Modifying Speech to Children: An Acoustic Study of Adults' Fricatives

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

Hannah M. Julien

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

Associate Professor Benjamin Munson, Adviser

May 2010

Acknowledgements

Thank you to Laura Fristad (my first grad school friend) and Heidi Stone (my fellow cyclist). I am honored to have become a speech-language pathologist alongside you both. Thank you to the Phonology Lab members, especially Kari Urberg-Carlson, Eden Kaiser and Marie K. Meyer for running subjects and for all the 8 million questions you were willing to answer. Thank you to Gerald Burke for your patience, fricative marking expertise and general willingness to take on a part in this project.

Thank you to Jan Edwards and Mary Beckman for allowing me to be a part of the larger HSD project. Thank you also to Mary for your time and willingness to serve on my committee. Thank you to Peter Watson for your invaluable feedback and support. Also, thank you for being such an enthusiastic lecturer; you are an inspirational professor who brings chocolate to class.

Thank you to Benjamin Munson, for being my advisor and mentor, and teaching me about how it all works (and works out in the end). If I am a great professor someday, it will be because I was first your student. Finally, thank you to my big brother, Kristian, who still gives me hugs when I am down, and to John and Susan Julien who taught me to do my best every day.

Abstract

This study examines the relationship between adult perception of children's speech and their subsequent productions to children. Twenty-two adults participated in the study. In the listen-rate-say task, first the participants heard a child's production of a speech sound and saw a picture representing the word the child was attempting. Second, they judged the child's accuracy using a visual analog scale (VAS). Finally, they were instructed to respond to the child by naming the image. The participants also participated in a "baseline" task in which they produced sentences that contained the target words, first in a clear speech style and then a conversational speech style. That is, for the second repetition of sentences, they were explicitly told to use a speaking style that would help a listener perceive the signal. Analysis focused on the mouse-click locations, fricative centroids and durations. Results revealed the average click location was correlated with the centroids of the fricatives being rated. The results were also not significantly different from those of Urberg-Carlson et al. (2008) even though in the present study the adults had knowledge of the target word they were rating. Within subjects ANOVA was used to examine whether centroid and duration differed between clear and conversational speaking style. Results revealed a revealed a significant main effect of fricative and style for duration. There was also a significant main effect of fricative for centroid. Hierarchical multiple regression was used to examine the relationship between the judgments of the children's speech and the immediate model the participant provided after making the judgment. For /s/, a small, but statistically significant proportion of the variance in M1 was accounted for by the mouse click location. Mouse-click location

accounted for a statistically significant proportion of the variance in duration for both /s/ and /ʃ/.   This suggests that for a subset of our participants, the duration of the fricative produced in response to a child was in part mediated by the adult's perception of that child's production. The data also suggest that the centroids of the responses were not mediated by the adults' perception in the same way as duration.

Table of Contents

List of Tables

List of Figures

Introduction

Undeniably, the process by which children learn language is both fascinating and complex. Language learners must learn both to perceive and to produce the speech sounds in a target language, the language of their speaking community. Additionally, they must be able to do this within the context of immense variation in the phonetic forms that they are exposed to during acquisition. This study attempts to further our understanding of this process by investigating the social dynamics of phonological acquisition. Specifically, it examines the relationship between adult perception of children's speech and their subsequent productions to children, which arguably serve as models to children during acquisition.

It is important to frame this relationship in the context of what we know about (a) children's speech development, (b) speech perception (including studies of measurement of this ability) and, (c) perception-production dynamics during interactions. Therefore, this paper will be framed with a brief review of literature relevant to this topic. First, studies of child-directed speech will be reviewed. Then, children's acquisition of speech sounds (i.e., children's productions) will be addressed by reviewing various studies (including those which highlight the gradual process by which children acquire speech sounds), and those which look at cross-linguistic differences in phonological acquisition. Finally, studies which have investigated the interplay between input and children's speech perception abilities will be reviewed. This review will motivate the two broad research questions. First, do adults produce speech differently in response to productions they judge to be inaccurate compared to ones they judge to be accurate? Second, how continuous is the nature of perceived accuracy and modification of their response?

Input

The speech and language input which children receive is undoubtedly a topic of critical importance for researchers in child development, speech and language development, and sociolinguistics. Much of the research is focused on the models preverbal infants receive. Infant-directed speech (IDS) has been studied within various theoretical frameworks, using different experimental tools. Investigators have shown that talkers use different intonation patterns—including higher overall pitch—and a more exaggerated (hyperarticulated) vowel space, and that infants show a preference for this type of speaking style (for a review, see Soderstrom, 2007). Soderstrom (2007) illustrates that the input infants (and children) receive is perhaps a function of the developmental stage of the child and the unique set of talkers within the environment. The present study attempts to give us a better understanding, at the acoustic level, of how listeners respond to children based on the accuracy with which they perceive their speech. This would provide a more detailed picture of how adult speech input may be a function of developmental stage of the child.

