EXPLORING VARIATION IN ACCURACY AND CONTRAST FOR SIBILANT FRICATIVES AT THE ONSET OF FRICATIVE ACQUISITION

A THESIS SUBMITTED TO THE FACULTY OF
THE UNIVERSITY OF MINNESOTA
BY

Hannele Buffy Marie Nicholson

IN PARTIAL FULFILLMENTS OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

Benjamin Munson, PhD.

April 2014

Acknowledgments

While obtaining the data presented in this thesis, I was supported by the Learning to Talk grant made possible by the National Institute for Deafness and Other Communicative Disorders and by the National Science Foundation. First and foremost, I owe thanks to my advisor Benjamin Munson who guided me from topic selection to the final stage of submission. Without the programming assistance of Patrick Reidy, Mary Beckman and Jeff Holliday, data collection and analysis would have been much more onerous. Thank you to Maria Swora for tirelessly recruiting families for the University of Minnesota section of the lab. I would also like to offer my thanks to all of the experimenters, past and present, at the University of Wisconsin-Madison and University of Minnesota whose patience and coaxing made obtaining speech from toddlers seem a trivial task. Thank you (in alphabetical order) to Jamie Anderson, Sara Bernstein, Ruby Braxton, Jamie Byrne, Cara Donohue, Tyler Ellis, Michelle Erskine, Kerri Engel, Colette Felion, Courtney Huerth, Isla Katz, Kelly Jorgenson, Sarah McGowan, Haley Webb, and Colleen Woyach. For assistance with segmentation, I am indebted to Jamie Byrne, Rose Crooks, Cara Donohue, Tyler Ellis, Michelle Erskine, Megan Flood, Amy Muzynoski, Sarah Schellinger, Bianca Schroeder, Janet Schwartz, Kristi Warndahl and Haley Webb. I would like to especially thank Rose Crooks for her assistance with turbulence tagging of the data used in this thesis. Thank you to my parents for their support and countless nights of grandma time. For countless cups of hot beverages and for being my companion through yet another thesis, I am indebted to Joseph Eddy.

To Torian whose [dʌvəɬ] *made it all relevant*

Abstract

Children's speech differs from adult speech in the many ways, including in its phonetic characteristics. A central question for researchers interested in child speech sound acquisition is when and how a child acquires robust adult-like contrasts. In this thesis, I present a protocol for the analysis of the English sibilant fricatives [s] and [ʃ]. Sibilant fricatives are of interest because they are late-acquired sounds that require articulatory-aerodynamic coordination, and are contrastively necessary in multiple languages around the world, English especially. Given the turbulent nature of the sound spectrum of fricative consonants, few agreed upon measures exist. Holliday, Reidy, Beckman and Edwards (In Preparation) propose that peak equivalent rectangular bandwidth is a psychoacoustically appropriate measure for modeling the robustness of phonological contrast between sibilant fricative types. The robustness measures put forth by Holliday et al. are applied to data from the speech of toddlers aged 28-39 months and are discussed.

Table of Contents

List of Tables

List of Figures

Figure Page

1 Introduction

Children's speech differs from that of adults. A central question in the investigation of phonological development concerns the factors that may influence these differences. According to the tenets of Generative phonology, children acquire phonological contrasts according to an innate and universal sequence (Jakobson, 1941/1968). An alternative model posits that children acquisition of new phonological contrasts is tied to their acquisition of new lexical items (Ferguson & Farwell, 1975; Vihman & Croft, 2007; Beckman & Edwards, 2000). These contrasts develop in multiple sensory domains at a multiple levels of abstraction away from raw sensory experiences. These multiple domains include perceptual knowledge, articulatory knowledge, abstract phonological knowledge and social-indexical knowledge (Beckman & Edwards, 2000; Edwards, Beckman, & Munson, 2004).

In this thesis, I will explain a methodology developed in order to study the robustness of contrast between the sibilant fricatives [s] and [ʃ] in the speech of two and three-year old children. The data were taken from a subset of 35 children participating in a longitudinal study on the development of the types of phonological knowledge listed in the previous paragraph. The overall purpose of this thesis is twofold. The primary purpose is to describe the development of the methods for measuring the robustness of contrast, as well as a justification for studying contrast development acoustically. A second ancillary purpose is to speculate about the potential implications for models of phonological development and how factors such as chronological age, vocabulary size and gender affect the acquisition of the /s/-/ʃ/ contrast. In the larger ongoing, longitudinal, multi-site study of phonological development, children over the age of 28 months return annually at three timepoints to complete a battery of standardized and non-standardized tasks. The data presented in this thesis are taken from 35 participants because this was deemed a suitable subset to use in order to develop a methodology for analyzing the full data set of approximately180 participants. Data collection is still in process for the second timepoint.

Preliminary results suggest that both age and vocabulary size are statistically related to two measures of the robustness of the /s/-/ʃ/ contrast, percent correctly classified by a logistic regression predicting fricative type from acoustic measures (%CP) and discriminability (*d(a)* or Cohen's d).

In Chapter 2, I review the literature pertaining to child language development, fricative production, methods for analyzing the acoustic measurements of fricatives and a summary of what previous studies have revealed. Chapter 3 describes the methodology employed in the current study, with a particular emphasis on of how the individual productions were analyzed. In Chapter 4, I present the results from a number of analyses of robustness of contrast measures. Finally, in Chapter 5 I discuss the relevance of the results and make predictions for a model of phonological development.

2 Literature Review

In this section, I present a review of background literature and motivate the need for the methods and analyses used in the present study.

2.1 Child Language

Children's speech perception and speech production differs widely from that of adult speakers of the same language. This is true from the earliest stages of acquisition through to adolescence. In terms of production, child speech differs from adults' speech in token-to-token durational and spectral variability (Kent & Forner, 1980; Smith, 1978), overall vowel duration (Lee, Potaminos, Narayanan, 1999), coarticulatory patterns (Nittrouer, 1993; 1995, Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, Studdert-Kennedy, & Neely, 1996), and spectral characteristics of fricatives (Li, 2012). In terms of perception, Hazan and Barrett (2000) found that children as old as 12 years of age perceived phonemic contrasts differently than adults. As children developed

from 6 to 12 years of age, their judgments in a categorical perception task increased in consistency. Adults still outperformed children aged 12, however, suggesting that the development of speech perception abilities continues into adolescence.

Further evidence attesting to the differences between a child's perceptual capabilities and an adult's comes from the work of Nittrouer and colleagues (Nittrouer, 1992; Nittrouer & Miller, 1997; Nittrouer, Manning and Meyer, 1993). Nittrouer (2002) proposed that there are developmental differences in the way children and adults perceive transitional cues between consonants and vowels in fricative-vowel sequences. In these sequences, both the spectrum of the frication noise and the shape of the formant transition provide cues to the fricative's place of articulation. Children were found to weight formant transition cues more heavily than fricative-noise spectra (Nittrouer, 2002). Mayo and Turk (2004) proposed that a child's cue weighting strategies differed depending on the segmental context. The children tested by Mayo and Turk used transitional cues for sibilant fricatives but were unable to perceive a difference between /de/-/be/ which suggests that segmental context may influence perceptual strategies. Mayo and Turk propose that a lack of refined auditory skills may explain the difference in perceptual strategies used by children compared to those used by adults.

In addition to developmental differences in speech perception, there is also substantial evidence of speech production differences between adults and children. The achievement of adult-like abilities occurs at different times for different tasks. According to phonetically transcribed data, English-learning children acquire the contrast between /p/ and /b/ before the age of 2 years (Sander, 1972; Templin, 1957). When more granular acoustic measurements including VOT, stop-gap measurements, and burst amplitude are used, however, it becomes evident that children develop adult-like voicing distinction in stages from 18-months through to 11 years of age (Whiteside & Marshall, 2001). Models that relied upon Sander (1972) and Templin (1957) provide a very coarse-grained view of phonological development. When finer-grained acoustic and perceptual analyses are used, such as those described by Whiteside and Marshall (2001), a

different pattern of development is evident. When phonetic transcription alone is used, it may appear that children acquire voicing contrasts early on but when more detailed cues are taken into consideration, it becomes apparent that this distinction develops well into early adolescence.

Many studies of child phonology have relied upon phonetically transcribed data (Sander, 1972; Wellman, Case, Mengert & Bradbury, 1931; Templin, 1957). Transcribed data is almost certainly affected by the transcriber's abilities and biases in her perception of the child's production (Scobbie, 1998; Edwards & Beckman, 2008). Li, Munson, Edwards, Yoneyama and Hall (2011) argue that acoustic analyses of the productions should be used together to provide a comprehensive picture of the actual contrasts a child is capable of making. If development occurs in multiple sensory domains and at multiple levels of abstraction, then we need a set of tools that can measure this. For this reason, any models of phonological contrasts should utilize acoustic measurements alongside measures of by adult perception. Thus, a method for analyzing sibilant fricatives should not rely solely on the use of traditional transcription methodology but should instead supplement that with reliable acoustic-phonetic measurements

In the tradition of Generative Phonology, Jakobson (1941/1968) applied the concept of markedness, originally proposed by Trubetzkoy (1939), to child phonological development. Markedness refers to a binary opposition between items where unmarked phenomena are default and simple compared to marked ones (Trubetzkoy, 1939, Johnson & Reimers, 2010). According to Jakobson, a child who is engaged in the process of acquiring a specific language's phonology (e.g. English) must work through a series of Universal contrasts as he or she acquires the contrasts needed for the given language. For example, in the infant stage, children begin by acquiring the simplest, unmarked contrasts such as the distinction between the vowel /a/ and the consonant /b/. As the child matures, he or she acquires contrasts according to an innate schedule of acquisition that moves from unmarked to more marked contrasts. Jakobson's view dominated phonological theory for several decades. Jakobson's hypothesis was not supported by data from

subsequent longitudinal studies of phonological development. These showed variable patterns of acquisition that deviate from Jakobson's proposed fixed order (Ferguson & Farwell, 1975; Menn, 1983; Johnson & Reimers, 2010). Furthermore, Jakobson's view discounted the importance of babbling by suggesting that it was not linguistically relevant. Subsequent work showed that babble is indeed very linguistically relevant. In a study of infants immersed in Swedish, English, French and Japanese language environments, Boysson-Bardies and Vihman (1991) observed babbled speech tended to reflect that phonetic structure of the language spoken by the infants parents. For example, infants in French language environments produced more labial consonants while soon-to-be Swedish-speaking infants produced more dentals (Boysson-Bardies & Vihman, 1991). Stoel-Gammon and Cooper (1984) found that prelinguistic babbling contained sounds that were later used in early words. Vihman and Keren-Portnoy (2011) found an association between the amount of time a child has been producing consonants in babbled speech and subsequent phonological memory for these sounds in nonword repetition tasks.

An alternative to a Jakobsonian model of development is one in which phonological patterns are viewed as emergent from development in other non-linguistic domains (like the development of motor control and articulatory acuity), and lexical growth (Ferguson & Farwell, 1975; Beckman & Edwards, 2000; Lindblom, 1992; Munson, Edwards & Beckman, 2005; Pierrehumbert, 2003; Vihman & Croft, 2007). As a non-nativist approach, this theory is rooted in the assumption that children are not born with innate rules or processes. Instead, acquisition of a phonological system occurs gradually as the child acquires new lexical items. A proposal for the way that this mechanism works in early phonological development is presented by Vihman and Croft (2007). In their model, a new word may be paired with a phonetic template that is specific for that child. As the processes of motor control and perception mature in the child, the templates are refined in an child-specific fashion depending on the target language and the individual child. Vihman and Croft (2007) point out that children typically use words that can be produced with

their current phonetic repertoire and that they tend to adapt adult productions to match their phonetic repertoires.

According to Ferguson and Farwell's (1975) predictions, if a phonological system emerges from the lexicon, then one may expect an association between individuals with larger vocabularies compared to individuals with smaller vocabularies. Edwards, Beckman and Munson (2004) found evidence to support these predictions. In their study, participants aged three to eight years repeated novel nonsense words (e.g. /petik/ or /næfkətu/) that contained either a high (/ik/ in /petik/) or low (/fk/ in /næfkətu/) frequency target sequence. Participant productions of these target sequences were transcribed according to their accuracy in terms of place, manner and voicing. In order to determine a lexical measure, participants were administered the *Peabody Picture Vocabulary Test (PPVT)* and the *Expressive Vocabulary Test* (EVT). Results showed that both vocabulary size and age were associated with phonological accuracy. Children with larger vocabularies were more accurate on both low and high frequency target sequences compared to children with smaller vocabularies. Accuracy for both low and high frequency sequences increased with age. Thus, it would appear from Edwards, Beckman and Munson's results that higher-order categorical phonological knowledge is acquired gradually as a child acquires a lexicon.

The finding that development appears to be different depending on the level of analysis used suggests the need for a richer model that takes into account development in multiple sensory domains and at multiple levels of analysis. One such model is presented by Beckman and Edwards (2000), and Edwards, Beckman, and Munson (2004). These investigators suggest that knowledge of phonological categories may be broken down into several knowledge dimensions including articulatory instructions, perceptual knowledge, knowledge of abstract phonological categories and social-indexical knowledge. Given that phonological acquisition is a gradual process that occurs over several years as the child ages, Beckman and Edwards (2000) propose that researchers should develop specific models to account for phonological development at

individual stages. These models should take into account the different sources of knowledge available to the child. For example, one could develop a methodology for the robustness of contrast of sibilant fricatives.

