Accuracy judgments of children’s sibilant fricatives:

Relationship between acoustic measures and perceptual judgments

Haley Nicole Webb

Submitted under the supervision of Dr. Benjamin Munson to the University Honors Program at the University of Minnesota-Twin Cities in partial fulfillment of the requirements for the degree of Bachelor of Arts, *summa cum laude* in Speech-Language-Hearing Sciences.

Spring 2014

Acknowledgements

I would like to thank the Learning to Talk Project at the University of Minnesota for supporting the completion of this thesis. Learning to Talk is funded by the National Institute on Deaf and Other Communicative Disorders and the National Science Foundation (NIH Grant # 5R01DC002932-13). I especially would like to thank my advisor, Dr. Benjamin Munson, who has given me continual advice and support throughout this entire project and has instilled in me a passion for research and a continued curiosity in the field. I am also indebted to Sara Bernstein, a graduate MA-SLP student at the University of Minnesota who reliability-checked my transcriptions and additionally provided immense amount of advice and planning tips to make sure I could finish on time. Thank you to Dr. Mary Beckman and Dr. Jan Edwards, both of whom provided advice on the direction of the thesis. In addition, thank you to Maria Swora, project manager at the University of Minnesota, and all of the wonderful examiners and segmenters at both sites at the University of Minnesota and University of Wisconsin-Madison. Thank you to Hannele Nicholson and Rose Crooks for all their hard work turbulence tagging and training me in segmentation and turbulence tagging. Outside of the lab, I would like to thank my friends and roommates, Amelia Ino, Jennifer McLean and Rosie Aasen, for putting up with my late night writings and occasional groans of frustration. Finally, I would like to thank my parents, my sister, and my boyfriend for calming me during times of anxiety and encouraging me throughout the entire process.

Abstract

As children learn language and acquire the different phonemes of their language, they often produce intermediate productions that are perceptually similar to the "correct" production. For example, productions of /s/ and /ʃ/ in children often sound very similar. As children age, they become more intelligible. Researchers have attempted to capture children's acquisition of the contrast between these two sounds using different acoustical methods. Nicholson (2014) created a new acoustic model using Peak ERB and individual variables that captured the contrast of 2-year-old children. This study expanded on Nicholson's findings by adding a perceptual component via phonetic transcription. The author transcribed productions from 15 children used in Nicholson's study. Percent correct was calculated for each target phoneme and correlated with age, EVT, and discriminability (an acoustic measure of robustness of contrast from Nicholson, 2014). This study found significant relationship between accuracy and age and accuracy of [ʃ] and discriminability.

Keywords: accuracy judgments, sibilant fricatives, robustness of contrast

Table of Contents

Section 1: Introduction and Background 5

Section 2: Literature Review 6

2.1 Fricative acquisition as measured by transcription 6

Table A 7

Figure 1 9

2.2 Effects of Listeners' Perceptions 9

Figure 2 16

2.3 Robustness of Contrast Measures 16

Section 3: Methods 26

3.1 Participants 26

Table A 26

3.2 Stimuli 26

3.3 Segmentation and Turbulence Tagging 27

3.4 Transcription 27

Figure 3 28

3.5 Reliability 28

Section 4: Results 28

Figure 4 29

Section 5: Discussion 31

Section 6: Conclusion 33

References 35

Section 1: Introduction and Background

A long-standing question that has puzzled theorists is how children acquire language and speech? Early theorists believed in a behavioral model of language development that posits children mimic adult speech and receive positive reinforcement for correct productions. Olmsted, an advocate for this theory, argued that the most difficult part of speech is perception not production. However, criticisms of the lack of consideration for the articulatory effort spurred new theories such as Jakobson's structuralist model (as cited in Vihman, 1996, p. 16-19). Jakobson presented the idea that children, regardless of their linguistic input, produce the same phonemes in a rule-governed order. Because the order is rule-governed, the presence of harder sounds such as fricatives (e.g. [s] and [ʃ]) and velars (e.g. [k] and [g]) imply that a child has already acquired easier sounds such labials (e.g. [b], [m]) and dentals (e.g. [d] and [t]). Similarly, in vowels, there is a determined sequence of acquisition: [a], [i], [e], [u]. Jakobson's theory, according to Ferguson and Farwell (1975), does not account for individual differences between children and rejects the importance of babbling, both of which have since been proven to be influential in language development.

With the arrival of the computer, more statistically based theories of acquisition emerged. Stemberger proposed a connectionist model that states the probability a child produces a certain phonological pattern depends on the frequency of which they heard it. This theory takes into account the ease of articulation of certain phonemes but discounts frequency of perception (as argued by Vihman, 1996, p. 44). A more contemporary theory, proposed by Vihman and Croft (2007), is similar in the idea of frequency of perception, but provides a more developmental approach. Vihman and Croft suggest that children develop language through the construction of templates based on adult productions. The templates are phonological patterns that the child seems to favor, perhaps due to articulatory constraints.

The variety of different phonological development theories provide the rationale for studies that attempted to discover how children learn speech and how their speech differs from adults. Fricatives are interesting to study because they can serve as a test case of a child's development of acquisition more generally. Fricatives require greater articulatory demand and therefore are one of the last sounds to be mastered by children. Because of their articulatory difficulty, children often produce a continuum of productions of a given fricative (i.e. sounds that are intermediate between adult targets). These intermediate stages show the gradual mastery but are unfortunately, difficult to capture with the tool most commonly used to study acquisition, phonetic transcription. The following section is a literature review that describes previous studies attempts at perceptually and acoustically measuring children's productions of fricatives. The review is split into three sections: Fricative acquisition as measured by transcription, effects of listener's perceptions, and robustness of contrast measures.

Section 2: Literature Review

2.1 Fricative acquisition as measured by transcription

Fricatives are a manner of speech sounds characterized by narrow constrictions and turbulent airflow. In English, each phoneme in the fricative manner category can be produced at four different places of articulation: labiodental (/f/ and /v/), interdental (/θ/ and /ð/), alveolar (/s/ and /z/), and palatal (/ʃ/ and /ʒ/). The fricatives /s/ and /ʃ/ are of interest to this paper because they are both sibilant voiceless fricatives (Ali & Van der Spiegel, 2001). Sibilant fricatives differ from non-sibilant fricatives in their acoustic characteristics: they are louder and have a more defined concentration of energy than non-sibilant fricatives.

Smit, Hand, Freilinger, Bernthal, and Bird (1990) completed a study that established articulation age-norms for different speech sounds by English-acquiring children. The study was cross-sectional and had 10 age groups: 3;0, 3;6, 4;0, 4;6, 5;0, 5;6, 6;0, 7;0, 8;0, 9;0 (years; months). Participants completed a picture-response type task and trained clinicians transcribed each utterance. Transcribers used a narrow transcription check-list system based off the system used in Shriberg and Kent (1982). The transcription system in Smit et al. (1990) contained some modifications from Shriberg and Kent's in that it added three new traits: post-alveolar distortion category, qualifiers "light" and "dark" for phoneme /l/, and "nasal release" for word-final obstruents. Transcriptions were coded and entered into the computer. A response that was considered "marginal" was also counted as correct in the coding process.