Kuhl and Andruski (1997) studied the linguistic input children receive. Their work focused on the vowel spaces of three different languages, English, Russian and Swedish. Acoustic measures revealed that for all three languages, mothers' speech to children was produced with more distinct vowels. The authors noted that the vowel formants were not just higher which would suggest a more imitative relationship, but that the vowel space became larger relative to adult-directed speech. The authors suggest that this expanded vowel space functions to support children's ability to learn language by making contrasts (for example vowels) more distinct and therefore discernible, and also by providing a greater number of clear instances or representations of the target. The results also showed an increase in fundamental frequency and vowel duration for all three languages.

Output

Children's phonological development (both perception and production) is also a well researched area. Research has shown that children acquire speech sounds gradually (for a review, see Hewlett & Waters, 2004). That is, their productions are not consistently wrong (or not present) and suddenly correct (or present). Numerous studies have demonstrated how children's approximations of speech sounds gradually become adult like, accurate, forms.

Work by Kewley-Port and Preston (1974) demonstrates the gradual nature of speech sound production. They recorded 3 children at regular intervals over a 2 year period. Recordings (elicited in natural contexts) and acoustic analysis focused apical English stop consonants (/t/ and /d/). Voice-onset time (VOT) is one distinguishing feature of stop consonants, and is defined as the time between the stop's release and the onset of vocal fold vibration. Long VOTs have been shown to be associated with voiceless stops in English, and short VOTs with voiced stops (Lisker & Abrahamson, 1964). The distribution of VOTs (from the infants) suggests articulatory coordination that differs from adult-like productions but the capacity to produce consonants with both long and short VOTs. Over time, the distribution of children's VOTs showed twice as many /d/ range consonants as /t/ range, whereas the adults' distribution was bimodal. As the children began to increase their expressive words, they the /d/ range consonants increased within the distribution, and while there were still errors (one part of the experiment included adults judging which sound the heard), the children's distributions began to resemble that of the adults. This seems to suggest the gradual process by which children acquire speech sounds.

Barton and Macken (1980) compared the voice onset time (VOT) of word-initial stops produced by children with that of adults. This research demonstrated the continuous nature of children's speech sound acquisition. The researchers compared the VOT values (from single words and words embedded within sentences) of two year old children with four year old children, and the children's values with the adults'. Results revealed that the four year old children's VOT values were, on average, longer than adults. Also, there was also a greater range of VOT values for the children than for the adults. Moreover, the researchers showed that the children's speech (measured by VOT values) seemed to follow a trajectory of first overshooting the adult value, and then more closely approximating it (compared to slowly approaching the appropriate/acceptable production). Barton and Macken (1980) showed that not only is the acquisition of speech continuous, eventually the children refined their productions to meet the community norms.   Children's productions exist within a speech community and are therefore perceived and interpreted by other talkers—siblings, peers, parents and other adults—within that community.

Work by Scobbie, Gibbon, Hardcastle and Fletcher (2000) includes discussion of covert contrast, providing further evidence of the gradual acquisition of speech sounds. A covert contrast occurs when a child produces a contrast between two phonemes; however, the contrast is perceived to be the same (interpreted by an adult as errored in some way) when in fact there is acoustic differentiation between the intended targets (as was first shown by Barton & Macken, 1980, see above). This would suggest both that the child has the ability to perceive the contrast they intend to make between two sounds and that the articulatory movements necessary to produce the difference do not (yet) produce a perceptible difference.

Children's gradual acquisition of speech sounds was also shown by Baum and McNutt (1990). They studied children's productions of /s/. The two experimental groups (children, ages 5-6 and 7-8 years old, who had been labeled as having a misarticulated /s/) were compared to age-matched controls that were judged to have typical articulation. /s/ and /Ө/-initial words were elicited in a prompted reading task. Results revealed greater variability in the spectral characteristics of /s/ in the experimental groups. Moreover, acoustic analyses revealed significant differences between the children's misarticulated /s/ production and their production of /Ө/. The results also showed acoustic similarities between the misarticulated tokens produced by the older group with the younger control group.

The gradual nature of children's phonological acquisition is also demonstrated by a greater amount of coarticulation in their speech. Research by Nittrouer, Studdert-Kennedy, and McGowan (1989) demonstrated that children's productions have more coarticulation than adults, and that this gradually decreases as children get older. They studied /s/ plus vowel and /ʃ/ plus vowel sequences. Results revealed that both fricatives had lower F2 frequencies when they were followed by /u/ than followed by /i/. The authors suggest that anticipatory lip rounding (which lengthens the vocal tract) for /u/ decreases the F2 frequency of the preceding fricative. This research also supports the idea that children are able to produce syllable units before they are able to produce larger segmental speech units. The differences in results between children (both a younger and older age group) and adults could not be attributed to differences in shape of the vocal tract.