While a child may have access to perceptual knowledge, there may not always be a direct mapping between perception and production. Berko and Brown (1960) showed that children are capable of perceiving contrasts that they cannot yet produce. This is illustrated by the well-known /fɪs/ phenomenon. This is the phenomenon in which a child may refer to a toy fish as /fɪs/ but promptly corrects an adult who says "Is this your /fɪs/?" by saying "No, my /fɪs/." Thus, the child can perceive that the adult production is incorrect. Beckman and Edwards (2000) explain the /fɪs/ phenomenon by explaining how the mapping between perception and production may not be linear. Although /s/ is perceptually salient and occurs in many English words, it is still one of the last sounds acquired by children. Beckman and Edwards point to the motor control required to distinguish /s/ from /ʃ/ as well as /s/ from /θ/ explains the potential discrepancy between perceptual and articulatory knowledge.

2.2 Fricative Production

Fricative consonants are produced by pushing air through a narrow constriction between two articulators (Fant, 1960; Ladefoged, 2001). The resulting sound is a turbulent noise with resonances in various frequency ranges. Sibilant fricatives /s/ and /ʃ/ typically exhibit higher-frequency spectral energy than interdental and labiodental fricatives such as /f/ and /θ/ when produced by adult speakers (Ladefoged, 2001). Owing to the sublingual cavity created during its production, [ʃ] has a primary peak frequency around 2500 – 3000 Hz, while the more anterior constriction for [s] with leads to a higher peak frequency around 4000-5000 Hz in adult speakers

(Jongman, Wayland, & Wong, 2000). These are illustrated in two spectrograms below. These were stimuli used in the current experiment, and were produced by the author.

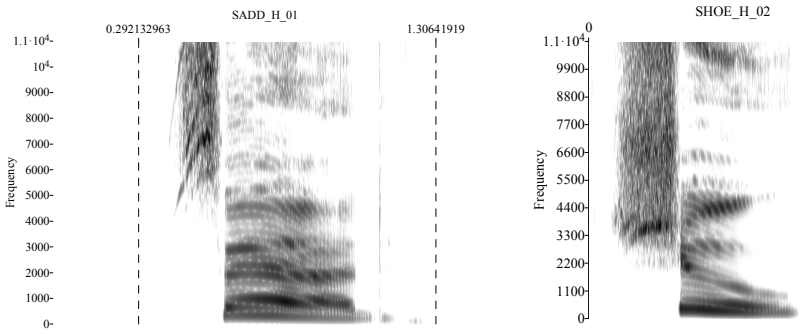

Figure 1. Stimuli from the Real Word Repetition experiment showing sibilant fricatives [s] for "Sad" (left) and [ʃ] for "shoe" (right). The frequency for [ʃ] exhibits lower energy than is seen with [s].

Owing in part to smaller overall oral cavities, the spectral frequencies for fricatives are higher when produced by children than when produced by adults (Li, Edwards, & Beckman, 2009; Nittrouer, Studdert-Kennedy & McGowan, 1989; McGowan & Nittrouer, 1988). Li, Edwards and Beckman (2009) report onset F2 values of English-speaking adults that range from 1500-3000 Hz for [ʃ] and 1500-2700 Hz for [s]. Centroid values, or the weighted mean frequency, ranged from approximately 4000-6000 Hz for [ʃ] and 7000-11500 Hz for [s]. Average values were not reported for children but a single child identified as having a clear contrast between sibilant fricatives showed centroid ranges from 4000-7000 Hz for [ʃ] and 6500-11500 for [s]. A child identified as exhibiting a covert contrast produced target [ʃ] centroid frequencies ranging from 4000-11000 Hz while target [s] ranged from 5000-9000 Hz.

Todd (2009) found that children with cochlear implants produced sibilant fricatives that were less distinct from one another compared to children with typical hearing. Targets for [ʃ] were closer to those of typical hearing peers than targets for [s]. This difference was attributed to poor frequency detection abilities in cochlear implants which made perception over 4000 Hz difficult. Since acoustic energy for [s] is typically above 4000 Hz while [ʃ] is below 4000 Hz, this

implicates the frequency-processing abilities of the implants themselves. Overall, both groups of children were more accurate when producing [ʃ] compared to [s]. Children with cochlear implants made [f] for [s] substitutions suggesting once again perceptual difficulties with [s].

2.3 Acoustic Measures Used to Characterize Fricatives

This section describes measures used to analyze fricatives. These vary across the field. There is no nearly universally agreed upon measure for fricative acoustics in the way that there is for vowels, where formant frequency and bandwidth are nearly universally used to characterize vowels. Section 2.3.1 presents one widely used method for characterizing fricatives, spectral moments analysis (Forrest, Weismer, Milenkovic, & Dougall, 1988; Jongman, Wayland, & Wong, 2000; Li, Edwards & Beckman, 2009). Spectral moments are calculated by treating the power spectrum as a random distribution of numbers, and calculating the first four statistical moments of this distribution: mean, standard deviation, skewness, and kurtosis. These measures have been useful in discriminating among different fricative places of articulation, at least for adults' productions. One disadvantage of spectral moments it that they are not a psychoacoustically realistic measure, as the auditory system does not compute statistics over a spectrum.

Section 2.3.2 presents an alternative measure that will be used in this thesis. Specifically, we will calculate the spectral peak in terms of Equivalent Rectangular Bandwidths, as was done by Holliday et al. (in preparation). This is based on Moore and Glasberg's (1987) model of the auditory system. Section 2.4 presents methods that will be used to characterize child-specific differences in the robustness of the contrast between /s/ and /ʃ/. These include measures from logit mixed-effects regression models, including individual-subjects slope measures and the percent correctly predicted in the model (Holliday et al., In Preparation), and discriminability

(Holliday et al., In Preparation; Romeo, Hazan, & Pettinato, 2013) Each of these measures will be explained in more detailed below.

2.3.1 Spectral Moments Analysis

Spectral moments analysis is an procedure in which statistical moments are used to characterize the shape of a power spectrum. If a power spectrum is treated as a random distribution of numbers, then the summary statistics that that characterize any random distribution of numbers can be used to describe the shape of the spectrum. The first moment (M1) or *centroid frequency* represents the mean frequency of the noise spectrum of the fricative. There is an inverse relationship between the length of the oral cavity anterior to the point of constriction and M1 (Jongman, Wayland, & Wong, 2000; Li, 2012). The second moment or M2 measures how widely dispersed the noise is over the spectrum or standard deviation of the spectrum (Li, 2012). The third moment (M3) or skewness calculates the amount of energy above and below the mean frequency while the fourth moment (M4) measures how peaked or flat the fricative spectrum is and is equivalent to its kurtosis (Romeo et al., 2013).

Nissen and Fox (2004) and Miccio et al. (1996) conducted spectral moments analyses of fricatives productions in children and adults. They observed that all four measures distinguished between the sibilant fricatives produced by children. This is in contrast to Jongman et al. (2000) who observed significant differences for only M1 (spectral mean), M3 (skewness), and M4 (kurtosis) when comparing sibilant fricatives produced by adults. Nittrouer (1995) obtained an age related result for M1 and M3. Adults showed greater differences in M1 values for sibilant fricatives than children did. Similarly, M3 values for sibilant fricatives were less positively skewed in adult production compared to child productions. Even at the age of seven, Nittrouer (1995) found that children did not produce a constriction for [s] that was narrower than that of [ʃ]. While many spectral moments may differ between /s/ and /ʃ/, two recent studies showed centroid

values alone are sufficient to distinguish between these two sounds (Li, Edwards & Beckman, 2009; Todd, 2009).

Finally, Li (2012) compared sibilant fricative productions in English and Japanese-speaking children aged 2-5 years to productions of adult speakers of those same languages. In her analysis, Li found that 35-month old English-speaking children's M1 values showed significant differentiation whereas Japanese-speaking children showed differentiation in a different phonetic parameter, F2 frequency of the following vowel at its onset. Li (2012) found that English-speaking children tended to produce sibilant fricatives closer to [s] than to [ʃ] whereas Japanese-speaking children tended to produce more [ʃ]-like tokens. Li attributes this to the frequency with which a child encounters each sound in their native environment. For English, [s] is about 6 times more frequent than [ʃ] (Edwards & Beckman, 2008). Similarly, English-speaking children are typically perceived to acquire [s] correctly and to make more [s] for [ʃ] substitutions.

2.3.2 Peak Equivalent Rectangular Bandwidth

As described by Reidy (2011), psychoacoustic experiments of hearing performed by Fletcher (1940) and Moore (1997) have determined that the human basilar membrane is unable to separate a given frequency component from a nearby component. As such, in order to model sibilant fricatives in a psychoacoustically valid way, an appropriate auditory measure must be developed as part of the auditory model that represents the basilar membrane as closely as possible (Reidy, 2011). The basilar membrane acts like a bandpass filter, or system that accepts a certain range of frequencies while excluding other frequencies outside this range. Important components of a bandpass filter are the *center frequency* and the *equivalent rectangular bandwidth (ERB).* The center frequency is the mean frequency within the range. The ERB refers to the area beneath the frequency curve as measured in rectangular bandwidths divided by the maximum value. Thus, if we take a spectrum and divide it into ERBs, calculate the loudness of each ERB and characterize

the shape of the ERB spectrum by picking its peak, we have a psychoacoustically plausible fricative measure.

2.4 Robustness of Contrast Measures

In previous sections, it was argued that models of phonological acquisition that relied upon phonetically transcribed data showed one pattern of development while studies that employed more granular acoustic measurements detailed another pattern of development. To further illustrate the need for measures of robustness of contrast, let us take a hypothetical example. Imagine that we have fictional data from two children, Annie and Brent. Brent's productions of /s/ and /ʃ/ are quite distinct from one another as characterized by the hypothetical peak ERBs shown for Brent in Figure 2 on the right. There is very little overlap between the two curves. Annie's productions of /s/ and /ʃ/, as shown on the left in Figure 2, show a greater degree of overlap of peak ERBs. Perceptually it would be much more difficult to discern a difference between Annie's sibilant fricatives in isolation than it would be for Brent's productions. If we were to rely upon phonetic transcription, our data would be highly suspect. If we were to use acoustic measurements to capture the robustness of the contrast, how can we reliably capture the difference between Annie's and Brent's respective patterns. Although acoustic measurements are objective by nature, there is no single measure that is appropriate for capturing the separation. Fortunately, a number of candidate measures exist. In this section, I will review these measures and explain their merits.

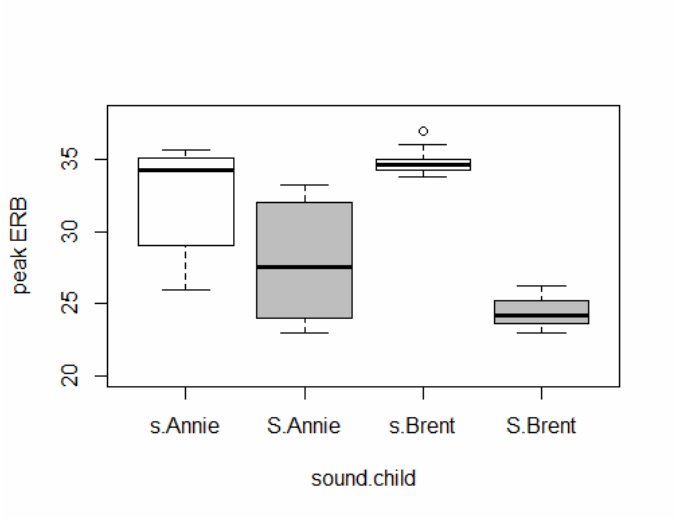

Figure 2. Fictional data portraying children with distinct and indiscriminable sibilant tokens

Holliday et al. (2010; In Preparation) proposed using peak ERB to measure the robustness of contrast in sibilant fricatives in children younger than 4-years of age. Sovinski (2011) validated the peak ERB measure by using a VAS paradigm to generate goodness ratings of sibilant productions. The robustness of contrast measure significantly predicted the goodness ratings for [ʃ] but not for [s] as judged by twenty college-aged adults.

2.4.1 Individual Slope Measure

Another robustness of contrast measure is the individual-slope measure as developed by Holliday et al. (In Preparation) for a generalized linear mixed model of regression. The Individual-slope measure plots peak ERB values against whether the target is [s] or [ʃ]. Thus, one would expect a steep sigmoidal plot between low Peak ERB values for [ʃ] and high peak ERB values for [s]. The regression model developed by Holliday et al. employs the `lme4` package described in Bates et al. (2013). Initially, there is a group-level comparison for [s] and [ʃ] peak ERB values. Each individual participant can be fit to the model by comparing individual adjustments to the group.

When applied to a data set of 80 child and adult talkers, Holliday et al. (In Preparation) found that separation between sibilant fricatives increased with age. Effects also differed based on the gender of participants, as adult females showed elevated peak ERB values for [ʃ] whereas adult males showed decreased values.