Smit et al. (1990) found no relationship between total score (i.e. amount of acceptable responses) and either state, population density or parental education. The only demographic variables with a significant relationship to total score were age and sex. Researchers found that total score increased with age and generally girls did better than boys. The data provided a general trend of accuracy across categories – nasals, glides, and stops were very accurate while fricatives, affricates, and liquids were less accurate at the same age level. Smit et al. (1990) created a table that listed the percentage of correct responses by phoneme and age. The author has recreated the relevant information in a table below. For the purposes of this paper, the results for word-initial /s/ and /ʃ/ for 3-year olds are shown.

Table A

Phoneme	Age (3 years)	
	Females	Males
/s/	75%	48%

/ʃ/	68%	44%

For 3 year olds, /s/ was more accurate than /ʃ/ in both males and females. There was also increase in percentage correct over age in both genders as well. In the discussion of Smit et al. (1990), the researchers recommended that the recommended age for intervention if the child is still inaccurately producing the phoneme /s/ is 7 years old despite their results suggesting that children master the phoneme /s/ at age 9.

Although Smit et al. (1990) used a modified transcription system (a checklist) to increase inter-transcriber reliability; Edwards and Beckman (2008) critiqued the use of transcription by citing evidence from the literature. The critique posited that transcription is not capable of capturing covert contrast and is susceptible to native-language listener biases. Fricatives often exhibit this covert contrast, acoustically distinct but perceptually similar productions, when children are mastering the phoneme. Edwards and Beckman cautioned the use of transcription due to the listener biases but recognized the practicality in clinical settings. Therefore, the researchers suggested that if transcription is used, to include intermediate productions to provide additional information to transcriptions.

Another method of transcription that is becoming more widely used in research is the use of a Visual Analog Scale (VAS). The visual analog scale provides the listener a continuum on which they can select an utterance sounded more like, for example, /s/ and more like /ʃ/. Figure 1 below shows can example of a VAS. Studies such as Munson, Johnson, and Edwards (2012) used the VAS because they wanted a system to represent children's intermediate productions and that is more reliable than IPA transcription system. The next section will begin with an overview of the Munson et al. (2012) study.

Figure 1

2.2 Effects of Listeners' Perceptions

Munson, Johnson and Edwards (2012) explored how changes in experience in the listener effects perceptions of children's productions. In this experiment, adults' ratings were compared to trained clinical speech-language pathologists to determine if clinical experience affects inter-rater reliability, sound differentiation, and correlation with acoustic measures and perceptual judgments. Munson et al. analyzed the ratings of both inexperienced and experienced listeners to test the reliability of a Visual Analog Scale (VAS) and determine if clinical training affects judgment accuracy.

The study had 42 listeners – 21 inexperienced (6 males, 15 women) and 21 experienced (1 male, 20 women). Experienced listeners were licensed speech-language pathologists with 2 – 40 years experience (M = 13 years). Experienced listeners worked in a variety of settings with different populations and disorders. All but three, who didn't provide any information, work with articulation disorders. Both groups of listeners completed a questionnaire on the amount of time they spend with children ranging from 1 (little or none) to 10 (a lot of time). Experienced listeners reported that they spend more time with children than inexperienced listeners: M = 5.7, M = 2.4 (Munson et al., 2012).

Both listener groups were asked to listen to consonant vowel (CV) sequences of children's productions of /s/ and /θ/, /t/ and /k/, and /d/ and /g/ that were elicited from real world and non-word repetition tasks. /d/ and /g/ stimuli were recorded from both monolingual English speaking children and monolingual Greek speaking children. Listeners were asked to rate the consonant they heard on a VAS. The VAS had each pair of stimuli on either end of the scale.

Click location in pixels was recorded and compared with-in group to examine probability density of rating for each group and inter-rater reliability. Using distributions of click locations regardless of target for each pair of stimuli, researchers found that experienced listeners clicked more towards /θ/, /k/, and /g/ on the scale and the distribution sof their click locations were more bimodal (have clicks at the endpoints) than the inexperienced listeners. Experienced listeners were also more reliable than inexperienced for all stimuli pairs. The greatest between-group difference was for /d/ and /g/ stimuli; researchers found that both groups accurately judged the children's productions and there were only small, insignificant differences between group ratings. In the /d/ and /g/ stimuli, both groups were susceptible to language bias where listeners rated the English speaking /d/ and /g/ in different categories than the Greek speaking /d/ and /g/. Additionally, researchers found a correlation between self-reported expertise and ratings for a particular transcription category. For example, listeners who claimed greater expertise rated the stimuli as more /s/-like for /s/, and overall had a greater ability to distinguish the two phonemes. Finally, researchers calculated how listener's perceptual judgments and acoustics measures align. For /s/ and /θ/ stimuli, a listener was more likely to rate the stimuli as /s/-like if the CV had a higher compact spectra, higher peak frequency, and was louder. For /k/ and /t/ stimuli, researchers did not find any meaningful results that could predict how a listener would judge the stimuli. If the peak ERB was lower and the stimuli was louder, the listener was more likely to rate the consonant /g/ in back-vowel stimuli. However, in front-vowel stimuli, a higher peak ERB and higher loudness would most likely result in a /g/-like rating (Munson et al., 2012).

Overall, Munson et al. (2012) found that experienced listeners were more reliable in their judgments and were better able to discriminate sounds than inexperienced listeners. Experienced listeners were also able to rate sounds closer to the endpoints when phonemes /θ/, /k/, and /g/

occur in atypical locations (i.e. a substitution). Although it is noteworthy that experienced listeners can discriminate speech sounds like /s/ and /θ/, it would be interesting to extend the study to other phonemes that are similar in articulatory complexity. Comparing /s/ and /ʃ/ would provide more information on phonological development because both sounds are among the last to be produced. /s/ and /ʃ/ have been used as examples in many phonological development studies before (Mann, Sharlin, & Dorman, 1985) to explain how children develop the ability to differentiate fricatives in speech and perception.

Current phonological development theories believe that children learn new sounds from their ambient linguistic environment. Statistical phonological development theories expand this to depend on the frequency the child hears the phoneme. The more the child is exposed to the phoneme, the greater ability for perceptual discrimination the child will have (Cristià, McGuire, Seidl, & Francis, 2011). Cristià et al. (2011) examined the statistical theory and found that children were able to learn some non-salient acoustic features. Their results suggested that, depending on the feature the child attends to and creates the distribution from, the statistical theory is a possible logical explanation for phonological development. For children, most of the ambient language from which they construct their frequency distributions consists of adult speech. Julien and Munson (2012) explored the idea of child-directed speech. Child-directed speech, or sometimes termed “motherese”, is high pitch tone and over exaggerated prosody that adults typically use when speaking with children. Julien and Munson investigated whether adults modify their speech depending on whether they perceived the child’s speech to be correct or incorrect.