Not only is children's speech acquisition gradual and marked with more coarticulation, the production of phonemes is also more variable than that of adults. Munson (2004) studied the spectral variability (measured as a trial-to-trial differences in spectral mean) of /s/ between children and adults. This work was based on previous work by the author (e.g. Munson, 2001). The investigator measured /s/ in /s/ + vowel (/a/ and /u/) and /s/ + consonant + vowel sequences (/w/ and /p/) elicited in a repetition task. Results revealed that children's production of /s/ has more spectral variation, and more temporal variability than adults. However, differences between the different age groups of children were not significant (accounted for perhaps by the small subject pool). The author suggests that more variability does not necessarily mean lack of skill in communication (that is, there is not a reason to view variability as a negative), and that coarticulation may be a way that children attempt to increase their intelligibility.

The previous section has shown that phonetic development is gradual. Children gradually learn to produce adult-like tokens of phonemes, and researchers have been able to measure differences between sounds that have been perceived to be identical. This provides a much different picture of speech sound development compared to if researchers or clinicians were to solely document children's speech using IPA-style transcription.

Cross-Linguistic Studies

Cross-linguistic studies have furthered our understanding of speech sound development in children including children's gradual acquisition of speech sounds. Stoel-Gammon, Williams and Buder (1994) studied native English-speaking and Swedish-speaking adult and children's productions of /t/. Perceivable and measurable acoustics differences (VOT, burst intensity and burst spectral diffuseness) were found in both the adult and children's productions in both languages. Moreover, trained listeners were able to determine the place of articulation for the stop (either dental or alveolar), and then put the talker into a 'language group.' While the children's productions were (acoustically) language specific. They were more challenging for the listeners to classify; more children than adults were misclassified according to their productions. This seems to suggest a universal, gradual process of the adult-form of a sound.

Li, Edwards and Beckman (2009) showed differences in the accuracy with which English-acquiring and Japanese-acquiring toddlers produced voiceless fricatives. The children were 2 year old and 3 year old monolinguals. The researchers elicited the target fricatives, /s/ and /ʃ/ for English, /s/ and /c/ for Japanese, all word-initial, through a repetition task. The productions were transcribed, and also acoustically analyzed. Results showed the English infants were more accurate (as judged through transcription) in their productions of the more anterior voiceless fricative than the more posterior one. The Japanese infants were more accurate in their production of the more posterior fricative than the anterior one. The results of this study also showed acoustic differences in sounds that were transcribed with the same phoneme, and that covert contrasts were observed in participants, both English speaking and Japanese speaking. Transcription allows for categorical judgments to be made, but only to the extent to which the listener perceives the sounds. Some of the participants were producing sounds with measurable acoustic differences but these differences were not accounted for through transcription. Moreover, the same phonemes exist differently across languages. That is, for example, the /s/ in Japanese is acoustically distinct from the /s/ in English.

Interplay between Input and Output

There is a paucity of research regarding the interplay of children's vocal and verbal output with adults' vocal and verbal output (models). Talkers tend to manipulate their speaking style depending on their audience, or the speech community within which they are communicating. This intersection of socio-linguistics, conversation analysis, language acquisition and laboratory phonology is not only interesting, but relevant to this paper only peripherally as children's speech acquisition fits into a broader understanding of language acquisition. It is important to note briefly the idea that children are not only responsible for learning the broader set of speech sounds within their native language *but also* they must learn which variation is important for lexical contrast, and which is important for indexical meaning (see work by Docherty, Foulkes, Tilloston & Watt, 2006).

Studies from adult laboratory phonology provide evidence phonetic detail in productions can change depending on social dynamics. Babel (2009) studied vowel accommodation between two talkers (a black man and a white man), and the extent to which automated social behavior affects this form of phonetic imitation. To collect baseline data, participants completed a single-word production task in which they read single words presented on a screen. Participants then completed the shadowing task in which they heard one of the two talkers, and were instructed to repeat the word. Participants then either completed the same task or the task but with the picture of the talker who had produced the tokens they heard and were meant to repeat. Depending on which iteration of the experiment they completed, they were asked to either identify the race of the talker, or they were asked to rate the attractiveness of the talker on a scale of 1 to 10. The talkers were overwhelmingly identified as being white and the ratings of their attractiveness were quite variable. Analysis focused on the location of vowel formants, and more specifically on how much a participant's production changed over the course of the exposure to the talkers' productions.