2.4.2 Percent Correctly Predicted

Once the generalized liner mixed model of regression has been built, the same data set can be presented to the model to see what predictions it makes about the identity of each target production (Holliday et al., In Preparation). Essentially, once each participant has an independent model, their data can be used to make a prediction about a given phoneme. For example, the model may output 1 if a token is predicted to be [s] and 0 if a token is predicted to be [ʃ]. Holliday et al. (In Preparation) then calculate the percent of tokens that were correctly predicted by the model and label this value as %CP. Thus, if the model were to predict all tokens with 100% accuracy, one would expect a %CP value of 1.0.

Using %CP on the same data set described above, Holliday et al. (In Preparation) found that as participant age increased, the separation between peak ERB values for /s/ and /ʃ/ also increased. Significant differences were found between %CP values of adults and children. Unlike for the individual-slope measure, gender differences were not obtained using %CP.

One advantage to the robustness of contrast measures developed by Holliday et al. is that they appear to be perceptually relevant. Sovinski (2011) conducted a visual-analog scaling study in which adults listened to these tokens and provided goodness ratings of sibilant fricatives for children under age 4-years old. Sovinksi (2011) found that the slope measure from Holliday et al. (In Preparation) is a perceptually valid measure for /ʃ/. Li et al. (2011) suggested that adults may have a larger perceptual space for /s/ than for /ʃ/. This may explain why adult listeners found it more difficult to distinguish between within-category differences for /s/ but were able to make

these distinctions for /ʃ/. Results of a VAS study that compared goodness judgments with robustness measures deemed that perceptually %CP was a reliable predictor of goodness (Holliday et al., In Preparation).

2.4.3 Discriminability

Another measure that has been used to model the robustness of contrast between fricatives is discriminability *d(a)* (Holliday et al., 2010; Holliday et al., In Preparation; Romeo et al., 2013). Discriminability is a concept from signal detection theory (McMillan and Creelman, 1996) that is widely used across academic disciplines. Essentially, for fricatives, *d(a)* is a measure of the amount of difference between the noise spectra of two tokens divided by the amount of dispersion or square root of the mean of the variance (Romeo et al., 2013). In Holliday et al. (In Preparation) *d(a)* is calculated by computing the distance between mean peak ERB values of [s] and [ʃ] divided by square root of the mean peak ERB variance for either [s] or [ʃ], respectively.

Holliday et al. (In Preparation) applied the discriminability measure to their data set. As observed with %CP measure, there was a significant effect for age. Peak ERB values were significantly different between adults and children. Gender differences were also found using the discriminability measure which showed that *d(a)* values were higher for adult females than adult males. A regression of discriminability against goodness ratings in a perceptual study revealed significant effects for [s] but less well for [ʃ]. Romeo et al. (2013) used the discriminability measure to compare the fricative productions of children aged 9-18 with adults. Children between the ages of 14 and 18 exhibited the same level of discriminability as adult speakers while younger children in the 9-10 year range showed less discriminability between contrasts. Romeo et al. also found a significant effect of gender with males showing less discriminability in fricative contrast than females.

2.5 Summary and Motivation for Present Study

Just as a child acquires knowledge about the world from a variety of different domains, language acquisition also occurs across different domains and at many levels of abstraction away from raw sensory experiences. One of the mechanisms that drives the acquisition of one type of phonological knowledge is lexical acquisition (Beckman and Edwards, 2000; Pierrehumbert, 2003; Vihman & Croft, 2007). Beckman and Edwards (2000) put forth a call for age-specific methodologies and models of acquisition in order to understand how children piece together phonological categories. While several studies have examined the robustness of contrast in sibilant fricatives, few studies have done so in a longitudinal fashion for children starting with children as young as 28 months. In this thesis, I will explain a methodology and some preliminary results that could be used to address these questions.

Studies of the robustness of the contrast between sibilant fricatives have utilized a number of measures to examine these differences. In the previous section, I have reviewed Individual Slope, %CP, and *d(a)*. When used to model sibilant fricatives, all three measures have been shown to increase significantly with Age for an older cohort of children (Holliday et al., In Preparation; Romeo et al., 2013). Romeo et al. (2013) found gender-related effects using the discriminability measure but Holliday et al. (In Preparation) contest this result suggesting that discriminability is a less stringent measure. In subsequent chapters, I will use these measures as potential descriptors of sibilant fricative data and will comment on their utility for the overall methodology of an analysis of turbulence.

In subsequent chapters, I will describe a protocol to analyze the sibilant fricatives obtained in the a real word repetition experiment in an ongoing longitudinal study of phonological development and lexical growth. In this study, participants were aged 28 to 39 months and as such were younger than any of the participants in previously described research. Some preliminary results will be presented across the measure types of peak ERB, slope, %CP and

discriminability. These data will be used to a) explain a methodology developed to study contrast between sibilant fricatives, b) discuss possible ramifications that each measure may have on a data set and c) discuss preliminary implications for a model of child phonological acquisition that we intend to pursue once all sibilant fricatives have been analyzed in the Learning to Talk database.

3 Methods

The data from thirty-nine participants were initially chosen for the analysis of sibilant fricatives. All participants were originally recruited for the Learning to Talk project from the area surrounding the University of Minnesota and University of Wisconsin-Madison campuses. Participants came in for two-hour sessions on separate days and were provided with reinforcement (stickers, toys, books) and breaks as necessary. A participant typically took two or three two-hour sessions to complete the measures needed for the first time point. Each visit happened on a different day and were generally within a month of the regular visit.

The Learning to Talk protocol consisted of measures of perception (Minimal Pair Discrimination, two Eyetracking tasks, Peabody Picture Vocabulary Test, a Hearing Screening, an Executive Function task) as well as measures of production (Real Word Repetition, Nonword Repetition, Expressive Vocabulary Test, Goldman-Fristoe Test of Articulation, and a Verbal Fluency task). The order of these tasks was quasi-randomized. For the children included in this study, the Real Word Repetition task occurred approximately an equal number of times in the first, second or third sessions.

Given that a goal of this study was to develop a measure that would be useful for children with a wide range of fricative production abilities, participants with a wide range of language abilities were chosen. Participants were chosen for this analysis based on their standardized scores on the Expressive Vocabulary Test (EVT). The EVT provides standardized scores for

children older than 30 months. Two of the participants were younger than 30 months when the EVT was administered. This fact was discovered after data analysis had begun. For this reason raw EVT scores were used. Five bins that designated ranges of scores were demarcated. Range 1 included five participants who scored between 0-14, Range 2 included fourteen participants who scored in the range of 15-29, Range 3 included six participants who scored in the 30-44 range, Range 4 included ten participants who scored in the range of 45-59, and Range 5 included two participants who scored in the 60-74 range.

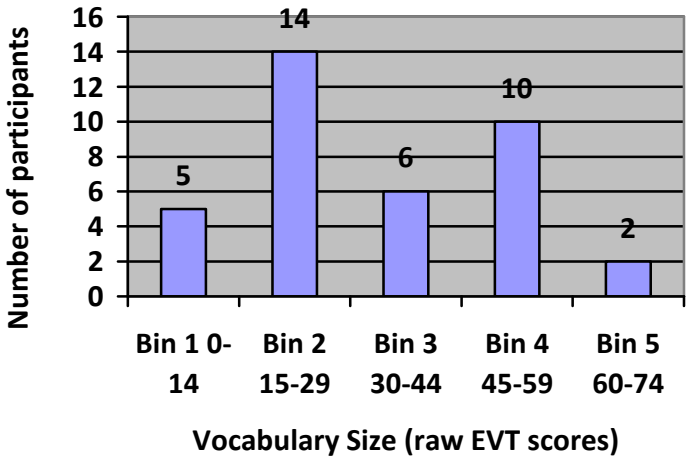

Figure 3. A histogram showing subject distribution by raw EVT score

Figure 3 depicts the distribution of participants by age and gender. The 39 participants ranged in age from 28 months to 39 months. Gender was nearly balanced: the data from twenty males and nineteen females was chosen for analysis. There were thirty Mainstream American English speakers and nine African American English speakers.

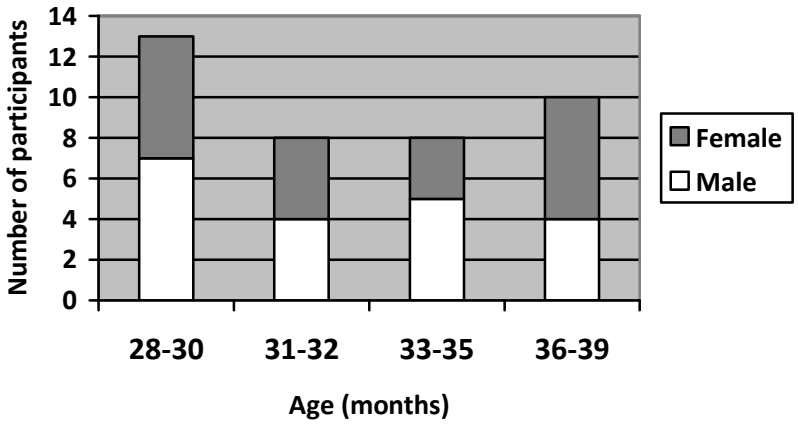

Figure 4. A histogram showing participant distribution by age and gender

3.1 Real Word Repetition Experiment

As part of the Learning to Talk battery of tasks, each participant completed a Real Word Repetition experiment that was presented via EPrime 2.0. In this task, the child was seated in front of a Planar HDMI PXL2430MW 24-inch touchscreen monitor approximately 60 centimeters and was requested to repeat 95 test items that he/she heard over Klipsch BT77 speakers at the University of Minnesota or Audix PH5 at the University of Wisconsin-Madison. A visual image accompanied each target word. For example, if the test item was "DOG", a picture of a golden retriever was displayed. Prior to the test items, the participant was trained on the nature of the task with four familiarization items (e.g. SHORTS, GIRL, COW and COLD). The experimenter attempted to elicit a production of the target word from the child. If possible, the experimenter was instructed to avoid saying the target word while prompting the child. During the familiarization period, the experimenter was not permitted to replay the auditory stimulus. During the experimental phase, the experimenter could repeat the test item up to two times before they moved on to the next item.

A visual reinforcement was used to help the child track his or her progress throughout the task. The participant chose an animal (e.g. hedgehog, bird, rabbit, frog) from an array. This animal then ascended through three colored bands on a ladder to demarcate how many more trials the participant needed to complete before the end of the experiment. Once the animal reached the top of the ladder, a recording of that animal's noise was played as a reward. In some cases, participants were reinforced with the option to place toy fish into a fish bowl after they repeated a certain number of test items. Some participants completed the task with their parent in the testing booth. Parents were instructed to avoid using the target items, if at all possible.

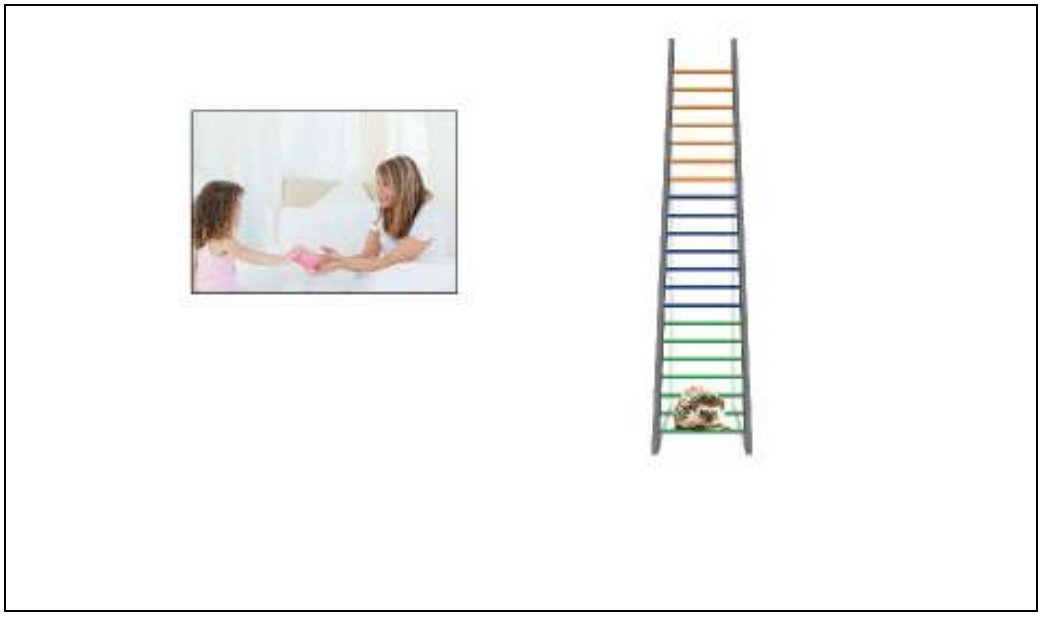

Figure 5. A screenshot of the visual reinforcement used in the Real Word Repetition experiment

The audio stimuli for this experiment were recorded by the author and included in the EPrime setup at both campuses. Laboratory staff selected the most representative tokens to include in the experiment. All stimuli were normalized to be presented at 70 dB. Appendix A lists all of the stimuli included in the Real Word Repetition experiment. Seven of these stimuli (*Dog,*

get, give, sheep, shoe, and *share*) were presented four times, one stimulus was presented three times (*duck)* and thirty-six of stimuli were presented twice. This was accomplished by dividing the stimuli across blocks in the EPrime script to prevent consecutive repetitions of the same target items. Of the 34 files analyzed in this experiment, ten of these files originally had consecutive items. The details of these files are included in Appendix A.