Julien and Munson (2012) included 18 women and 4 men (Mean age: 25.8 years) with no history of speech, language or hearing problems. Participants filled out a questionnaire on the amount of time they spend with children weekly, ranging from 1 (none) to 10 (20 hours or more). Scores ranged from 1 – 5 meaning that at most, participants spent 10 hours a week with children.

Adults completed two tasks: the Clear-Speech Task and a task that Julien and Munson called the Listen-Rate-Say Task. The purpose of the Clear-Speak Task was to obtain a baseline of the adults' productions of 30 words used in the following task (Listen-Rate-Say Task). Participants read 30 sentences twice. The first time, the adults were given no prompt. The second time adults were asked to "speak clearly". Upon completion, adults completed the Listen-Rate-Say Task where they were instructed to listen to a child's production of a word that contained either the /s/ sound or the /ʃ/ sound and rate the production on a Visual Analog Scale (VAS). The stimuli for the task were Consonant-Vowel (CV) sequences of a fricative (/s/ or /ʃ/) and 150ms of the following vowel elicited from 22, 2- to 3-year old children. The fricatives were analyzed using conventional acoustic measures: first four spectral moments, onset of second-formant frequency, and fricative intensity and duration. The stimuli differed in all of these acoustic features. After listening to child's production and rating the production on the VAS, the adults were asked to "say the word as if you were responding to the child whose production you just rated". Adults' responses were recorded for acoustic analysis (Julien & Munson, 2012).

In the Clear-Speech Task, Julien and Munson (2012) found substantial differences between clear-speech (prompted) and baseline speech (unprompted). Each fricative was distinct from the other by properties of the centroid, or M1, duration, and vowel dispersion. Fricative duration and vowel dispersion were greater in clear-speech than baseline speech. This means that when asked to speak clearly, adults changed the duration of the fricative and made the vowels more extreme in the vowel space. Cohen's *d,* a measure of effect size, was calculated to determine the factor that differentiated the two speech styles the most. The largest effect size was for fricative duration and the smallest for the M1 of the fricative. This indicates that when asked to “speak clearly,” adults were more likely to lengthen the duration of the fricative rather than change the spectral properties. In other words, adults did not change the contrast between the two fricatives they produced though they changed the dispersion for vowels.

Julien and Munson (2012) completed several multiple regressions to examine how adult’s click location correlated with the acoustics; they discovered that the M1 and duration of the stimulus accounted for 53% of variance in click location. This indicated a moderate to strong relationship with the acoustics of the stimuli, specifically the M1 and duration of the stimulus.

Lastly, Julien and Munson (2012) investigated if adults’ responses changed as a result of their perception. Using a two-factor ANOVA, researchers found that M1 changed as a result of the target fricative regardless of the judgment. They found that the /ʃ/ fricative duration changed the most as a result of an adult judging the production as incorrect. /ʃ/ was longer when the adult perceived the child’s production as incorrect than when the adult perceived the production as correct. Phoneme /s/ duration also changed but the change was not deemed statistically significant. In addition, relative vowel durations were significantly longer if the production was judged less accurate than productions judged more accurate. Overall, Julien and Munson (2012) found that adults changed the duration of their fricative and vowel in response to children who they judged to produce a word inaccurately. It is interesting that the study did not find differences in the M1 in relation to a perceived accuracy category since the M1 is one of the primary acoustic features used to distinguish /s/ and /ʃ/. Results suggest that the purpose of child-directed speech is to draw attention to the inaccuracy rather than showing the correct manner of producing the sound.

Julien and Munson (2012) provided interesting information on adults' feedback on children's productions. The study was conducted to determine how adults' speech changes in response to children's correct and incorrect productions; however, the design of the methods lacked the naturalness needed to effectively examine this phenomenon. Adults responded to a computer and did not receive any visual input or pragmatic context of having a real child in front of them. This task might have different results if it were done with real children due to that loss of pragmatic context.

Mann et al. (1985) hypothesized that perception of speech sounds, specifically sibilants, precedes a child's ability to produce the speech sound. Therefore, they investigated whether children without an articulation disorder can better differentiate sibilants in each vowel context ([u] and [eI]) than children with an articulation disorder. The study had two groups: one with articulation mastery of the sibilant fricatives /s/ and /ʃ/ and one without. The first group sampled three ages – 5 years, 7 years and adults. All participants in the first group had to have articulation mastery of /s/ and /ʃ/ as determined by their productions of the words "Sue", "shoe", "save" and "shave". In the second group the researchers looked only at second graders (mean age: 7.6 years) with difficultly producing /s/ and /ʃ/. A speech-language pathologist chose the second graders because, in both elicited and spontaneous speech, the child incorrectly produced /s/ and /ʃ/, the child correctly produced all other speech sounds, and the child was in speech therapy for at least a year.

Both groups were asked to listen to 45 words that contained a computer-generated fricative (/s/ or /ʃ/) followed by naturally produced portion to create words such as "shoe", "Sue", "shave" and "save". The computer-generated fricative stimuli were on a continuum with slight changes in their center frequency. Nine fricatives were generated and placed on a continuum with 1 being the most /ʃ/ and 9 as the most /s/. In the experiment, adults were asked to listen to the words and write which fricative they heard. Children were given a 2-alternate forced choice response task where they were asked to choose between two pictures. One picture corresponded to the word they heard and the other differed in fricative (i.e. "Sue" and "shoe").

Researchers found that all groups began to discriminate a difference in the stimuli around stimulus 5 or 6 when the following vowel was [eɪ] and 4 or 5 when the following vowel was [u]. Meaning that if the vowel was rounded ([u]), the participants rated more /ʃ/ responses than /s/ responses than if the vowel was [eɪ]. To the researcher's surprise, the analysis showed that even children with an articulation disorder rated the fricatives similarly to children with no articulation disorder. Mann et al.'s (1985) finding contradicted the previous idea that children learn speech sounds by modifying their speech based on what they hear, because children with articulation disorders rated sibilants the same as those children without articulation disorders, yet they still have difficulty with /s/ and /ʃ/.

Mann et al. (1985) explored the idea of covert contrast. Covert contrast is the phenomenon of having acoustically different but perceptually similar productions. There is a wide range of perceptually acceptable productions of a certain phoneme. This phenomenon occurs frequently in children's speech as they are learning to acquire what adults would judge as "correct" articulation of a phoneme. For example, fricatives have covert contrast in children's speech due to their hard articulatory nature and relatively acoustically similar properties (i.e. both high frequency turbulent consonant sounds). In the Mann et al. study, the stimuli (1-9) were acoustically different, they had different frequency poles, but despite their acoustically different poles, participants rated a number of stimuli as the same phoneme. Figure 2 below displays a function to depict the change in rating a stimuli /ʃ/ across participants. The frequencies on the xaxis are the values of the pole 2 used in the Mann et al. study, the y-axis the percent of 'sh' responses.