The participants also completed an Implicit Association Task (IAT; Greenwald et al. 1998) in which they first are asked to judge whether a first name is 'black or white,' and then whether a word is ‘good or bad’ (i.e., rainbow, cancer; Babel, 2009). Then, the participants were asked to categorize words and names. The categories in which they could put the words were a combination of the first two tasks. So, for example, they saw ‘black’ above ‘good’ and ‘white’ above ‘bad’ and then were asked to put the name or word into a category. The combinations of black/white and good/bad were reversed and counterbalanced for the order in which they were presented. Results revealed that participants did not accommodate to one talker more than the other, and that male and female participants seemed to accommodate to the same degree, despite the fact that both of the talkers (to which they were potentially imitating) were male. The results also showed that accommodation was present for some vowels (/ae/ and /a/), but not for others (/u/, /o/ and /i/), and driven, to a certain extent, by social values. For example, individuals who showed a pro-black bias accommodated more for the black talker. Female participants accommodated more when they rated the talker as more attractive. Males accommodated more in the social conditions, as opposed to the asocial conditions. Babel suggests that accommodation is neither solely an automatic process nor a socially driven phenomenon.

Pardo (2006) studied the extent to which talkers demonstrated phonetic convergence over the course of a natural conversation. Phonetic convergence is defined as ‘an increase in segmental and suprasegmental similarity of the speech of one talker to another’ (2384). Talkers participated in a task in which they gave their conversational partner directions based on a map. The pairs were female-female and male-male; there were no male-female dyads. In an ABX design, listeners judged the talkers’ speech

(speech prior to participating in the map task, after the map task, and for a repetition task) in comparison to their conversational partner's production (recorded during task). Results revealed that the talkers' speech did indeed become more perceptually similar; phonetic convergence was influenced by the role of the talker during the task and their sex, and that this lasted after the interaction was complete. In the female dyads, the talker who gave the instructions demonstrated more convergence (as judged in the perception task) to the receiver. Oppositely, in the male dyads, the receiver of instructions converged more to the giver. Overall, males demonstrated more phonetic convergence than females. The results do not completely align with previous studies regarding how talker sex influences accommodation (i.e. Namy *et al*, 2002); the author uses this as support for the idea that phonetic convergence does not occur solely as a function of perception and production. Obviously, there are multiple variables implicit in any communicative act which may influence the magnitude and direction of phonetic convergence. Pardo asserts variables such as talker sex and conversational role may be idiosyncratic to each communicative pair. We interpret Pardo's findings as evidence that phonetic accommodation in dyads is much more task and role specific than would be suggested by more generic model such as the one described by Bell (1994).

It is important to note that there are most likely multiple variables which may influence the extent to which phonetic convergence emerges between talkers. It is within the scope of this investigation to discuss phonetic convergence in that it is a product of a dyadic interaction. This investigation attempts to further our understanding of the interplay between adult model and child talker.

The next section will focus on the dynamics of adult-child dyads, and how variation in children's perception abilities can be linked to variation in adults' productions. Liu, Kuhl and Tsao (2003) combined a focus on infant directed speech with measurement of children's speech discrimination abilities. They recorded native Mandarin-speaking mothers as they spoke to their infant and as they spoke to another adult native-speaker. The infants' (who were grouped into either the 6-8 months old, or 10-12 months old group) speech perception abilities were measured in a discrimination task using a head-turn paradigm. Analysis focused on the vowel space (measurements of formant values were extracted from the speech samples) of the adult, and the speech sound discrimination of the infants. They found that the mothers not only utilized expanded vowel space when speaking to their infant but also that Mandarin-acquiring infants whose mothers utilized a more expanded vowel space had better speech perception of a phonemic fricative contrast.

Recent research by Cristià (2009) further explores how adults' speech is correlated with their children's speech perception abilities. More specifically, she wanted to know if higher emotion/affection characteristics of adults' IDS yields better speech perception in children, or alternatively, if clearer 'phonetic boundaries' in adults' speech yields better speech perception in their children. She studied the relationship between the extent to which caregivers contrasted /s/ and /ʃ/ and the perception abilities of their children. Two different age groups of infants were studied: 4-6 month olds and 12-14 month olds, all were native English-acquiring monolinguals. The adult participants were the primary caregivers to the infants. The infants' perception was studied using the

Visual Habituation procedure (Maye, Weiss, & Aslin, 2008; Narayan, 2006; Werker, Shi, Desjardins, Polka, & Patterson, 1998). The adults' speech was recorded during natural interactions with their child and with the experimenter. The target phonemes were elicited during these interactions by having the dyad play with toys with the target phonemes (/s/ and /ʃ/), as well as control phonemes (/p/ and /b/) as the first sound. Results were mixed. It seemed that on average the caregivers to the older children more clearly differentiated /s/ from /ʃ/ (as measured by the difference between mean peak locations). Only some of the caregivers showed an increase in vowel space size. Moreover, the adults' speech to their child could not be consistently classified into either demonstrating more emotion/affection characteristics compared to more clear phonetic detail. Interestingly, the results showed a relationship between the infants' perception abilities, and the difference between their caregiver's /s/ and /ʃ/ but not the pitch excursion (the measure of emotion/affection).