Visual stimuli in the experiment were color photographs chosen from databases of stock images. Each picture was approximately 5 inches tall by 7.5 inches wide. Depending on the original size and pixilation of each image, the dimensions of each image could vary. All visual stimuli were presented against a white background.

3.2 Segmentation

During the analysis phase, each Real Word Repetition file was segmented by a trained member of the Learning to Talk lab. Segmentation was completed using Praat software (Boersma & Weenink, 2013). A script was created to segment each production. The script loaded the relevant Real Word Repetition recording and created a TextGrid or text document that was time-aligned with the recording. The script then prompted the segmenter to locate the first Familiarization production in the waveform and select it for segmentation. The script then inserted boundaries in the TextGrid and asked the segmenter to label the production as either a Response, VoicePromptResponse, UnpromptedResponse or NonResponse. Segmenters continued until all 99 items were segmented.

Segmenters were prompted to decide whether a production was a Response, VoicePromptResponse, UnpromptedResponse or NonResponse. A Response was defined as an attempt at the target item. In a VoicePromptResponse, the child makes an attempt at the target item after being prompted by the experimenter, parent or another adult (e.g. "*Say dog"*). A production was considered an UnpromptedResponse if the child produced the target without an

auditory stimulus. For example, a child might repeat several consecutive productions of the same target word after the initial prompt (“*duck duck duck”)* or the child might whisper the first production and then be asked by the experimenter to speak louder. Finally, in a NonResponse the child did not make an attempt to say the target item or said something that was unrelated to the target item (e.g. “*I want a drink”*).

After a file was completely segmented, it was checked for accuracy by the experienced segmenters. The experienced segmenter made changes to the TextGrid as necessary and stored both copies of the file in separate locations. The revised TextGrid was then deemed ready for Turbulence tagging.

3.3 Turbulence Tagging

Turbulence tagging was also completed in Praat software via a script specially designed for this purpose by Patrick Reidy. Initially, the author consulted a turbulence tagging protocol written for the Paidologos project by Syrika and Li (2009). A copy of the protocol developed for the Learning Talk project is included in Appendix B.

The author and one specially trained student completed turbulence tagging. The script located only the test items that began with a sibilant fricative. The tagger was then prompted to listen to the target item and make a judgment about whether the production was a sibilant fricative, sibilant affricate, non-sibilant fricative, non-sibilant plosive or other (e.g. an approximant). If the tagger labeled a production as a sibilant fricative or sibilant affricate, the script then prompted the tagger to insert point tiers that corresponded with the onset of turbulence (turbOnset), voicing onset time (VOT) of the vowel, and the end of the vowel (vowelEnd). Taggers inserted turbOnset tags at the beginning of aperiodic noise in a band above 1000 Hz. VOT tags were inserted at the first glottal pulse following the period of frication. In some cases there was a hiatus between the offset of turbulence and VOT. In these cases, the tagger could

insert a point at the turbulence offset (turbOffset). If the target production began with another consonant besides a sibilant fricative, the tagger could insert a consOnset tag to denote consonant Onset. Finally, a vowelEnd tag was inserted at the point where the second formant ended.

As detailed in the protocol in Appendix B, challenging cases can be divided into two types: production related cases where some element of articulation is involved and environmental cases where an event not under the participant's control occurred. Production-related challenges included a short vowel prior to the onset of the fricative (e.g. /əsık/ for *sick*). These cases were handled by inserting a consOnset tag at the onset of the vowel and a turbOnset tag at the onset of /s/. A second challenging case included epenthetic stops that occurred between the initial consonant-vowel pair of the target. These were handled by tagging the onset of turbulence with turbOnset, marking the end of turbulence with turbOffset and then inserting a VOT annotation at the onset of the vowel. The epenthetic stop was located in the hiatus between turbOffset and VOT. The same procedure was used on productions where there was a silent hiatus between the end of turbulence and the onset of a vowel. Another production-related challenge involved fricated vowels or cases where the turbulence persisted after the onset of the vowel.

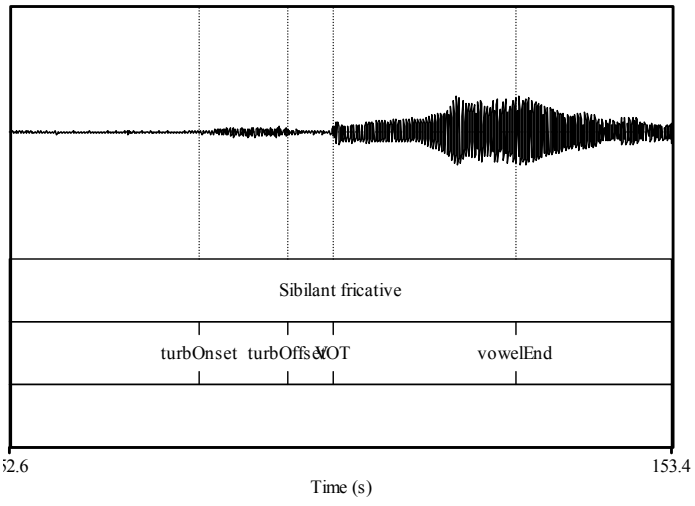

Figure 6. Example of a token "*shower"* that had been tagged for Turbulence

In such instances, the tagger inserted a VOT tag at the first peak indicating the onset of the vowel. If the speech waveform clipped because the participant used a loud voice to say the production, the turbulence was tagged as usual. Productions with epenthetic stops, initial epenthetic vowels, silent periods, and clipped or fricated vowels were included in present analyses. Finally, there were occasionally emerging fricatives where the constriction changed during production (e.g. the child starts with /θ/ but ends with /s/). These cases were labeled as non-sibilant fricatives and were not included in present analyses.

Environmental challenges included static or buzzing noise in the acoustic signal. In this case, the tagger inserted a "BackgroundNoise" annotation. If the static prevented the tagger from discerning between a sibilant and non-sibilant fricative, the tagger defaulted conservatively towards non-sibilant fricative. In the event of overlapping speech between the participant and an adult in the room, an "OverlappingSpeech" tag was inserted. These cases were excluded from the analyses.

Results in this thesis were taken from 35 participants. Of these 35 files, 28 or 80% were tagged by the author while the remaining seven were tagged by Rose Crooks, an undergraduate employee in the lab. Training was completed by first walking Rose through the Turbulence Tagging manual available on the Learning to Talk Wiki space. The author and Rose then independently tagged two files, one easy file and one difficult file, and discussed discrepancies until they were in agreement. For the easy file, agreement between the two coders with respect to consonant type (e.g. *sibilant fricative, non-sibilant fricative, non-sibilant plosive, other)* was 100%. Both coders labeled each of the 33 target productions as "sibilant fricative". For the more challenging file, agreement between the coders with respect to consonant type was 76.5%. This was deemed acceptable agreement between coders.

3.4 Analysis of Turbulence

Following the completion of tagging, the acoustic analysis of turbulence was conducted using scripts written in R by Patrick Reidy, Mary Beckman and Jeffrey Holliday and used in Holliday et al. (In Preparation). In the first script, buildRWRdataframe-sibilants.R, a data frame was plottcreated by looping through and entering in information for the tokens labeled as sibilant fricative or sibilant affricate. Next, the extractRWRspectra-sibilants.R script inserts a column into the data frame containing the peak ERB values from the middle 40 milliseconds of every target sibilant fricative. Finally, the robustnessRWRsibilants.R script links the peak ERB data with the EVT and other information for each participant. This script then produces box plots of the descriptive statistics and completes a logistic regression to see whether it is possible to predict if the target was /s/ or /ʃ/ given the peak ERB values.

4 Results

4.1 Descriptive Statistics

4.1.1. Age

The Real Word Repetition experiment attempted to elicit 33 sibilant fricative targets from each of 39 participants. Four participants were removed from data analysis because they produced only one or zero sibilant fricatives. Overall, the 35 participants in this experiment produced a total of 520 alveolar sibilant fricatives /s/ and 601 post-alveolar fricatives /ʃ/ for a total of 1121 sibilant fricative tokens. As previously mentioned, the participants ranged in age from 28 months to 39 months. Figure 7 shows the proportion of target items that were tagged as sibilant fricatives across age groups.

Next, we calculated peak ERB values for each target tagged as a sibilant fricative. Since there were an unequal number of tokens from each participant, we calculated the median peak ERB values for [s] and [ʃ] targets by subject. The distribution of median peak ERB values plotted against age quartile groups is shown in Figure 8. These data show that the difference between peak ERB values for /s/ and /ʃ/ increased for children in the older age groups. Largely, this was due to a decrease in the peak ERB frequencies for [ʃ]. As evident in Figure 8, older children had lower peak ERB values for [ʃ] compared to younger children. At age 28-30 months, peak ERB was roughly 32 Hz whereas for the children aged 36-39 months of age peak ERB for [ʃ] decreased to 28 Hz. Conversely, peak ERB for [s] remained at approximately 33 Hz across the age range. One possible explanation is that as children age, they begin to gain more control over their articulators and are better able to mark the contrast between [s] and [ʃ].

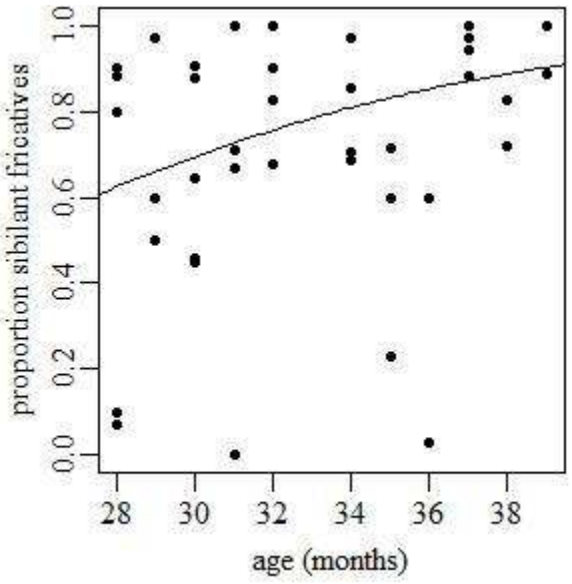

Figure 7. The proportion of targets tagged as sibilant fricatives plotted against Age

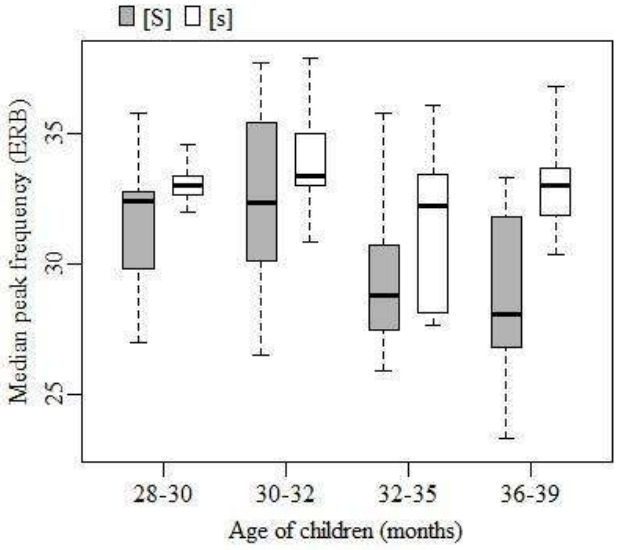

Figure 8. Boxplot showing Median peak ERB (Hz) of sibilant fricatives [s] (white) and [ʃ] (grey) against Age of children (months). Peak ERB values were lower for [ʃ] in the 36-39 month age range compared to 28-30 months.

4.1.2. Vocabulary Size

Expressive vocabulary scores were obtained for each participant from the EVT. Raw scores ranged from zero to 70 across the 35 participants. The proportion of targets tagged as sibilant fricatives on the Real Word Repetition experiment were plotted against raw EVT scores. These data did not reveal any statistically significant results ($p = 0.15$) but are nevertheless depicted in Figure 9. In Figure 10, median peak ERB frequency was plotted against vocabulary size. As for age, [s] remains relatively constant between 30 and 35 Hz.

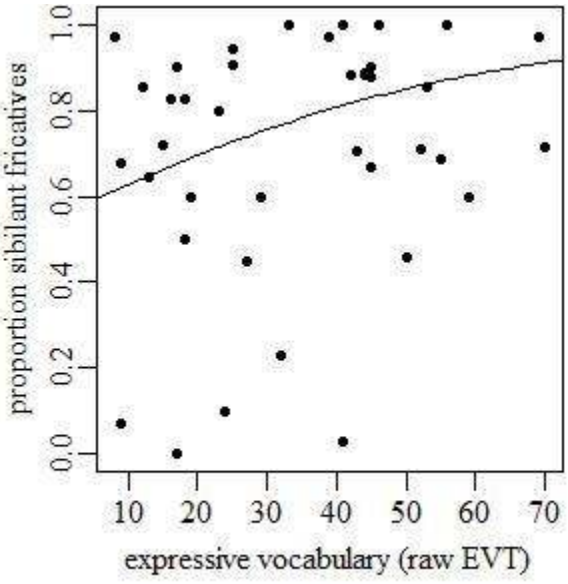

Figure 9. The proportion of targets tagged as sibilant fricatives plotted against expressive vocabulary (raw EVT scores).