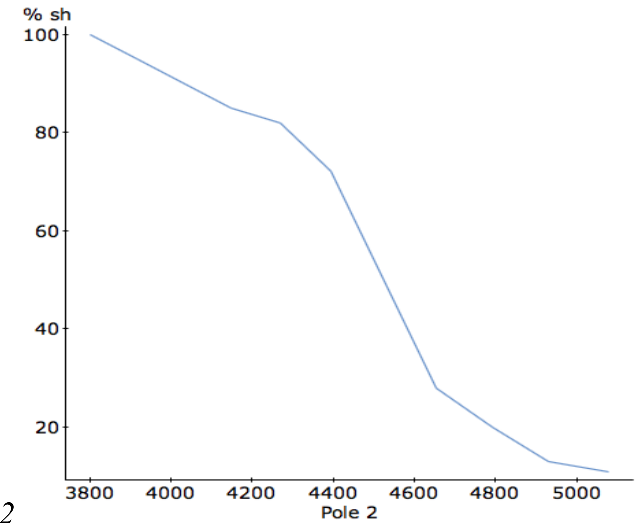


Figure 2

Notice that even as the pole frequency shifts, the perception of /ʃ/ stays relatively constant until around 4400 Hz the perception shifts, and less participants hear /ʃ/. This figure shows covert contrast – the acoustics are changing, but perception remains constant until a certain frequency.

The next section reviews studies that have attempted to measure and identify reasons for children's lower robustness of contrast in sibilant fricatives.

2.3 Robustness of Contrast Measures

Early studies investigating children's fricatives hypothesized that the differences in children and adults' productions were the result of anatomical differences such as vocal tract size (Bickley, 1980 as cited in McGowan & Nittrouer, 1987). McGowan and Nittrouer (1987) posited that anatomical differences cannot account for the higher amplitudes of lower frequencies in children's productions. Therefore, McGowan and Nittrouer examined the spectral properties themselves and, from their data, provided another possible explanation for the differences.

Eight children and 4 adults participated in the study. The children's ages ranged from 3 – 7 years with 2 children (1 male and 1 female) in each age group. The groups were divided into children aged 3 years, 4 years, 5 years, and 7 years. The children were assessed by a speech-language pathologist to have normal articulation of the fricatives /s/ and /ʃ/. Participants completed a picture response task where they repeated 10 tokens of the following disyllables: /ʃiʃi/, /sisi/, /ʃuʃu/, /susu/. The examiner modeled how to say the target disyllable and encouraged each participant to practice prior to running the experiment. Researchers recorded the responses and determined the 2nd and 3rd formant's (F2 and F3) frequency peaks in two locations – File A and File B. File A was extracted 100ms prior to the first vowel and went 80ms into the first vowel. File B was extracted 80ms before the second vowel and went 100ms into the second vowel (McGowan & Nittrouer, 1987).

In adults, researchers discovered that for all but five comparisons of F2 values, the F2 for /ʃ/ is higher than the F2 for /s/, suggesting that /ʃ/ has a further back constriction than /s/. Additionally, females create a constriction further back than males (F2 was higher) for all utterances. For both males and females, F2 was higher in the context of /u/ than for /i/. In children, vowel F2 values were more variable than those of adults. In other words, adults' F2 values were relatively similar for each fricative in every vowel context, whereas children's F2 values changed in each vowel context. Children also differed from adults in amplitude or intensity of F2. Children had a higher F2 amplitude than adults, and F2 amplitude was higher for /ʃ/ than for /s/. McGowan and Nittrouer hypothesized that the children's differences in intensity are a result of differences in glottal and anterior constriction. McGowan and Nittrouer explain the physics behind the aerodynamic workings of constrictions, demonstrating that having a larger anterior constriction, results in a relatively larger glottal power. The researchers do not have any data to support their claim but make inferences off other data in other studies for /p/ and /b/ intraoral pressure. Regardless, McGowan and Nittrouer discredited the idea that the size of the vocal tract was the sole reason behind the differences in F2 in children's and adult's speech and provided support for the idea that articulatory patterns in children differs from that of adults.

As a follow up to the earlier study, Nittrouer, Studdert-Kennedy, and McGowan (1989) examined the amount of co-articulation and contrast of /s/ and /ʃ/ occurred in the vocalic context of [i] and [u]. The study had 8 adults (20 – 21years) and 32 children split in 4 groups based on age: 3 yrs, 4 yrs, 5 yrs, and 7 yrs. All children were evaluated by a speech-language pathologist for correct productions of /s/ and /ʃ/.

All participants completed a perception experiment using the stimuli and pictures that are used in the elicited speech sample task. In the speech sample task, participants were asked to respond to pictures of the words "she", "see", "shoe" and "sue". Prior to starting the speech sample task, the experimenter provided an example of speaking the disyllables and encouraged the participants to model them. Ten tokens of the disyllables "ʃiʃi", "sisi", "ʃuʃu" and "susu" were collected from each participant. Centroid values of /s/ and /ʃ/ were taken at two different time intervals: 100ms before the onset of the first vowel (VO1-100ms) and 30ms before the onset of the second vowel (VO2-30ms) (Nittrouer et al., 1989).

At VO1-100ms, Nittrouer et al. (1989) found that adults' centroid values for each fricative were similar across vowel context with /s/ centroid around 8kHz and /ʃ/ centroid around 6kHz. In children, the /ʃ/ centroid value was higher in frequency, and the /s/ centroid value was lower in frequency than the mean adult centroid value for each respective fricative. This difference shows that children have a lower robustness in contrast between /s/ and /ʃ/ compared to adults. However, they did find that the contrast increased with age as a result of the /ʃ/ centroid lowering in frequency. Researchers also looked at the effects of coarticulation and found that there were vowel effects for /s/ but not for /ʃ/. This meant that depending on the vowel, the fricatives' centroid value would shift due to coarticulation. In /s/, the effect of coarticulation was most noticeable for [u] rather than [i]. At VO2-30ms, the same pattern of centroid values for /ʃ/ and /s/ was present. These values were to a lesser degree due to the fricative's location in the second syllable and articulation effects from the previous vowel.

To analyze the vowels, the second formant (F2) was collected from 32 tokens from each subject. Researchers found that F2 is higher in children than adults and decreases with age due to changes in vocal tract length. F2 was also higher in /ʃ/ than in /s/ and higher before [i] than before [u]. Using F2, researchers were able to identify that the effects of vowels context changes with age. As children get older the effects of coarticulation decrease (Nittrouer et al., 1989)

Overall, the study found that centroid values showed an increase in contrast of fricatives with age and stable vowel context effects. F2 values showed no change in contrast of fricatives while vowel context (effects of coarticulation) decreased with age. Nittrouer et al. (1989) concluded that because the children's perceptual organization, as measured by the perceptual study in Nittrouer, Studdert-Kennedy, and McGowan (1987), is unable to distinguish differences between phonemes children's production must rely on their perception of phonemes.