Cristià (2009) notes, generally, infants spend a large amount of time with the same small number of talkers. This is an important consideration in terms of the variation (or potential invariance) that children may receive as acoustic input. As children develop, they have more opportunities to interact with different communicative partners. Speech sound development is a gradual, emergent process. The speech models, embedded in the communicative interactions with these varied partners, seem to be emergent and varied as well.

Perception of Children's Speech

The next section discusses research related to speech perception and perceptual learning. This study is, in part, focused on adult's perceptions of children's speech, and as such it is important to discuss the amount of phonetic detail that listeners are able to perceive and to understand how adults' perceptions correspond to acoustic detail within a speech signal. Moreover, it is important to discuss factors that may influence adults' perception of children's speech. As highlighted in the previous section(s), children's speech sound acquisition is a gradual and emergent process. Therefore, the framework for the present study must consider how adults may judge children's productions as they approximate more adult like forms.

Nygaard, Sommers and Pisoni (1994) demonstrated that familiarity with a person's voice (facilitated through perceptual learning) increases an individual's ability to recognize novel words. Listeners participated in task in which they learned the voices of 10 different talkers over a 9-day training period. They were then tested on a set of new words (which had not been used during the training tasks). Results revealed that the listeners were better able to identify words (with noise was added to the task) if they were produced by a familiar voice compared to an unfamiliar voice. The authors suggest this result demonstrates that the encoding of information about a talker facilitates later decoding of phonetic information. When one considers the small group of communicative partners with whom a child initially interacts, it seems that the coupling of these two processes better facilitates speech sound learning in children.

Research suggests that adults have the ability to perceive fine details within phonetic categories. Sharf, Ohde and Lehman (1988) investigated the degree to which listeners were able to identify a distorted /r/ token from a synthesized continuum. The researchers used two different training and feedback conditions to determine if either increased listeners' ability to identify the distorted /r/ token. Results showed that training did not increase listeners' performance and that only some listeners were able to identify the distorted /r/ token. Identification of intermediate, allophonic tokens of the /r/ sound seems to suggest that categorical perception of phonemes could be more related to an experimental paradigm, not the perception abilities of listeners. Results also revealed a relationship between identification of distorted /r/ sound and discrimination of /t-d/ continuum but not /w-r/ continuum. This study demonstrated that individuals' perception of the vowel-like consonants /w-r/ is not categorical.

Kraljic and Samuel (2005) studied how auditory experience may alter listeners' perceptions. After completing a lexical decision task (which served to expose the listener to the talker's production of /s/ and /ʃ/, as well as the ambiguous s-sh sound created from each individual talker) participants completed a categorization task. Some participants completed an 'unlearning' phase of the experiment in which they completed a silent, visual task in between the listening (lexical decision task) and the perception (identification) task. Results showed that listeners' perceptions were influenced depending on the ambiguous sound to which they had been exposed in the lexical decision task. That is, if a listener was exposed to ambiguous productions of / ʃ / during the lexical decision task, they labeled more tokens as /ʃ/ during the categorization task. Moreover, there was an even larger effect for the group who completed the unlearning task which seems to indicate that perceptual learning is not necessarily (and solely) an immediate perceptual phenomenon. The results also demonstrated that listeners seemed to learn in a speaker-specific way. Also, their results seem to suggest that perceptual learning is, in part, influenced by the acoustics (and variations of those acoustics) of the signal.

More recent work (i.e., Kraljic & Samuel 2007) further expands this idea and suggests that perceptual learning varies for different phoneme contrasts. That is, perceptual learning of some phoneme contrasts (for example, /t/ - /d/ in their study) seemed to result in more coarse shifts in perception. Contrastingly, perceptual learning of fricatives was finer in that the listeners seemed to maintain speaker-specific representations of phonemes. Thus, perception seems to be, at least in part, driven by auditory experience that does not function as pure priming or adaptation.

Schellinger, Edwards, Munson and Beckman (2008) showed that naïve listeners are able to make less-categorical judgments about children's speech (and the errors within it) that correlate with the type of error/substitution which the child made. This not only provides support for covert contrasts within children's speech but also furthers our understanding of adult perception.