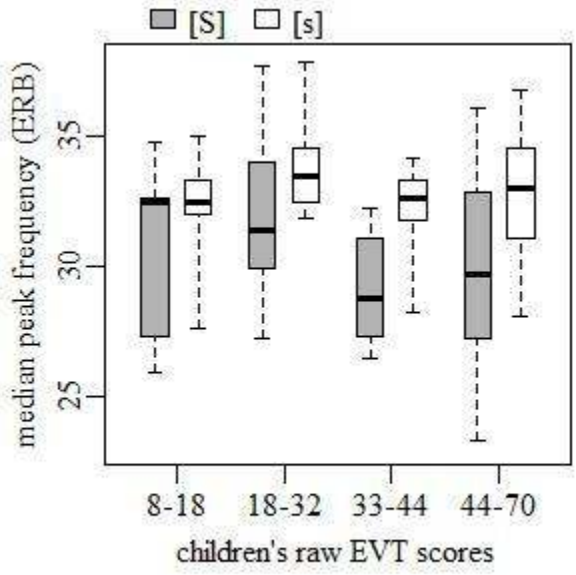

Figure 10. Boxplot showing Median peak ERB (Hz) of sibilant fricatives [s] (white) and [ʃ] (grey) against raw EVT scores. Peak ERB values were lower for [ʃ] in the 44-70 EVT score range compared to the 8-18 month range.

4.2 Robustness of Contrast Measures

4.2.1. Age

Recall from Section 2.4.1 that the glmer() model first predicts whether a target token was [s] from centered peak ERB input and random slopes. Variables are added that include individual intercepts and slopes, log odds based on those individual slopes, a variable for whether the prediction is correct, and the total slope. Next, %CP or percent correct is added to the data. When %CP was plotted against age, %CP gradually increases from age 28 months to 38 months ($p<.05$). Holliday et al. (In Preparation) also found a significant effect of Age, albeit for a much older population. Given that the current data set only included 35 participants, it is best to interpret the results with some degree of caution. Next, a mixed-effects model that plots individual-slope against age did not reveal a significant result for age ($p = .07$). The results in Figure 12 show only a slight increase in the individual-level slope as children age but this was not significant.

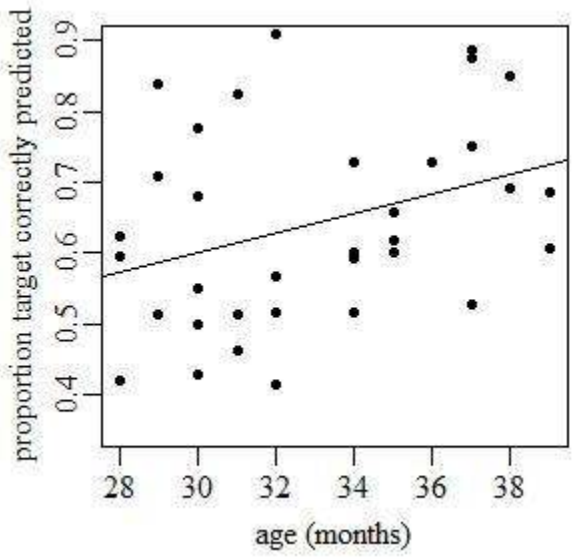

Figure 11. A scatterplot showing %CP for sibilant fricatives plotted against Age (months). The trend line shows that %CP increases significantly along with Age.

	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	0.189115	0.222724	0.849	0.4019
age	0.013738	0.006714	2.046	0.0488

Table 1. Output results of the %CP mixed-effects model of Holliday et al. (In Preparation) for 35 Learning to Talk participants. Age is a significant predictor of [s] versus [ʃ] at the 0.05 level.

Recall that Holliday et al. (In Preparation) found significant results for age for %CP for an older cohort of children. The results presented here fail to confirm this for children aged 28 to 39 months of age.

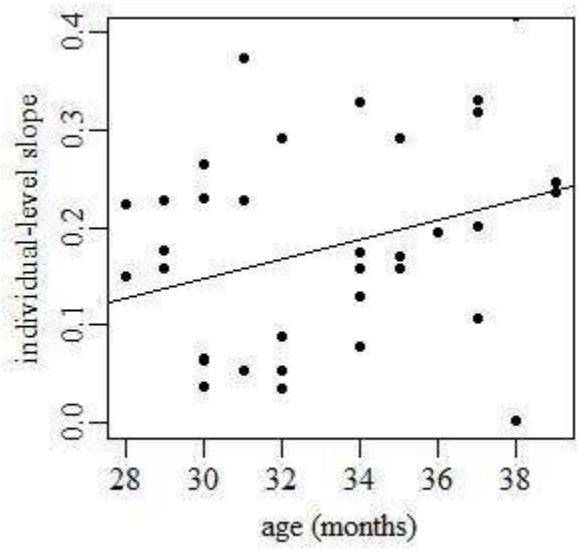

Figure 12. Scatter plot of individual-level slopes for peak ERB (BLUPs) from the logistic regression lmer as a function of the child's age in months

	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	-0.155440	0.181362	-0.857	0.3976
age	0.010075	0.005467	1.843	0.0744

Table 2. Output results of the individual-slope mixed-effects model of Holliday et al. (In Preparation) for 35 Learning to Talk participants.

4.2.2. Raw Vocabulary Score

The glmer() model for %CP and individual-slope were then plotted against vocabulary size as measured by raw EVT scores. As shown in Figure 13 and Table 3, %CP were not significant ($p > .05$) predictors for this sample subset of the data. The data for individual slope results are shown in Figure 14 and Table 4 and also reveal no significant findings. One possible explanation for these results may be the fact that vocabulary size was not properly controlled when subjects were chosen.

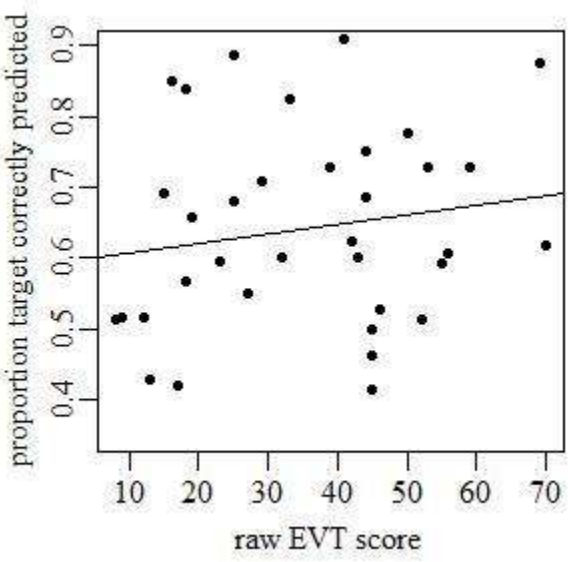

Figure 13. A scatterplot showing %CP for sibilant fricatives plotted against raw EVT score. The model was not significant.

	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	0.594834	0.054406	10.933	1.66e-12
Raw EVT	0.001348	0.001386	0.972	0.338

Table 3. Output results of the %CP mixed-effects model of Holliday et al. (In Preparation) for 35 Learning to Talk participants. Vocabulary size (raw EVT) is a not a significant predictor of [s] versus [ʃ] at the 0.05 level.

Holliday et al. (In Preparation) found that %CP was correlated with receptive vocabulary scores but that %CP was not correlated with expressive vocabulary. While receptive vocabulary scores were not included in the model, these results provide further support for the null effect observed by Holliday et al.

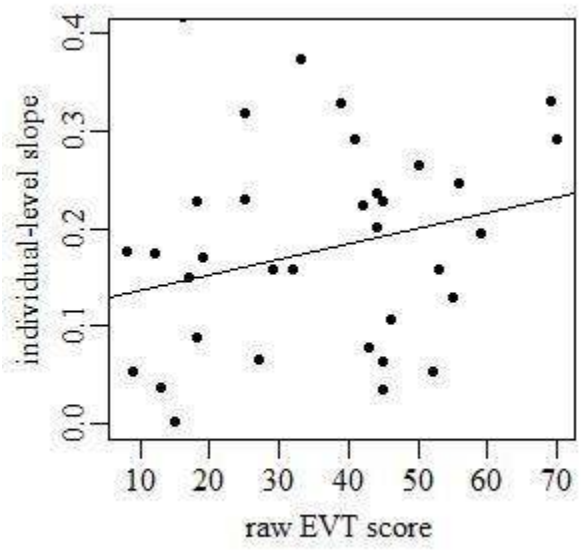

Figure 14. Scatter plot of individual-level slopes for peak ERB (BLUPs) from the logistic regression lmer as a function of the child's vocabulary (raw EVT)

	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	0.121244	0.043125	2.811	0.00824
Raw EVT	0.001578	0.001099	1.436	0.16035

Table 4. Output results of the individual slope mixed-effects model of Holliday et al. (In Preparation) for 35 Learning to Talk participants. Vocabulary size (raw EVT) is a not a significant predictor at the 0.05 level.

4.3 Discriminability

The discriminability measure from Romeo et al. (2013) was also applied to these data. The regression with Discriminability as dependent variable and Age as an independent variable also revealed a significant association. Discriminability, which measures how large the variance is between [s] and [ʃ], increased significantly with age ($p < .02$). These results suggest that the older children in the selected study had a more robust contrast between fricatives than the younger children.

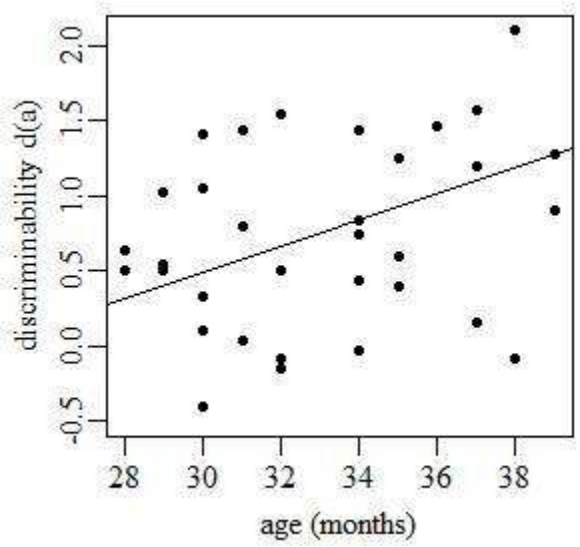

Figure 15. A scatterplot showing *d(a)* against Age (months). Age was found to increase significantly with *d(a)*

	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	-2.16764	1.17153	-1.850	0.0732
Age	0.08844	0.03532	2.504	0.0174 *

Table 5. Output results of the discriminability model. Age is a significant predictor at the 0.05 level.

Discriminability also increased significantly with vocabulary size ($p = .02$). Children who had a larger expressive vocabulary as measured by the EVT were found to show a more robust contrast between [s] and [ʃ].

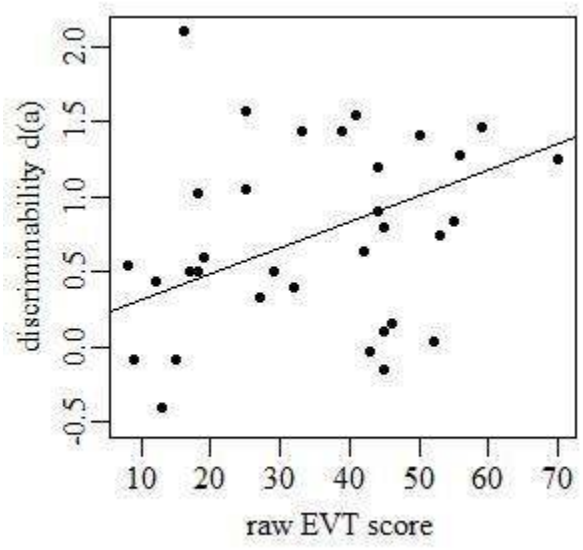

Figure 16. A scatterplot showing *d(a)* against raw EVT scores. Raw EVT scores were found to increase significantly with *d(a)*

	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	0.142363	0.274105	0.519	0.6070
Raw EVT	0.017216	0.006985	2.465	0.0191*

Table 6. Output results of the discriminability model. Vocabulary size is a significant predictor at the 0.05 level

As discussed previously by Holliday et al. (In Preparation), discriminability appears to be a less stringent measure than %CP. As such, these results should be treated with caution. When age and vocabulary size were entered into the same General Linear model, neither Age ($p = 0.229$) nor vocabulary ($p = 0.26$) remains significant. A likely explanation for this fact is that Age and vocabulary size are correlated with one another, and that their simultaneous inclusion in the model cancels out the other's effect.