Mann et al. (1985) and other studies have since disproven Nittrouer et al.'s (1987) conclusion; children's production of phonemes does not rely on their perception. Recall that Mann et al. had two groups of children – one with an articulation disorder and one without. Both groups performed similarly on tasks of perception. Despite Nittrouer et al.'s (1987) disproven conclusion, Nittrouer et al. (1987) did find results that captured the differences between children's and adult's speech.

Li, Edwards, and Beckman (2008) further investigated the best acoustic measure to differentiate between children's sibilant fricatives. Li et al. (2008) conducted a cross-linguistic study in both English and Japanese but for the purpose of this paper only the results for English will be discussed. The study had three age groups of participants: 2 year-olds, 3 year-olds and adults. All groups of participants completed a listen and repeat-style activity. Participants listened to words that contained word-initial fricatives (/s/ or /ʃ/) and repeated the words back to the computer. Researchers controlled for variance from co-articulation by having fricatives followed by all of the vowels in all edges of the vowel space [ɪ, e, a, o, u]. All stimulus words were recorded from a native-speaker of English in "child-directed speak."

After collecting children's responses, a native-speaker of English transcribed all responses as either correct or incorrect. Researchers set 75% or higher accuracy as the criterion for mastery of the distinction between /s/ and /ʃ/. As predicted, more 3 year-old children mastered the distinction between /s/ and /ʃ/ than 2 year olds. Interestingly, in both ages, children were more likely to correctly produce /s/ than /ʃ/. This fits with the notion that a common substitution error for /ʃ/ is /s/; this suggests that /s/ is an easier phoneme to acquire than /ʃ/. Referring back to Vihman and Croft (2007) phonological development theory, /s/ is a higher frequency phoneme in the ambient language than /ʃ/ and therefore is learned and master prior to /ʃ/ (Li et al., 2008).

Researchers acoustically analyzed the productions of the adults and children. Most of the analysis was cross-linguistic and the researchers found that the acoustic features used to differentiate fricatives in English can also be used to differentiate fricatives in Japanese. In English, the primary acoustic measure to differentiate between the sibilant fricatives in adults is the centroid. However, the second spectral moment and the onset of F2 frequency also capture the differences between the two phonemes. After comparing the target fricative to the normalized value for each acoustic feature, centroid alone was enough to differentiate between /s/ and /ʃ/. In children who mastered the contrast between fricatives, the centroid was the primary acoustic measure for differentiating between fricatives. However, even though the children were considered to have mastered the contrast between /s/ and /ʃ/, the amount of contrast was much less than that of adults (Li et al., 2008). However, later researchers suggest that the acoustic measure such as spectral mean and even the highest spectral peak are not accurate at showing the robustness of contrast of fricatives. Therefore, Holliday, Beckman and Mays (2010) proposed a different acoustic analysis method of fricatives. They proposed using Peak ERB or the frequency of the loudest peak and use a different analysis spectrum to capture high frequencies. Their goal was to better capture the acoustic differences of fricatives /s/ and /ʃ/.

Peak ERB is a new method of acoustic analysis. Previously, researchers used acoustic features such as centroid, F2 frequency, and spectral moments (Li et al., 2009; Mann et al., 1985; McGowan & Nittrouer, 1987; Nittrouer et al., 1989). However, as Nicholson (2014) described, those measures do not reflect the way the basilar membrane responds to sound. The ear passes sound through a band-pass filter. Peak ERB acts like a band-pass filter by segmenting equivalent rectangular bandwidths (ERBs) of the frequency waveform and calculating the loudness of each ERB and choosing the loudness peak.

Holliday et al. (2010) had two age groups of Japanese and English speakers – adults and children (2 – 5 years). There were a total of 82 children in each language group and 17 English-speaking adults and 20 Japanese-speaking adults with a fairly even distribution of males and females. Participants completed a picture prompted word repetition task where the participants' productions of word initial fricatives were recorded and acoustically analyzed. In each production, areas of turbulence were marked off and researchers calculated Peak ERB and the compactness index (CI) of each fricative. Researchers extracted Peak ERB and CI from two different times during the production of the fricative, 90 ms before the vowel and 10ms before the vowel. In the description that follows, any measure marked “f” represents the former time interval and any measure marked “v” represents the latter time interval.

Percent accuracy predicted at each time interval was calculated for both languages. For the purpose of this paper, only the English results will be discussed. As a group, Peak ERB-f of English-speaking adults predicted 79.1% of the phonemes produced. Results were even higher for females (88.4%). The researchers did not find any benefit of adding the CI-f to the analysis. This means that using only Peak ERB-f, researchers can predict an adult’s production of /s/ and /ʃ/ with 79.1% accuracy. The researchers referred to this as the “Community Norm Model” (Holliday et al., 2010).

Holliday et al. (2010) then decided to examine results on an individual level and determine the best model to predict an individual speaker’s productions. Using the community norm model on each speaker, the mean accuracy was 89.5%. For 8 out of the 17 adults, the community norm model had the best-predicted accuracy score. For the rest, either Peak ERB-f and CI-f together or CI-f alone gave the highest predicted accuracy score. Once again, women had a higher perceived accuracy score. The mean individual perceived accuracy score for women was 94.4% (for men, 84.1%). The researchers attributed the difference in perceived accuracy to women typically having a higher frequency /s/ sound than men, thereby creating a larger contrast between the two fricatives.

In children, 38 of the 65 children who produced contrast between fricatives had the best statistical fit using Peak ERB-f alone. However, for 13 of the 65 children none of the models were able to provide a predicted accuracy score. Looking at the age range of children in the study, the researchers found an age effect – the younger the children, the lower the accuracy. There was also a large difference between genders – consistent with adult results, girls had better accuracy than boys (Holliday et al., 2010).

Holliday et al. (2010) found a more consistent and accurate way of representing the contrast of fricatives /s/ and /ʃ/ in adults than earlier studies. However, the majority of children who were modeled using Peak ERB-f had low predicted accuracy scores and there was still a large amount of children whose productions were not modeled at all. Results of this study are promising in the use of Peak ERB as an acoustic analysis method.

Nicholson (2014) expanded upon Holliday et al.'s finding by using Peak ERB, along with individual slopes, discriminability, and percent correctly predicted (%CP), to analyze the robustness of contrast of children's productions of the sibilant fricatives /s/ and /ʃ/. These new variables, individual slopes, discriminability, and %CP, provided additional information of the contrast in regards to the target fricative. For example, the individual slope measure graphs a child's Peak ERB against whether the target was /s/ or /ʃ/. Both discriminability and %CP take the target fricative into account in their calculations as well. Separately, discriminability and %CP were both found to have perceptual correlates in subsequent studies (Nicholson, 2014) creating a more perceptually relevant analysis for Nicholson's (2014) study.