Moreover, researchers have demonstrated that judgments of more continual contrasts have been shown to correlate with certain acoustic properties in the speech signal. Urberg-Carlson, Kaiser and Munson (2008) investigated three different, non-categorical methods which allowed listeners to signal their perception of children's productions of /s/ and /ʃ/. The first was a measure of Reaction Time to a forced choice decision (F-CRT) about the phoneme the participant heard. The second was a Direct

Magnitude Estimation (DME) in which listeners were asked to provide a number based on their perception of the phoneme which was presented to them. The researchers noted that the instructions for DME are complex and it is difficult to get reliable results from listeners. The third method was Visual Analog Scaling (VAS) in which listeners were presented with a child's production of an /s/-/ʃ/ phoneme and asked to judge where it fell on a continuum. The continuum was a horizontal line with the words 'the 's' sound' and 'the 'sh' sound' at opposing ends. The researchers compared listeners' responses for the three tasks to acoustic measures of the fricatives. Results revealed a strong correlation between ratings from the VAS task and the fricative acoustics. Urberg-Carlson et al. argue that VAS provides the best method for judging 'category goodness.' Subsequent studies have demonstrated that judgments made by VAS were not significantly altered with varying amounts of delay added between the presentation of the sound and the prompt to judge the phoneme (Munson, Kaiser, Urberg-Carlson, 2008).

Research has shown that adults' perception of children's productions is influenced by the native language of the adult. Li, Munson, Edwards, Yoneyama, and Hall (in press) studied native English-speaking and native-Japanese speaking adults' perceptions of /s/ and /ʃ/. The stimulus items were produced by native speakers and included productions from children and adults. The stimuli were accurate productions and substituted productions (/s/:/ʃ/ and /ʃ/:/s/) (as determined by transcription) and were balanced to reflect the typical patterns of acquisition in each of the languages. That is, the English speaking adults were presented with more /s/ for /ʃ/ substitutions and the Japanese-speaking adults were presented with more /ʃ/ for /s/ substitutions. Their results showed that the participants relied on different aspects of the acoustic signal, and that this was a function of their native language. Both groups perceptions were correlated with the centroid frequency of the fricative (M1), the English-speaking adults also relied on F2 onset while the Japanese-speaking adults relied on F2 onset and M2. This finding in perception complemented data from cross-linguistic production studies on fricatives (i.e., Li et al. 2009). Also, the English-speaking adults were more likely to accept an intermediate token as /s/, while the Japanese-speaking adults were more likely to accept intermediate tokens as /ʃ/. Their work highlights that speech sound acquisition is not only in the signal produced by the children, but also in the perception (and linguistic experience) of the listeners, and that if transcription is the only method used to capture the development of speech sounds in children we risk misinterpretation of these trajectories.

Native language plays a role in listeners' perception. Moreover, work by Munson, Edwards, Schellinger, Beckman and Meyer (2010) suggests that listeners' perception of children's speech may be a function of both expectations about developmental variation in it, as well as sociolinguistic variation of the adults' speech within their speaking community. Their work was a follow up to the small biasing effect of carrier phrase reported in the work by Schellinger et al. (2008). The researchers used three different biasing conditions to see how adults' perception of /s/ produced by children may be influenced. The first condition blocked the tokens according to carrier phrase type (either a carrier phrase meant to bias the listen to think the child was younger, 'weawwy yike' or older 'really like'). The second condition used different instructions, omitting the mention of ‘misarticulation’ with respect to the productions which the listeners were judging. The third condition acoustically modified the carrier phrases to see if a carrier phrase which more closely matched the target word would lead to greater perceptual biasing. Listeners were asked to respond ‘yes’ or ‘no’ to whether or not what they heard was ‘the /s/ sound.’ They were told that they may hear ‘s’ incorrectly produced and it may sound like ‘th.’ Their results show that those participants in the second condition were more likely to accept intermediate tokens as variants of /s/ when they were not biased to believe they were listening to misarticulations. Munson et al. suggest that this is perhaps a function of the immense sociolinguistic variation in /s/. The results also showed that carrier phrase had a larger effect on the intermediate tokens (/theta/-like stimuli) than on /s/. When appended to the ‘really like’ carrier phrase, they were less likely to be treated as errors, and the authors suggest that this may because listeners were willing to accept them as normal variation as opposed to misarticulations from a child (those that were appended to the ‘weawwy yike’ carrier). This was observed only in the third condition, when the carrier phrase’s f0 had been modified to more closely match the stimuli. This is a contrast to other studies by these authors that have shown that adults are more likely to accept intermediate tokens when they are biased to believe they were from a child compared to an adult.

Before researchers can begin to answer questions that relate to the *degree* to which a learner’s linguistic environment has on her speech acquisition, we must first better understand the nature of the link between child output and adult input. Perhaps adults who perceive children to be nearly adult-like in their productions of words respond differently than to productions that are perceived to be further away from the intended target. This would seem to suggest hyperarticulated models in response to productions which are most divergent from the target. Alternatively, as adults are indeed able to perceive fine grained acoustic detail within a speech signal, perhaps their models serve to reinforce a child as they more closely approximate the adult like form. That is, given that we know children to acquire speech more continually than categorically, perhaps adults respond with reinforcement which supports the child's productions by closely approximating the child's production (irrespective of perceived accuracy). The results of this study have implications that will potentially further our understanding of typical speech sound development. Moreover, we hope to further strengthen the body of evidence that demonstrates adults have the ability to perceive fine acoustic detail when provided with an appropriate task that allows them to make fine grained judgments.