5 Discussion

The limited data set of 1121 tokens taken from 35 participants was used to develop a methodology for an analysis of sibilant fricatives. The primary purpose of this thesis was to develop a protocol for tagging turbulence in consultation with primary investigators and student employees on the grant. This protocol documented how to handle anomalous productions that included epenthetic stops, fricated vowels, emerging sibilant fricatives as well as environmental challenges such as overlapping speech or static noise in the waveform. The turbulence tagging protocol is general enough that it could be used with any data set of fricatives but conservative enough that tokens were easily selected for analysis.

Preliminary results revealed that the %CP measures was correlated with age while the individual-slope measures was not. Implications of a significant effect for age, if one exists, for this data set would suggest children begin to acquire a contrast between sibilant fricatives by lowering peak ERB values for [ʃ] within this age range. This may be due to an increasing control of articulatory musculature and movement around 30 months of age. Another age-related explanation may rest with how well the child is able to discriminate minimal pair tokens. A future analysis may attempt to plot robustness of contrast measures against measures of discrimination from the Minimal Pair Discrimination task in the Learning to Talk project.

In contrast to age, neither %CP nor individual-slope measures were found to correlate with vocabulary size. A plausible explanation may be the fact that the subject pool was not well-balanced with respect to raw EVT scores. Of course, it may also be the case that the older children who showed a contrast between [s] and [ʃ] were in the process of acquiring new vocabulary items that began with these sounds and as such had more impetus to make the distinction between sounds. To test this possibility in the Learning to Talk data set, a future analysis could incorporate parental reports of expressive vocabulary from the MacArthur CDI to see which sibilant sounds the child says reliably.

As argued by Holliday et al. (In Preparation) %CP is an appropriately conservative measure to use when measuring robustness. Although significant effects were obtained for both age and vocabulary size using the Discriminability measure, there are reasons to be suspect of such results. Discriminability measures the distance between categories while %CP represents how well the model was able to predict the targets. Based on these factors and the fact that only 35 children were analyzed in the present data set, it is best to interpret any significant findings with caution. A future analysis of the full data set of 180 children is currently in progress.

6 Conclusion

In summary, the present thesis has outlined steps taken to analyze turbulence in sibilant fricatives. I have described how scripts were developed by Patrick Reidy to automate the tagging of turbulent intervals. As with any aspect of segmentation or annotation of acoustic data, there are often problematic instances where the tagger is uncertain about how to correctly annotate the data. In the Methods section, I have provided a list of the solutions developed in consultation with other Learning to Talk members to uniformly handle some of the more difficult cases. Finally, this thesis describes three measures used to analyze robustness of contrast and some preliminary results. The %CP measure appears to be the most reliable measure for the reasons offered above.

Bibliography

Bates, D., Maechler, M., and Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2. http://CRAN.R-project.org/package=lme4.

Beckman, M. E. and Edwards, J. (2000). The Ontogeny of Phonological Categories and the Primacy of Lexical Learning in Linguistic Development. *Child Development, 71(1),* 240-249.

Berko, J. and Brown, R. (1960). Psycholinguistic Research Methods. In P. Mussen (Ed.) *Handbook of Research Methods in Child Development,* (p. 517-557). Englewood Cliffs, NJ.: Prentice-Hall

Boysson-Bardies, B. de and Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language,* 67, 297-319.

Edwards, J. and Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes. *Language Learning Development, 4(1),* 122-156.

Edwards, J., Beckman, M. and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research,* 47, 421-436.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague, Netherlands: Mouton.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 419-439.

Forrest, K., Weismer, G., Milenkovic, P. and Dougall, R.N. (1988). Statistical Analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*, 115-124.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics,* 28, 377-396.

Holliday, J.J., Reidy, P.F., Beckman, M.E., & Edwards, J. (In Progress). Quantifying the robustness of the English sibilant fricative contrast in children. Draft in Preparation for submission to *Journal of Speech-Language Hearing Research*

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals.* The Hague & Paris: Mouton

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108(3),* 1252-1263.

Johnson, W. and Reimers, P. (2010). *Patterns in Child Phonology.* Edinburgh: Edinburgh University Press.

Kent, R. and Forner, L. (1980). Speech segment duration in sentence recitation by children and adults. *Journal of Phonetics,* 157-168.

Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages*. Massachussets: Blackwell Publishers.

Lee, S., Potaminos, A. and Narayanan, S. (1999). Acoustic of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America,* 103(5), 1455-1468.

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

Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. *Journal of the Acoustical Society of America, 129(2),* 999-1011.

Li, F. (2012). Language-Specific Developmental Differences in Speech Production: A Cross-Language Acoustic Study. *Child development*, *83*(4), 1303-1315.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In Ferguson, C. E., Menn, L., and Stoel-Gammon, C. (Eds.) *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press.

Mayo, C. and Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased towards transitions. *Journal of the Acoustical Society of America,* 115(6). 3184-3194.

McGowan, R.S. and Nittrouer, R. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *Journal of the Acoustical Society of America, 83,* 229-236.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. *Language production*, *2*, 3-50.
Miccio, A. W., Forrest, K. and Elbert, M. (1996). Spectra of voiceless fricatives produced by children with normal and disordered phonologies. In *Pathologies of Speech and Language: Contributions of Clinical Linguistics and Phonetics*, (Ed.) T. Powell. ICPLA, New Orleans, LA. 223- 236.
Moore, B.C.J. and Glasberg, B.R. (1987). Formulae describing frequency selectivity as a function of frequency and level and their use in excitation patterns. *Hearing Research, 4,* 209-225.
Munson, B. (2004). Variability in /s/ production in children and adults: evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research,* 47, 58-69.
Munson, B., Edwards, J., and Beckman, M. E. (2011). Phonological representations in language acquisition: climbing the ladder of abstraction. In Cohn, A., Fougeron, C. and Huffman, M. (Eds.), *Oxford Handbook in Laboratory Phonology*. (p. 288-309). Oxford University Press.
Munson, B., Edwards, J., and Beckman, M.E. (2005). Phonological knowledge in typical and atypical speech-sound development. *Topics in Language Disorders,* 25, 190-206.
Munson, B., Kaiser, E., Urberg-Carlson, K. (2008). Assessment of phonetic skills in children 3: Fidelity of responses under different levels of task delay. Paper presented at the 2008 ASHA Convention, Chicago, 20-22 November, 2008.
Nissen, S. L., and Fox, R.A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America, 118*, 2570-2578.
Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics,* 20, 351-382.
Nittrouer, S. (1993). The emergence of mature spectral patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research, 36,* 959-972.
Nitrrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America, 97(1),* 520-530.
Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America,* 112(2). 711-719.
Nittrouer, S., Manning, C. & Meyer, G. (1993). The perceptual weighting of acoustic cues change with linguistic experience. *Journal of the Acoustical Society of America,* 94, S18652.
Nittrouer, S. and Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America,* 101, 2253-2266.
Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32,* 120-132.
Nittrouer, S., Studdert-Kennedy, M. and Neely, S. T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *American Speech-Language Hearing Association,* 379-389.
Nycz, J. (2013). New contrast acquisition: methodological issues and theoretical implications. *English Language and Linguistics, 17(2), 325-357.*
Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of Phonology. *Language and Speech.* 46(2-3), 115-154.
Reidy. P. (2011). *New Measures for Old Fricatives*. Unpublished Manuscript. Ohio State University
Romeo, R., Hazan, V., and Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America, 134(5)*, 3781-3792.
Sander, E.K. (1972). When are speech sounds learned?. *Journal of Speech Hearing Disorders, 37,* 55-63.
Scobbie, J.M. (1998). "Interactions between the acquisition of phonetics and phonology". In Papers from the 34th Annual Regional Meeting of the Chicago Linguistic Society, Volume II: The Panels, edited by M.C. Gruber, D. Higgins, K. Olson, and T. Wysocki (Chicago Linguistics Society, Chicago), pp. 343-358.
Smit, A.B., Hand, L, Freilinger, J. J., Bernthal J. E. ., and Bird, A. (1990). The Iowa Articulation Norms Project and Its Nebraska Replication. *Journal of Speech, Language, and Hearing Research*, 779-798.
Smith, B. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics,* 6(1). 37-67.

Sovinski, R. (2011). Perceptual Validation of an Acoustic Robustness of Contrast Measure. *Unpublished Masters Thesis.* University of Wisconsin-Madison.
Stoel-Gammon, C. and Cooper, J. (1984). Patterns in lexical and phonological development. *Journal of Child Language,* 11, 247-271.
Syrika, A. and Li, F. (2009). *Instructions for Aligning Fricatives*. Unpublished protocol for the Paidologos project. University of Wisconsin-Madison.
Templin, M. (1957). *Certain Language Skills in Children,* Vol. 26 (University of Minnesota, Minneapolis), pp. 19-60.
Todd, A., (2009). *Acoustic characteristics of /s/ and /ʃ/ in children with cochlear implants.* Unpublished Master's thesis. University of Wisconsin-Madison.
Trubetzkoy, N. S. (1939). Grundziige der Phonologic. *Travaux du cercle linguistique de Prague*, *1*.
Vihman, M. (1996). *Phonological Development: The Origins of Language in the Child*. Wiley-Blackwell.
Vihman, M. & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, 45(4). 683-725.
Vihman, M. & Keren-Portnoy, T. (2011). The role of production practice in lexical and phonological development – a commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language*, 38, 41-45.
Wellman, B., Case, I., Mengert, I., & Bradbury, D. (1931). "Speech sounds of young children," Univ. Iowa Stud. Child Welfare 5, 1–82.
Whiteside, S.P. & Marshall, J. (2001). Developmental trends in voice onset time: some evidence of sex differences. *Phonetica,* 58(3). pp. 196-210.

Appendix A

This is a list of the stimuli presented in the Real Word Repetition experiment.

WORD	REPETITIONS
cake	2
car	2
cat	2
cookie	2
candy	2
coat	2
cup	2
daddy	2
dish	2
dance	2
dinner	2
dog	4
door	2
duck	3
get	4
good	2
go	2
garbage	2
give	4
gum	2
kitchen	2
kitty	2

sad	2
share	4
sheep	4
shoe	4
shovel	2
shower	2
sick	2
soap	2
sock	2
soup	2
scissors	2
sun	2
sandwich	2
table	2
tape	2
tickle	2
teddy	2
tummy	2
tongue	2
toast	2
tooth	2

This is a list of sibilant fricative targets that were elicited consecutively in the Real Word Repetition experiment.

Subject	Word	Token Numbers
010L	SOCK	76/77
012L	SHARE	68/69
600L	SHEEP	75/76
604L	SANDWICH	98/99
606L	SHOWER	35/36
612L	SHEEP	85/86
613L	SANDWICH	7/8
613L	SOCK	91/92
619L	SHOE	67/68

Appendix B

Turbulence Tagging Protocol
Learning to Talk Project
Fall 2013

STEP-BY-STEP INSTRUCTIONS FOR TAGGING TURBULENCE

0. Obtain Praat from praat.org

1. Open Praat
Load the TurbulenceTagging.praat script written by Pat Reidy. To load the script, go to "Praat" in the Praat Objects window, select "Open Praat script..." and then select the script. As of Sept 2013, this script is stored on Tier2 at this location: Tier2:\DataAnalysis\RealWordRep\TimePoint1\TurbulenceTagging

2. Adjust the Settings for the Spectrogram
From Spectrum-Spectrogram settings use
View range (Hz) 0.0 and 9000.0 for adults but 0.0 and 11000.0 for children
Window length(s) 0.005
Dynamic range (dB) 40.0

3. Follow the prompts
Enter your initials. Select “Continue”.

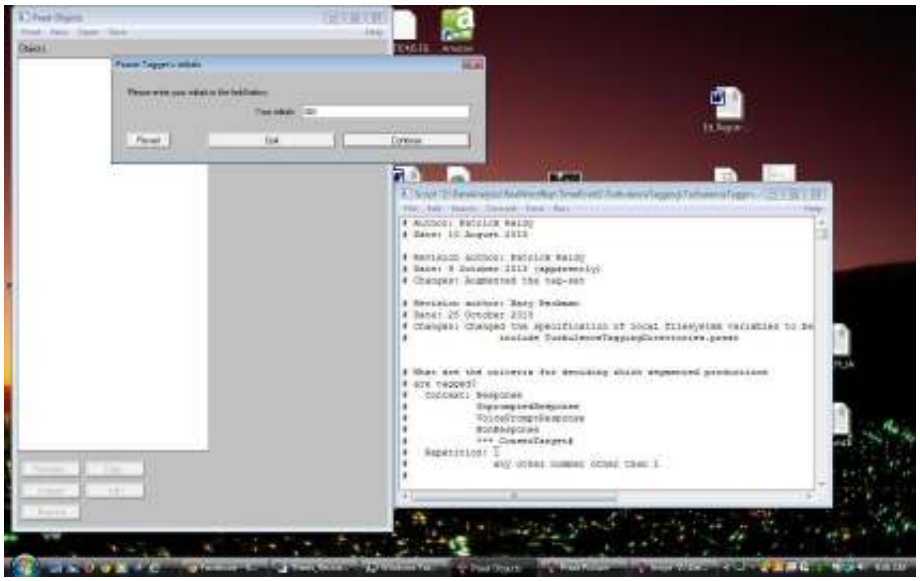

Figure A. Image showing sample initials entered into the user interface

Select a file for tagging from the dropdown menu. Press “Continue”.