Participants were recruited from a larger study called "Learning to Talk." "Learning to Talk" is a federally funded project investigating word learning in children from different economic and linguistic backgrounds. Thirty-nine participants' data from the larger study were analyzed based on their raw vocabulary score as measured by the Expressive Vocabulary Test (EVT). Participants were children aged 28 to 39 months; there were 20 males and 19 females.

Participants were divided into five groups based on their raw EVT scores. Group one's scores ranged from 0 – 14, group two's ranged from 15 – 29, group three's from 30 – 44 and groups four and five, 45 – 59 and 60 – 74 respectively. Each participant completed a real word repetition task where they listened to a computer play a well-known word and repeated the word. The child's productions were recorded and then segmented and turbulence tagged. Further explanation of the methods involved in segmentation and turbulence tagging can be found in the methods section of this paper or in Nicholson (2014).

Nicholson (2014) found no significant effects of vocabulary score with Peak ERB and percent correct predicted. In other words, vocabulary was not a factor or indicator of contrast between sibilants. Nicholson did find that differences in peak ERB values of /s/ and /ʃ/ increased with age; the effect was due to a decrease in Peak ERB of /ʃ/ and constant Peak ERB of /s/. Percent correct predicted score also increased with age – as the children age their robustness of contrast increases.

When selecting participants out of the larger study, "Learning to Talk," Nicholson (2014) did not control for vocabulary size and therefore, could have skewed the results to have no effect between EVT scores and robustness of contrast. Nicholson's paper lacks a perceptual component to affirm that the decrease in Peak ERB of /ʃ/ is perceptually salient and meaningful.

For decades, researchers have investigated the best method for analyzing children's fricatives. Unfortunately, fricatives are extremely complex to measure in children because they are acoustically similar due to their difficult articulatory nature; children have a low robustness of contrast. Holliday et al. (2010) had the best results using Peak ERB to examine the differences between fricatives in adults and children. Recall that he found above 80% predicted accuracy for fricatives in adults. However, the predicted accuracy score in children was much lower and for some children, the acoustic method was not even feasible. Nicholson (2014) further investigated how to capture children's robustness of contrast by using Peak ERB frequency with other individual variables. Although she did not have as conclusive of results using percent correctly predicted, errors could result from improper control of subject's vocabulary score.

The motivation behind the study reported in this paper stems from the idea that both acoustic and perceptual measures provide a complete representation of, in this case, a child's robustness of contrast of /s/ and /ʃ/. Acoustic measures are considered the "gold standard". They eliminate listener bias and can be replicated by different researchers. However, acoustic measures are often impractical for clinical and everyday use. Within the scope of this paper, acoustic models have yet to accurately capture the children's robustness of contrast. Perceptual measures are more clinically relevant and ecologically valid because they are measurements used in everyday life – clinicians use perceptual judgments on children's articulation of speech sounds and parents give feedback to children based on their perception of the goodness of the production (Julien & Munson, 2012). However, perceptual judgments are prone to listener biases – one listener's perception of a phoneme will differ from another's perception. Therefore, using perceptual measurements in conjunction with acoustic measurements provides the most detailed report of findings.

The goals of this paper are to extend the findings of Nicholson (2014) by adding a perceptual component, to provide ecological validity to Nicholson's acoustic analysis methods, to provide additional evidence of correlation of vocabulary size and robustness of contrast through perceptual means. The author will use the same participants analyzed in Nicholson paper and transcribe and provide accuracy measures of the children's productions. This paper will also examine the correlations between judgments, vocabulary size and age, and how the findings align with those found in Nicholson's paper.

Section 3: Methods

This section describes the methods used to obtain the children's productions of the sibilant fricatives, the analysis processes undertaken to prepare for transcription, and the transcription methodology itself. The recordings of children's productions and segmentation were done as part of a larger study called "Learning to Talk" that was described earlier.

3.1 Participants

The participants for this project are 15 of the participants used in Nicholson (2014) and were recruited for a larger NIH funded project, "Learning to Talk." There were 8 females and 7 males with ages ranging from 28 – 39 months (M = 33.5 months) and all had normal hearing. Participants were split into five bins based on their raw EVT score using the same criteria as Nicholson (2014). The five bins score ranges were as follows - Bin 1: 0 – 14; Bin 2: 15 – 29; Bin 3: 30 – 44; Bin 4: 45-59; Bin 5: 60-74 (*see table below)*.

Table A

Bin 1	Age	Bin 2	Age	Bin 3	Age	Bin 4	Age	Bin 5	Age
001L	28	629L	30	018L	35	615L	36	600L	37
036L	29	066L	38	611L	32	012L	30	603L	35
013L	32	025L	27	623L	31	608L	39		
605L	30			646L	34				

3.2 Stimuli

The data for each participant were obtained through a Real Word Repetition Task used for the larger study. The task presented 99 trials of real words that are high frequency words for

2- and 3-year-old children. The computer presented the words in the child's native dialect (Standard American English or African American English) and the child was instructed to repeat what the computer said. The child's responses were recorded on a Marantz Recorder.

3.3 Segmentation and Turbulence Tagging

After the Real Word Repetition Task, each participant's responses were segmented by experienced segmenters using a Praat script. The script prompted segmenters to specify the context of the child's response as Response, Voice Prompt, Unprompted or Non Response (see Nicholson (2014) for a more detailed explanation of each context). After segmentation, an experienced segmenter with over a year of experience, checked each TextGrid file. Any repetitions segmented as Response and Unprompted Response were turbulence tagged. In stage one of turbulence tagging, the tagger is prompted to specify whether the initial phoneme is a sibilant fricative, sibilant affricate, non-sibilant fricative, non-sibilant plosive, or other (e.g. [l], [m]). If the phoneme is judged as a fricative, the tagger specifies boundaries for turbulence onset and Voice-Onset Time (VOT) and vowel end. If there is a period of breathiness between the turbulence and the onset of the vowel a turbulence offset is marked. All of the files were turbulence tagged using an additional Praat script. Nicholson completed all but one of the turbulence tagged text grids included in this paper.

3.4 Transcription

After a file was segmented and turbulence tagged, the author went through each file and manually transcribed each interval of sibilance using the WorldBet phonetic symbol system (Hieronymus, 1993). However, due to limitations of transcriptions that have been discovered in many research papers (Julien & Munson, 2012; Li, Edwards & Beckman, 2008) the author created intermediates to create a continuum and provide more accurate representations of the children's productions. The four codes used to denote the different types of fricatives were: /s/, /ʃ/, /s:ʃ/ for sounds that are not quite /s/, and /ʃ:s/ for sounds that are not quite /ʃ/ (*Figure 3)*. The WorldBet symbol [hl] was used to represent any sounds that were lateralized or contained some sort of distortion to their production.