In addition to furthering our knowledge of general child development, a more complete understanding of typical speech development has important clinical implications. One of the roles of a speech-language pathologist (SLP) is to help children who misarticulate certain speech sounds. If one subscribes to a more naturalistic approach to therapeutic interventions, knowledge of the relationship between perception, modeling and children's speech sound acquisition would be foundational to their creation of an intervention for a child with errored/atypical speech productions.

As previously mentioned, this study addresses two broad questions. First, do adults produce speech differently in response to productions they judge to be inaccurate compared to ones they judge to be accurate? Second, how continuous is the nature of perceived accuracy and modification of their response? Therefore, results which demonstrate that adults systematically alter their responsive productions (as measured by fricative duration and/or centroid frequency) would provide evidence against the null hypothesis, that adults do not produce acoustically different models in response to productions they perceive to be inaccurate.

Methods

Participants

Twenty four participants (19 female, 5 male) participated in the study. They ranged in age from 19 to 49 years (M=25.83 years, SD=8.42). According to self-report, they had no history of speech, language or hearing problems. Data from two of the subjects were not included. A recording error prevented inclusion of one of the subject's productions. The other productions which were not included contained marked dysfluencies and unintelligible speech.

Stimuli

The 200 stimuli were children's productions of the voiceless alveolar and voiceless palatoalveolar fricatives. These recordings were taken from a study by Li et al. (2009). The productions were elicited during a scripted imitative-naming task in which the children saw a picture (item representing the target word) and were provided a verbal model of the target word. Thus, the researchers were able to control the initial fricative(s) plus (different) vowel combinations via their target words. The words were high frequency and most likely not 'new' words (receptively or expressively) to the children who were prompted to produce them. The fricatives were excised from the remainder of the target word.   Therefore, the stimuli in this experiment were not entire target words, but rather the initial CVs from the target words. See Figure 1 for stimuli acoustics.

Procedures

Perception-production component.

Each trial consisted of three parts which we have called the listen-rate-say task (LRS).   See Figure 2 for a schematic representation of each trial. First, the participant was presented with a stimulus item. Each stimulus item was paired with the same picture which had been used to elicit the stimulus words.   The auditory component was presented over *Sennheiser HD280 Pro* Headphones. The visual component was presented on a computer monitor. The orthographic representation of the word appeared directly below the picture. The item was presented on the screen for 2.0 seconds.

After being presented with the stimulus item and picture/word, the participant was prompted to make a perceptual judgment. This judgment was made using Visual Analog Scaling (VAS). Text that read, 'The 's' sound' appeared on one end of a horizontal line and 'the 'sh' sound' appeared opposite. The participant used the computer mouse to manipulate the cursor's location on the screen and clicked to note the place where they believed the sound to belong. See Figure 3 for a picture of the screen which participants saw when they made their judgment.

Immediately following their judgment, participants were prompted to respond to the stimulus item. The pictured item did not appear on the screen a second time. Instead, the prompt for the participant (text) appeared on the screen, 'now, respond to the child.'

Recordings were made using a *Marantz Professional CDR300* CD Recorder (model no. CDR300/U1B) and a *Shure Dynamic SM48* microphone. Participants completed 5 listen-rate-say practice items, followed by 200 actual recorded trials. See Appendix B for the instructions provided to the participants.

Clear speech-conversational speech elicitation component.

In order to better interpret the acoustic measurements elicited during the production part of the experiment, the final component was a speech elicitation section. This gave us baseline data on individual subjects' ability to alter their production of fricatives in response to a command that they speak more clearly. This allows us to better interpret individual differences in participants' performance on the LRS task. The same words (or word approximations) from which the children's tokens had been segmented became key words in a set of 30 sentences. (See Appendix A for sentences). The participant read the sentences from the computer screen. First, the participants were instructed to 'read the sentence.' Immediately following the final sentence, the participants were asked to read the set of sentences a second time. They were instructed to read them as if they were speaking to someone whom they perceived to be a second-language learner of English, or an individual with a learning disability. The participant read the sentences a second time. Each session lasted between 30-45 minutes. Participants were paid $10 for their participation.

Analysis

The Praat (version 5.1.29) signal processing software (Boersma & Weenick, 2002) was used to acoustically analyze the production data. The fricatives from both the baseline clear-conversational task and the listen-rate-say task were initially analyzed using a script which marked deviations from baseline amplitude. Each fricative boundary was then manually checked and marked by the author and an undergraduate student who had specialized coursework in speech acoustics, and extensive experience using the Praat software. Fricatives were marked by their onset (defined as the start of the aperiodic frication noise visible on the spectrogram) and offset (typically this was the cessation of the aperiodic signal, though in cases where there was formant-like structure in the aperiodic signal, this was taken to reflect overlap of the glottal-opening gestures of the fricative with the vowel, and was included as part of the vowel).

