transcription accuracy. The study also found that general articulatory ability and receptive vocabulary size at age 2 ½ to 3 years predicted children's acquisition of the /r/ vs. /w/ contrast one year later. These results are consistent with previous literature. Future research is needed to determine whether VAS listener ratings could be clinically useful and to understand the nature of the relationship between vocabulary size and speech production ability.

REFERENCES

American Speech-Language-Hearing Association. (2001). *Speech Sound Disorders: Articulation and Phonological Processes.* Available from http://www.asha.org/public/speech/disorders/SpeechSoundDisorders.

Bernstein, S. (2015). Individual differences in the acquisition of the /t/ - /k/ contrast: A study of adult's perception of children's speech. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/174721.

Boersma, P., & Weenink, D. (1992). *Praat, a system for doing phonetics by computer.* Web site: www.praat.org.

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., Sabo, D. L., & Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development, 74*(2), 346-357.

Cowley, J. A., & Youngblood, H. (2009). Subjective Response Differences between Visual Analogue, Ordinal and Hybrid Response Scales. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(25), 1883-1887.

Dalston, R. M. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America, 57*(2), 462-467.

Delattre, P. C., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics, 44,* 29-68.

Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test (4th ed.). Minnesota: Pearson.

Edwards, J., Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics, 22*(12), 937-956.

Edwards, J., Gross, M., Chen, J., MacDonald, M. C., Kaplan, D., Brown, M., & Seidenberg, M. S. (2014). Dialect Awareness and Lexical Comprehension of Mainstream American English in African American English-Speaking Children. *Journal of Speech, Language, and Hearing Research, 57*(5), 1883-1895.

Wilson, C. Y. (1992). Acoustic Measures for linguistic features distinguishing the semivowels in American English. *The Journal of the Acoustics Society of America, 92,* 736-757.

Goldman, R., & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation (2nd ed.). Minnesota: Pearson.

Hagiwara, R. (1995). Acoustic realization of American /r/ as produced by women and men. *UCLA Phonetics Laboratory Working Papers,* 90.

Harrison, L. J., & McLeod, S. (2010). Risk and protective factors associated with speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech, Language, and Hearing Research, 53*(2), 508-529.

Hewlett, N., & Waters, D. (2004). Gradient change in the acquisition of phonology. *Clinical Linguistics & Phonetics, 18*(6-8), 523-533.

Hoffman, P. R., Bengoa, D., & Schuckers, G. H. (1985). Misarticulating and Normally Articulating Children's Identification and Discrimination of Synthetic [r] and [w]. *Journal of Speech and Hearing Disorders, 50,* 46-53.

Hoffman, P. R., Stager, S., & Daniloff, R. G. (1983). Perception and Production of Misarticulated /r/. *Journal of Speech and Hearing Disorders, 48,* 210-215.

Holliday, J. J., Reidy, P. F., Beckman, M. E., Edwards, J. (2015). Quantifying the Robustness of the English Sibilant Fricative Contrast in Children. *Journal of Speech, Language, and Hearing Research, 58,* 622-637.

Jamieson, D. G., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology, 16*(3), 201-210.

Johnson, A. (2016, in preparation). *Child-Level Factors & Acquisition of the /t/-/k/ Contrast* (Master's thesis). University of Wisconsin-Madison, Madison, WI.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (eds.), *Phonological development: models, research, implications,* pp. 65-90. Parkton, MD: York Press.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105*(3), 1455-1468.

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics, 37*(1), 111-124.

Locke, J. L. (1980). The inference of speech perception in the phonologically disordered child. Part II: Some clinically novel procedures, their use, some findings. *Journal of Speech and Hearing Disorders, 45,* 445-468.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word initial stop consonants. *Journal of Child Language, 7,* 41–74.

McAllister Byun, T., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex Versus Bunched in Treatment for Rhotic Misarticulation: Evidence From Ultrasound Biofeedback Intervention. *Journal of Speech, Language, and Hearing Research, 57*(6), 2116-2130.

Maxwell, E. M., & Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of misarticulating child's voice contrast for stops. *Applied Psycholinguistics, 3,* 29-43.

McCormack, J., McLeod, S., McAllister, L. & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology, 11*(2), 155-170.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A Model for the Prediction of Thresholds, Loudness, and Partial Loudness. *Journal of the Audio Engineering Society, 45*(4), 224-240.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships Between Nonword Repetition Accuracy and Other Measures of Linguistic Development in Children With Phonological Disorders. *Journal of Speech, Language, and Hearing Research, 48,* 61-78.