Figure 3

/s/ /s:ʃ/ /ʃ:s/ /ʃ/

For target /s/ words, /s/ and /s:ʃ/ was correct. Any other sound was considered incorrect: /ʃ:s/, /ʃ/ or affricates or other fricatives. For target /ʃ/ words, /ʃ/ and /ʃ:s/ was considered correct and sounds – /s:ʃ/, /s/, affricates and other fricatives – were incorrect. All files were checked after being transcribed by the author again to ensure self-reliability and that no mistakes were made while manually entering the transcriptions. The checking occurred after discussing the reliability-checked files mentioned below.

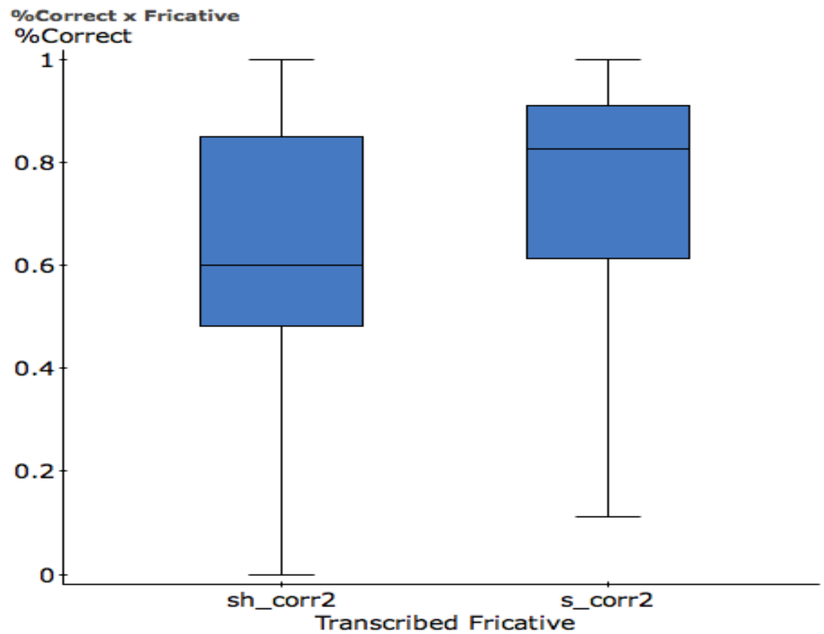
3.5 Reliability

Due to the subjective nature of transcriptions, the author had an experienced graduate student reliability check, the subjectively "most difficult", files in each bin. Initially, reliability was 79.1% for five files. Both transcribers, the author and the graduate student, sat and discussed each utterance that they disagreed on. After discussing each one, reliability increased to 89.5%. Considering the reliability check was done on files deemed "most difficult", the author considered 89.5% acceptable. Disagreements that were not resolved stayed with the author's transcription.

Section 4: Results

A series of statistical analyses were performed to understand the factors that affected the accuracy of children's productions of fricatives. This included statistical tests of the difference in percent correct accuracy for target /s/ and target /ʃ/, as well as statistical associations between the accuracy measures and other measures, including age, vocabulary size, and the measures of acoustic differentiation between /s/ and /ʃ/ documented in Nicholson (2014). As mentioned in the methods, for target /s/ both /s/ and /s:ʃ/ were considered correct and for target /ʃ/ both /ʃ/ and /ʃ:s/ were considered correct.

Figure 4



As this figure shows the mean percent correct of /ʃ/ was 58% with a standard deviation of 32% and the mean percent correct /s/ was 71.8% with a standard deviation of 26%. Children produced /s/ correctly for target /s/ more than they produced /ʃ/ correctly for target /ʃ/.

A two-tailed paired-samples t-test was conducted to examine whether the difference between percent correct /s/ and percent correct /ʃ/ was statistically significant. The t-test showed no significant difference between percent correct /s/ and percent correct /ʃ/ (t[15]= 1.86, p=.082). However, the significance level was approaching the conventionally used .05 level. The failure to achieve statistical significance may have been a product of the relatively small sample size. In the larger study with a larger sample size, the difference between percent correct /s/ and /ʃ/ may be statistically significant.

Following the paired t-test, correlations between percent correct for both phonemes and age and EVT score were calculated. Both percent correct /s/ and percent correct /ʃ/ were correlated with age at a significance level of $\alpha = 0.05$ and both correlations suggested a positive relationship: as a child ages their ability to articulate /s/ and /ʃ/ increases ($r = .61$ and .65 respectively). There was also a moderately strong correlation between EVT scores and percent correct for /s/ and /ʃ/ ($r = .53$ and .61 respectively, $\alpha = 0.05$) suggesting a similar relationship – as a child's vocabulary size increases their ability to articulate /s/ and /ʃ/ increases as well. However, upon further examination, Age and EVT were also highly correlated ($r = .52$, p = .04) suggesting that relationships between percent correct, age and EVT are due to the interaction of the two variables. Therefore, a partial correlation between accuracy and EVT controlling for age was performed. The partial correlation did not show any statistically significant correlations between percent correct /s/ and /ʃ/ and EVT. However, the correlation value for percent correct /ʃ/ and EVT ($r = .43$) is strong and therefore provides evidence for the author to believe that the results may be significant in the larger study ("Learning to Talk") when the entire sample is analyzed (n=180), as this would lead to less variable data than in this smaller subset. If the results were statistically significant, it would confirm the hypothesis that as children learn more words their ability to create contrast between sounds (e.g. /s/ and /ʃ/) increases.

Partial correlations controlling for age between robustness of contrast measures (i.e. discriminability) from Nicholson (2014) and percent correct /s/ and /ʃ/ were also calculated. The correlation between percent correct /s/ and discriminability was not significant ($r = .25$, $p = .42$) however, the correlation between percent correct /ʃ/ and discriminability was significant at the α

= 0.05 level (*r* = .57, *p* = .04). This finding suggests that children who had better contrast between /s/ and /ʃ/ were more accurate in their productions of /ʃ/. This finding aligns with Nicholson (2014) findings that the Peak ERB of /ʃ/ decreased with age.

Lastly, paired sample t-tests were conducted to examine how transcribed target productions of /s/ and /ʃ/ differed quantitatively. Not surprisingly, phoneme /s/ was transcribed more for target /s/ than for target /ʃ/. Similarly, phoneme /ʃ/ was transcribed more for target /ʃ/ than for target /s/. The intermediate phoneme /ʃ:s/ was transcribed more for target /ʃ/ than for target/s/; although hypothesized, the difference in significance in phoneme /ʃ:s/ and [s:ʃ] suggests an interesting finding. The lack of statistical significance in [s:ʃ̹] being transcribed more for target /s/ than target /ʃ/ suggests that [s:ʃ̹] was transcribed relatively similar in both target utterances. This finding may arise from children's depalatalization errors when producing /ʃ/. Additionally, /ʃ/ is usually acquired latter than /s/ in English, and as shown from previous results, decrease in /ʃ/ peak ERB indicates an increase in the child's robustness of contrast.