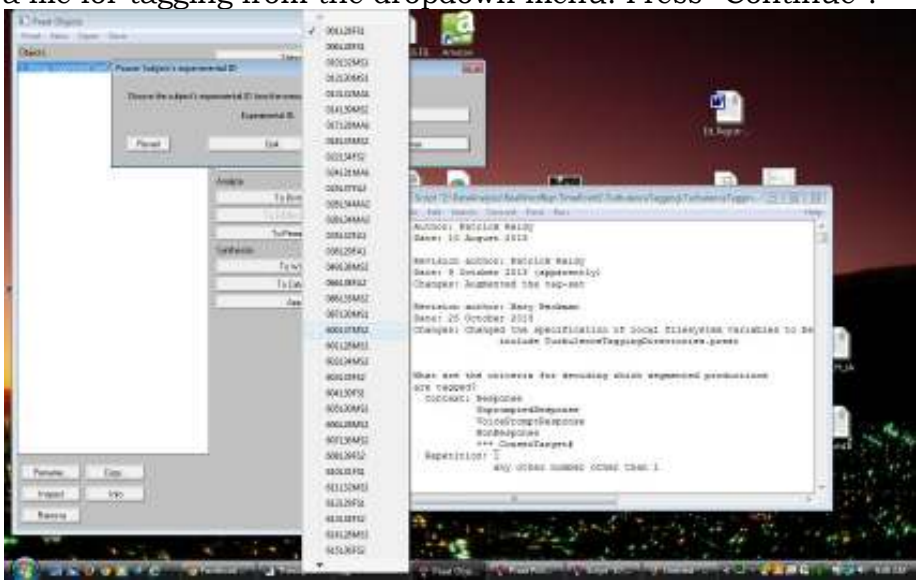

Figure B. Image showing selection of a file to tag

4. Listen to the target word and determine whether the initial consonant is a sibilant fricative, sibilant affricate, non-sibilant fricative, non-sibilant plosive or Other

	What you may observe
Sibilant Fricative	- Period of turbulence above 1000 Hz - Sounds like /s/ or /ʃ/
Sibilant Affricate	- Visible burst in addition to turbulence - Production sounds like /ʧ/ or /ʤ/
Non-sibilant Fricative	- Turbulence is dispersed below 1000 Hz - Production sounds closer to /f/ or /Θ/
Non-sibilant plosive	- Visible burst - May hear a stop in place of fricative
Other	- Absence of turbulence - Hear glide-like production or production begins with a vowel

5. Insert Required Notes

If there is overlap between the target stimulus and either a BackgroundNoise or OverlappingVoice, mark this in the second drop down menu. If the response was a Malaprop, indicate which word was said by typing this in to the third field. See the Troubleshooting guide for more specific instructions.

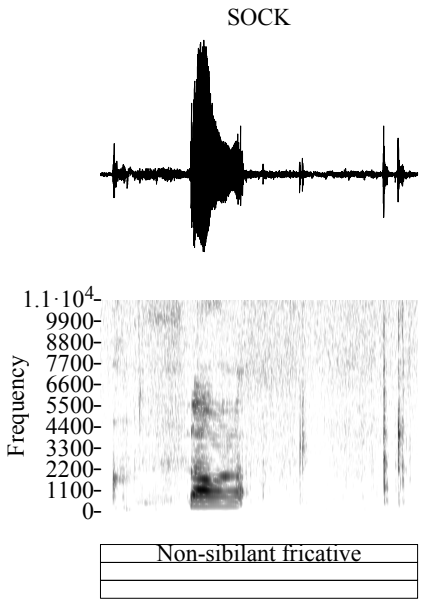

Figure C. Non-sibilant fricative (TH) example
Note how dispersed the turbulence is compared to a sibilant fricative

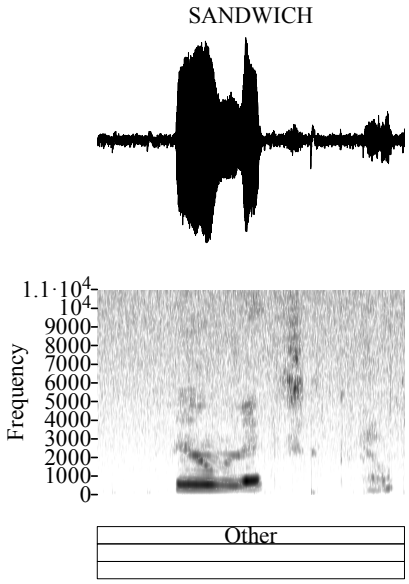

Figure D. Example of OTHER.
Note the absence of a sibilant fricative.

6. Determine if consOnset tag is needed
If you selected "sibilant fricative", the script will ask you if it is necessary to insert "consOnset". The consOnset tag should be inserted when the word begins with another consonant besides the target fricative. This includes cases where there is a clear burst visible in a sibilant affricate. If you would like to insert "consOnset", select "Yes" and then move the boundary to the onset of the consonant.

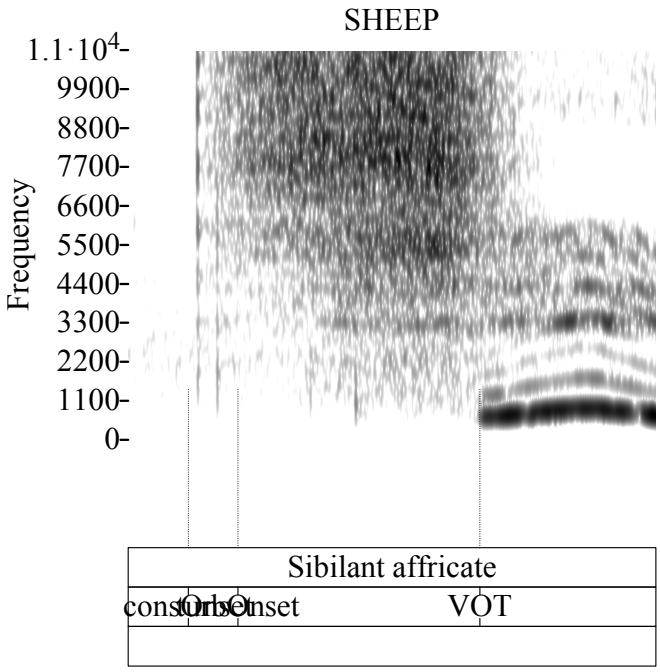

Figure E. Sibilant affricate /ʧ/ example showing consOnset tag prior to burst

7. Insert turbOnset tag
The script will then automatically insert "turbOnset" tag. Move the tag to the onset of turbulence in the spectrogram. Li and Syrika (2009) wrote "Mark the beginning of the fricative at the beginning of aperiodic high-frequency noise characteristic of voiceless fricatives. This is indicated by both clear increase in frication noise in the waveform and by the presence/occurrence of white noise in a frequency band above 1000Hz ".

8. Insert VOT tag
To mark the onset of the vowel, the script will insert a "VOT" tag. Move the VOT tag to "the first zero crossing of the periodic glottal pulses of the vowel. Make sure it is consistently the first zero crossing of an upshooting pitch cycle and one that follows a clear downswing." (FangFangand Asimina's description)

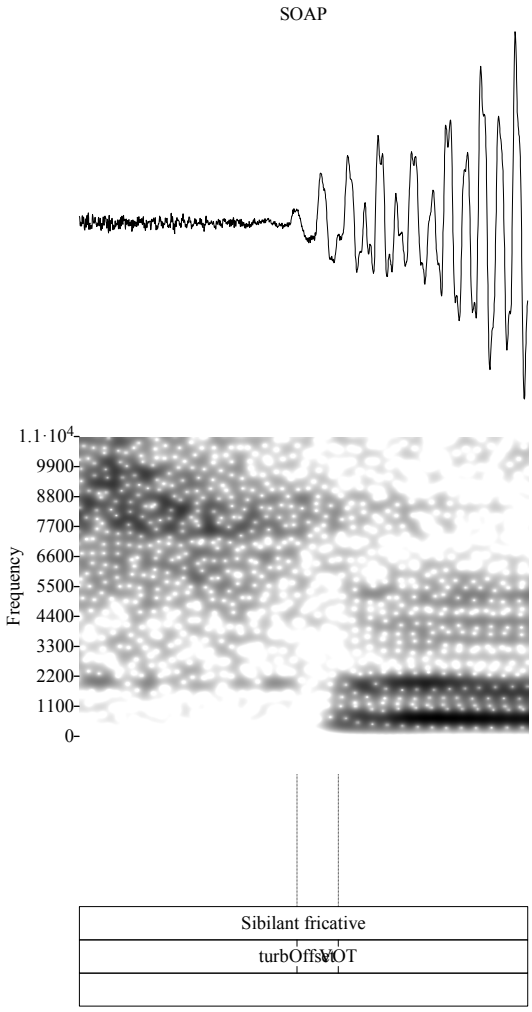

Figure F. Example showing placement of VOT tag

9. Determine if turbOffset tag is required

To mark the offset of turbulence use "turbOffset". The script will ask whether it is necessary to insert a "turbOffset" tag. It can be tricky to determine in cases where there is a period of aspiration or a fricated vowel. See "Criteria for marking the end of a fricative" below" for an explanation for both typical and atypical cases. If you would like to insert "turbOffset", select "Yes" and move the tag accordingly on the point tier.

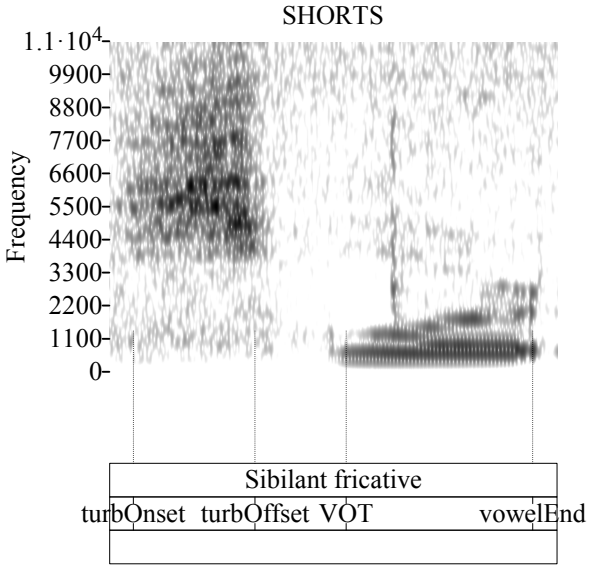

Figure G. Example showing use of turbOffset
Note the period of silence between turbOffset and VOT

10. **Insert vowelEnd tag**
Finally, the script will insert a "vowelEnd" tag. Move the "vowelEnd" tag to the end of the second formant. See "Criteria for marking the end of the vowel" below for more explanation.

11. **Move on to next target**
The script will then proceed to the next target word. Repeat steps 3-9 until you have reached the end of the file.

Troubleshooting for Challenging Cases

1. *Production Related Challenges*

    a. <u>Vowel precedes target consonant</u>

If a child produces a short vowel prior to the target consonant, the tagger should choose to insert the “consOnset” tag to denote the onset of the epenthetic vowel. Insert the “turbOnset” tag at the onset of turbulence.

    b. <u>Epenthetic stop between target consonant and vowel</u>

In the event of an epenthetic stop between target consonant and vowel, follow these steps:

i. Place turbOnset at the onset of turbulence as usual
ii. Choose to insert turbOffset. Place turbOffset at the end of turbulence, but prior to the burst for the epenthetic stop. If no burst is visible, place turbOffset tag at the end of turbulence.
iii. Place VOT at the first peak of the upswing as normal

c. <u>Sibilant affricate with burst</u>

If the child produces a sibilant affricate, the ConsType tier tag should be sibilant affricate.

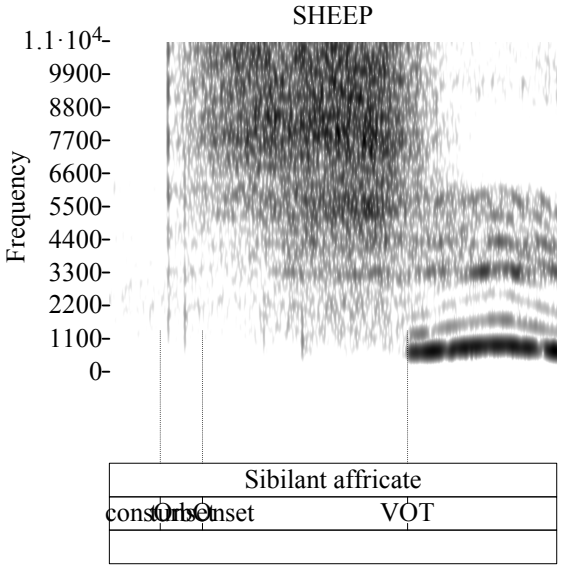

Figure H. Example of a sibilant affricate with a visible burst

i. If there is a visible burst, insert consOnset prior to the burst portion.
ii. Place the turbOnset tag after the burst
iii. Place VOT at the first peak of the upswing as normal
iv. If there is no visible burst, label the target as an affricate and insert turbOnset and turbOffset tags.

d. Silent hiatus or aspiration between fricative and vowel

A silent hiatus or period of aspiration is treated analogously to epenthetic stops.

i. Place turbOnset at the onset of turbulence as usual
ii. Choose to insert turbOffset. Place turbOffset at the end of turbulence, but prior to the hiatus/aspiration.
iii. Place VOT at the first peak of the upswing as normal

e. Fricated vowels

A fricated vowel is an acoustic event where the frication continues after the onset of voicing. The tagger should zoom into the spectrogram so that each peak is visible. Insert VOT at the peak of the first upswing as normal. A note may be made of FricatedVowel and inserted into the Notes tier.

f. Loud talkers and clipped productions

Occasionally, the participant may repeat the word with excessive amplitude. Since the increased amplitude is unlikely to cause clipping in the fricative, these cases can be tagged as normal. A note can be made and inserted on the relevant Wiki table.

g. Quiet talkers and whispered productions

Some participants may have a tendency to whisper when repeating the target production. As a result, the experimenter may have asked the child to produce several repetitions of the same target. If this was the case, the tagger should tag each production as a sibilant fricative. A note "Quiet" may be inserted in the Notes tier.