Munson, B., Edwards, J., Schellinger, S.K., Beckman, M.E., & Meyer, M.K. (2010). Deconstructing Phonetic Transcription: Covert Contrast, Perceptual Bias, and an Extraterrestrial View of Vox Humana. *Clinical Linguistics and Phonetics, 24*, 245-260.

Munson, B., Schellinger, S.K., & Urberg-Carlson, K. (2012). Measuring speech-sound learning using visual analog scaling. *Perspectives in Language Learning and Education, 19,* 19-30.

National Institute on Deafness and Other Communication Disorders. (2010). *Statistics on Voice, Speech and Language.* Available from http://www.nidcd.nih.gov/health/statistics/Pages/vsl.aspx.

Ohde, R. N., & Sharf, D. J. (1987). Effect of Formant Transition Rate on the Differentiation of Synthesized Child and Adult /w/ and /r/ Sounds. *Journal of Speech and Hearing Research, 30,* 215-222.

Ohde, R. N., & Sharf, D. J. (1988). Perceptual Categorization and Consistency of Synthesized /r-w/ Continua by Adults, Normal Children and /r/-Misarticulating Children. *Journal of Speech and Hearing Research, 31,* 556-568.

Preston, J. L., & Edwards, M.L. (2010). Phonological Awareness and Types of Sound Errors in Preschoolers With Speech Sound Disorders. *Journal of Speech, Language, and Hearing Research, 53,* 44-60.

Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool Speech Error Patterns Predict Articulation and Phonological Awareness Outcomes in Children With Histories of Speech Sound Disorders. *American Journal of Speech-Language Pathology, 22,* 173-184.

Reidy, P. F. (2014). *Moving targets and unsteady states: "Shifting" productions of sibilant fricatives by young children.* Paper presented at the 168th Meeting of the Acoustic Society of America, Indianapolis, IN.

Reidy, P. F. (2015). A comparison of spectral estimation methods for the analysis of sibilant fricatives. *Journal of the Acoustical Society of America, 137*(4), 248-254.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37,* 347-357.

Rvachew, S. (2006). Longitudinal Predictors of Implicit Phonological Awareness Skills. *American Journal of Speech-Language Pathology, 15*, 165-176.

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research, 49*(1), 74-87.

Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of Phonemic Perception Training on the Speech Production and Phonological Awareness Skills of Children With Expressive Phonological Delay. *American Journal of Speech Language Pathology*, *13*(3), 250-63.

Sharf, D. J., & Ohde, R. N. (1983). Perception of Distorted "R" Sounds in the Synthesized Speech of Children and Adults. *Journal of Speech and Hearing Research, 26,* 516-524.

Sharf, D. J., & Ohde, R. N. (1984). Effect of Formant Frequency Onset Variation on the Differentiation of Synthesized /w/ and /r/ Sounds. *Journal of Speech and Hearing Research, 27,* 475-479.

Sherman, D., & Gieth., A. (1967). Speech Sound Discrimination and Articulation Skill. *American Journal of Speech, Language, and Hearing Research, 10,* 277-280.

Shriberg L. D. (2009). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In Paul R., and Flipsen P., (eds). *Speech Sound Disorders in Children*. San Diego, CA: Plural Publishing.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of Speech Delay in 6-Year-Old Children and Comorbidity With Language Impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1461-1481.

Slawinski, E. B., & Fitzgerald, L. K. (1998). Perceptual development of the categorization of the /ɹ-w/ contrast in normal children. *Journal of Phonetics, 26*(1), 27-43.

Smit, A. B., Hand. L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*(4), 779-798.

Snowling, M., Bishop, D. V. M., & Stothard, S. E. (1999). Is Preschool Language Impairment a Risk Factor for Dyslexia in Adolescence? *Journal of Child Psychology and Psychiatry, 41*(4), 587-600.

Svensson, E. (2000). Comparison of the quality of assessments using continuous and discrete ordinal rating scales. *Biometrical Journal, 42*(4), 417-434.

Tyler, A. A., Figurski, G. R., & Langsdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech and Hearing Research, 36,* 746-759.

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication, 26,* 203-226.

Williams, K. T. (2007). Expressive Vocabulary Test (2nd ed.) Minnesota: Pearson.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe., A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *Acoustical Society of America, 123,* 4466-