Section 5: Discussion

Previous studies dating back to the 1980s, have attempted to capture children's robustness of contrast acoustically (Mann et al., 1985; McGowan & Nittrouer, 1987; Nittrouer et al., 1989). Holliday et al. (2010) was able to distinguish fricatives /s/ and /ʃ/ in adult's speech and a majority of children's speech. However, there was still a need to find an acoustical model that could depict a child's robustness of contrast. Nicholson (2014) developed an acoustic model that incorporated Peak ERB used in Holliday et al. (2010) with individual variables (individual slopes, discriminability, and percent correctly predicted) and found interesting correlations between discriminability (measure of robustness of contrast) and age. This study had three goals: 1. add a perceptual component to the model in Nicholson (2014), 2. examine the correlation of perceptual judgments, age, and EVT scores, and 3. examine the correlation of perceptual judgments and Nicholson's (2014) discriminability measures.

The results of this study suggest that the acoustic model in Nicholson (2014) has ecological validity. The strong correlation between discriminability and percent correct /ʃ/ aligns with the findings in Nicholson (2014). The correlation indicates that children who had better contrast between /s/ and /ʃ/ had a high percent correct /ʃ/. Furthermore, the results suggest that a larger contrast between speech sounds is a result of age. For example, as a child ages, they gain greater control over their articulatory muscles and therefore are able to articulate /ʃ/ with greater precision. This result could also be a result of vocabulary size. The correlation between EVT and percent correct /ʃ/ was strong but not statistically significant. However, the author believes that in the larger study ("Learning to Talk") this relationship will become statistically significant. A significant relationship between vocabulary size and percent correct /ʃ/ would provide evidence for vocabulary size factoring into children's phonological development due to the significant relationship between percent correct /ʃ/ and discriminability.

Continued research on differences between children and adults' speech is still needed. Possible explanations of the correlation between percent correct /ʃ/ and discriminability could be physiological (e.g. a child's ability to articulate the differences increases) or developmental (e.g. the vocal tract length increases providing the lower fundamental frequency of /ʃ/). All of these possible explanations would need to be explored by future research. In addition, this study was based solely on the author's judgments and therefore not as ecologically valid as if the judgments were obtained from a sample of untrained listeners. Although Munson et al. (2012) found experienced listeners are better able to perceive intermediate productions better than inexperienced listeners, Julien and Munson (2012) found that adults, or inexperienced listeners, alter their speech directed to children; therefore, research looking into inexperienced listeners' feedback to children's productions and the resulting robustness of contrast is needed to confirm the validity of the experiment.

Another limitation of the study is low inter-transcriber reliability. Unfortunately, transcribing is subjective and therefore allows for large variability between transcribers. Smit et al. (1990) used a modified transcription system (a checklist) that limited the number of phonemes the transcriber could select as a possible transcription. The mean inter-transcriber reliability in Smit et al. was 76.4% and 78.2% from Iowa and Nebraska transcribers respectively. However, this reliability looked at if the transcribers selected the same phoneme, not the agreement on whether the production was "acceptable" or an "error". The accuracy inter-transcriber reliability was 94% of 100 responses. Again, this reliability represents the amount of agreement on whether the production was "acceptable". The author suggests using a modified transcription system that is being used in the larger "Learning to Talk" study. This transcription system utilizes what Smit et al. (1990) did, limiting possible phoneme choices, by guiding the transcriber to the appropriate transcription symbol. For each production, a script in Praat prompts for decisions on place, manner, and voicing to transcribe the appropriate symbol.

Section 6: Conclusion

The following study examined the relationship between perceptual judgments and an acoustic analysis model suggested in Nicholson (2014). The study found a significant relationship between discriminability, a measure of robustness of contrast, and percent correct /ʃ/ to suggest a relationship between a child's articulatory mastery of /ʃ/ and their overall contrast between phonemes /s/ and /ʃ/. These findings, combined with Nicholson's, allow for a greater understanding of a child's robustness of contrast and its relationship to age and EVT. By utilizing both acoustic measures and perceptual judgments, the findings have more validity because they were not only apparent in the acoustic measures but were also perceptually salient. The alignment between the two suggest a strong relationship between percent correct /ʃ/ and discriminability.

References

Ali, A. M. A., Van der Spiegel, J., & Mueller, P. (2001). Acoustic-phonetic features for the automatic classification of fricatives. *The Journal of the Acoustical Society of America*, *109*(5), 2217-2235.

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of phonetics*, *39*(3), 388-402.

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

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 419-439. Retrieved from http://links.jstor.org/sici?sici=0097-8507%28197506%2951%3A2%3C419%3AWASIEL%3E2.0.CO%3B2-A

Holliday, J. J., Beckman, M. E., & Mays, C. (2010). Did you say susi or shushi? Measuring the emergence of robust fricative contrasts in English-and Japanese-acquiring children. Submitted to InterSpeech 2010.

Hieronymus, J. L. (1993). ASCII phonetic symbols for the world's languages: WorldBet. *Journal of the International Phonetic Association*, 23.

Jakobson, R. (1968). *Child language: Aphasia and phonological universals*. (1 eds.). The Hauge, Netherlands: Mouton Publishers. Retrieved from http://books.google.com/books?hl=en&lr=&id=i2gUutbrJLkC&oi=fnd&pg=PA7&dq=Jakobson, 1968 Child language, aphasia and phonological universals&ots=dLvdpO6zK7&sig=eTiQejXusO4HI1x1Fc29mrzQuGE

Julien, H. M., & Munson, B. (2012). Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech, Language, and Hearing Research*, *55*, 1836-1849. doi: 10.1044/1092-4388(2012/11-0131)

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*, 111-124. doi: 10.1016/j.wocn.2008.10.001

Mann, V. A., Sharlin, H. M., & Dorman, M. (1985). Children's perception of sibilants: The relation between articulation and perceptual development. *Journal of experimental child psychology*, *39*, 252-264.

McGowan, R. S., & Nittrouer, S. (1987). Differences in fricative production between children and adults: evidence from an acoustic analysis of/ʃ/and/s. *The Journal of the Acoustical Society of America, 83*, 229-236.

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, *21*, 124-139. doi: 10.1044/1058-0360(2011/11-0009)

Nicholson, H.B. (2014). Exploring variation in accuracy and contrast for sibilant fricatives at the onset of fricative acquisition. (Unpublished Master's thesis). University of Minnesota, Minneapolis.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, *32*, 120 – 132.

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*(779-798).

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell Publishing.

Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, *45*, 683-725. doi: 10.1515/LING.2007.021