Extreme care may need to be used when deciding where to place turbOnset and VOT in these cases. It may help to zoom in to the spectrogram.

h. Emerging Fricatives

If a participant begins the target production with one fricative (e.g. /h/) and finishes it with a sibilant fricative, this should be treated in a similar fashion to a sibilant affricate. Insert the "consOnset" tag at the onset of the non-sibilant fricative and "turbOnset" at the onset of the sibilant fricative. Insert a "turbOffset" tag, if needed, and "VOT" tag as normal.

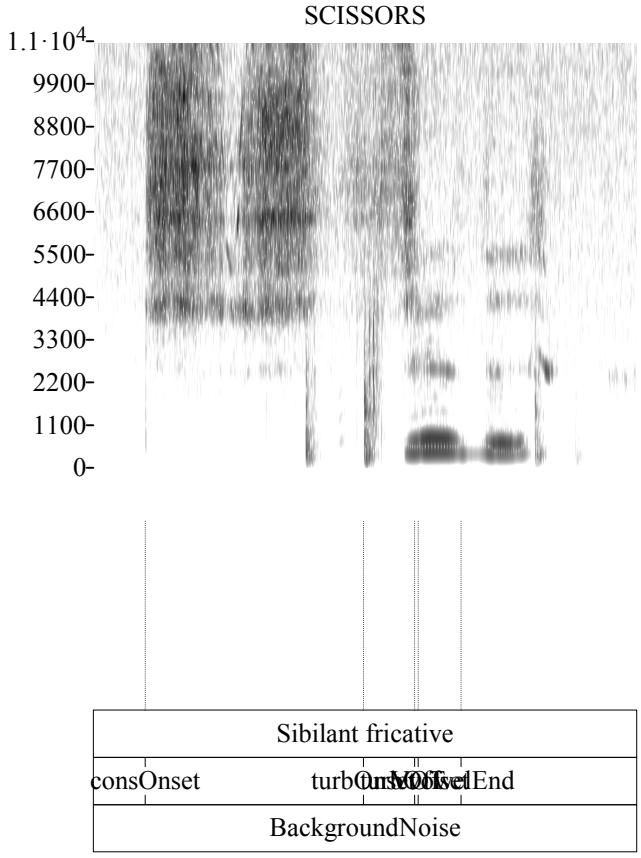

Figure I. Example of an Emerging Fricative where consOnset is used

If the production begins with a sibilant fricative and ends in a non-sibilant fricative (e.g. /h/), treat this case similarly to a hiatus. Insert the "turbOnset" tag at the onset of sibilance and the "turbOffset" tag at the onset of the non-sibilant fricative.

i. Malaprop production

If the participant produces another word (*"Toast")* other than the intended target (*SHORTS)*, choose the ConsType tag that matches the consonant produced by the child (e.g. "Non-sibilant plosive" in the case of *toast*). Be sure to insert the Malaprop note from the drop-down menu as shown in Figure J. Type the child's production (TOAST) into the Malaprop field.

If the child produces a malaprop that begins with a sibilant fricative, this should be labelled as a sibilant fricative in the ConsType tier. It is very important in this case that a Malaprop note be inserted from the drop-down menu and that the Malaprop be written into the Malaprop field. This will

alert the analyst that this production requires special attention. You may tag the production with turbOnset, turbOffset, VOT and vowelEnd tags as appropriate.

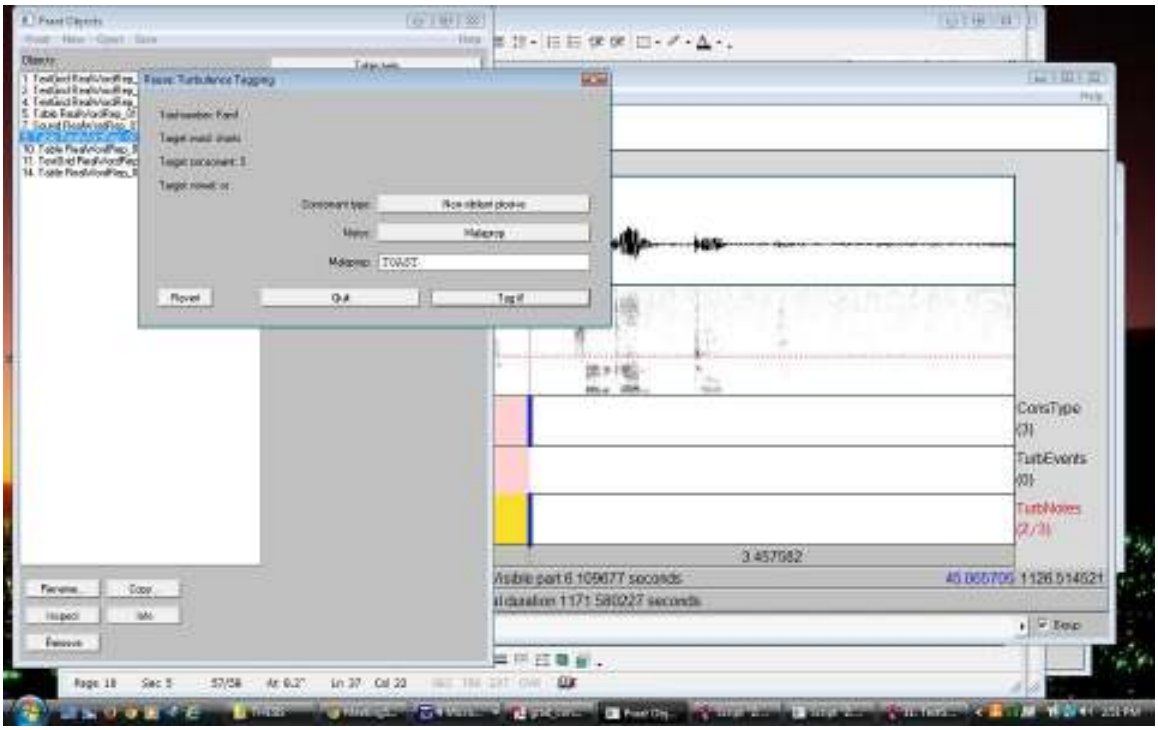

Figure J. Image depicting the steps to take in the case of a Malaprop

2. *Environmental Challenges*

    a. Static or buzz in waveform

If the soundfile was recorded with a buzz, this may be evident on the spectrogram. Try to place the boundaries as normal but if this is impossible, make a note so the file can be listened to later by another tagger. Insert a note indicating “Static” in the notes tier.

    b. BackgroundNoise

If a background noise (e.g. tapping of computer keys, foot kicking table) overlaps with the target consonant and vowel, select the “BackgroundNoise” note from the drop down menu.

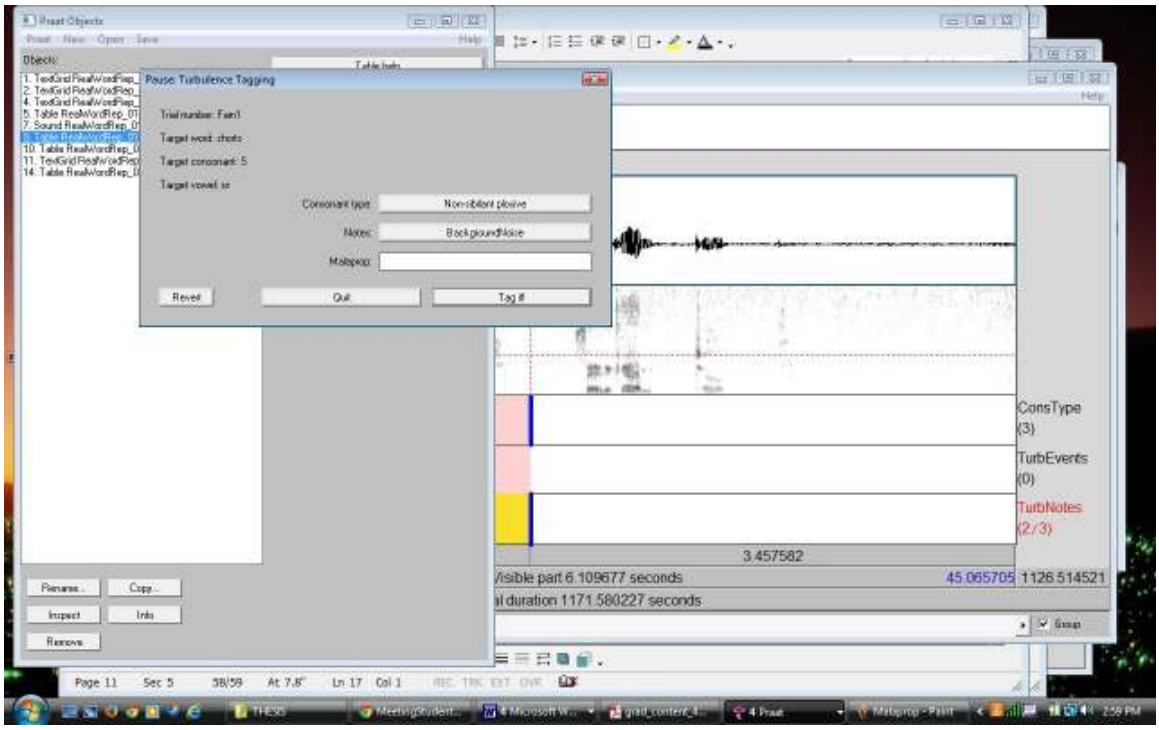

Figure K. Image depicting BackgroundNoise selection from the drop down menu

c. Overlapping Voice

Occasionally, the computer stimulus or speech from an experimenter or a parent in the room may overlap with a target production from the participant. If the speech overlaps the target consonant and vowel, select and insert the "OverlappingVoice" note from the drop down menu. Attempt to tag as normal being careful to distinguish the child's production from any speech from another adult.

In the example below, the first period of turbulence is the final /s/ produced in the computer stimulus. Note how the "turbOnset" tag was inserted after the stimulus /s/.

If the stimulus had overlapped the target consonant any further, the data would not have been valid. It would have been impossible to distinguish the adult fricative from the child's fricative. The tagger should insert the "OverlappingVoice" note and proceed.

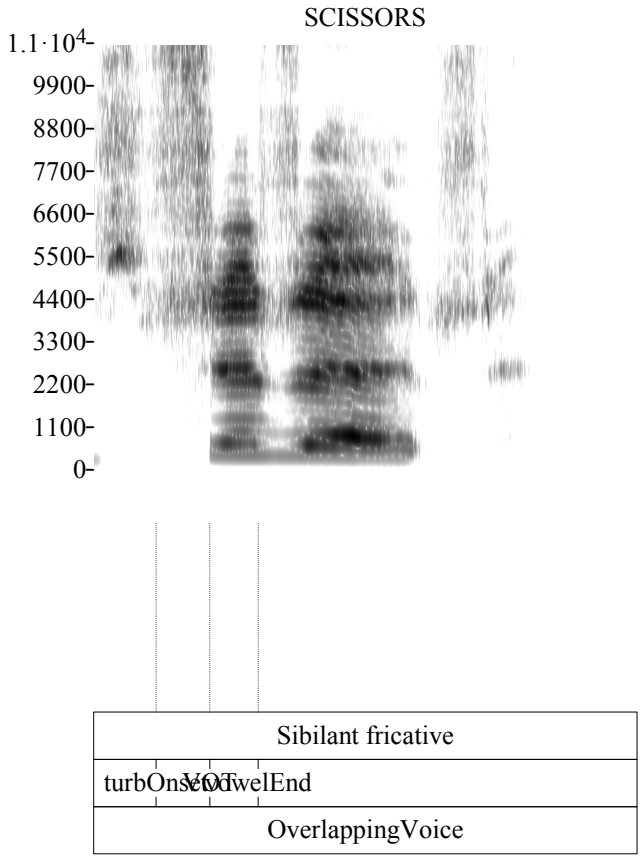

Figure L. Example of overlapping speech