The first spectral moment (M1) of each fricative was one measurement extracted from a 40 ms window centered around the midpoint of the fricative. Spectral moments analysis (Forrest, Weismer, Milenkovic, & Dougall, 1988) treat the power spectrum as a random distribution, and summarizes it using the mean (M1), standard deviation (M2), skewness (M3), and kurtosis (M4) of the fricative. Centroid is another term to describe the mean frequency of the fricative. Spectral moments have been shown to differentiate between English /s/ and /ʃ/ for both adults (e.g., Jongman et al., 2000) and children (e.g., Fox & Nissen, 2005). The fricative /s/ has a higher frequency M1 than /ʃ/. It also has a more negatively skewed spectrum (i.e., a lower M3) than /ʃ/. The duration of frication energy was also extracted from the data. A longer duration is associated with slower, clearer speech. Therefore, for each subject we have the centroids and durations from the 205 listen-rate-say productions, 31 productions from the conversational speech and 31 productions from the clear speech task.

The location of the mouse click was also logged, as it indicates the adults' perception of how accurate the children's productions were. We used these data two ways. First, we examined whether the ratings in the LRS task were similar results to those of previous studies using VAS obtained by Urberg-Carlson et al. (2008). The listeners in Urberg-Carlson et al. were not aware of the words from which the fricative-vowel sequences were extracted. It is possible that the LRS task, in which the targets were known, would elicit different VAS ratings. Second, we wanted to examine the relationship between the adults' perception of the children's speech, as indexed by their mouse-click location, and the acoustics of their subsequent production, as indexed by the M1 and duration of their fricatives.

In the first analysis, within-subjects ANOVA was used to examine whether M1 and duration differed between the clear and conversational speech elicitation conditions. More specifically, we were interested to know if individuals produced acoustically different fricatives when given explicit instructions that would elicit a clear speaking style. Moreover, we used this analysis to inform our understanding of the range among individual talkers. That is, a baseline task was necessary in order to appropriately calibrate the measurements from the LRS task. That is, individual productions in the LRS task were analyzed relative to talkers' productions of the same words from the baseline clear versus conversational speech task.

In the second analysis, hierarchical multiple regression was used to examine the relationship between the various acoustic measures extracted from the LRS task (i.e., M1 and duration), and the location of the mouse click for the LRS task, controlling for the acoustic measures taken from the baseline task. We used this method of analysis because we were interested in determining how much of the variance in the spectral characteristic of the fricatives produced in the listen-rate-say task was accounted for by the spectral characteristic in the clear task, and in the conversational task. In addition to the ANOVAs, the individual subjects' data was examined qualitatively, to assess the extent to which individual talkers' patterns mirrored that of the group.

Results

Results from the various analyses are described below. The first set of results compares the data from the present study with the results from those of Urberg-Carlson et. al (2008). Second, results from the ANOVAs, comparing the clear and conversational speech styles are presented. Finally, the results of the hierarchical multiple regressions in which the mouse-click location data was analyzed with respect to the adults' productions are presented.

The first analysis examined the similarity between the VAS ratings in this study and those made by Urberg-Carlson et al. (2008), reviewed in the introduction. Recall that participants in Urberg-Carlson et al. were not aware of the words that the children were attempting, nor were they required to provide a model production after making their ratings. It is possible that the ratings by listeners in this study were substantially different from those in Urberg-Carlson et al. simply by virtue of these differences. Average VAS click locations for the 200 stimuli from the two studies were subjected to a paired-samples t-test. They did not differ significantly (t[170] = 0.924, p > 0.10). The two measures were strongly significantly correlated (Pearson's r = 0.944, p< 0.001). (See Figure 4 which shows the relationship between average VAS rating from the Urberg-Carlson et al. study and the present study). The average click location in the LRS task was significantly correlated with the centroids of the stimulus items (the children's productions which were being rated). We fit a sigmoid to a scatterplot relating click location in the LRS task with the centroid frequencies of the tokens being rated. A nonlinear regression found that the sigmoidal function (see Figure 7) fit the data significantly, accounting for 56% of the variance. This was statistically significantly different from chance, F[2,197] = 125, p < 0.001. These findings suggest that differences in task demand do not affect ratings in VAS tasks like that used by Urberg-Carlson et al. and in this study.

The second analysis examined differences between the clear and conversational speech tokens. A two-way analysis of variance (ANOVA) was conducted with style (clear and conversational) and fricative (/s/ and / ʃ /) as the within-subject variables. For the first ANOVA, M1 from the LRS task was the dependent measure. Results revealed an a significant main effect of fricative (F[1, 21] = 852.8, p < 0.001, ŋ^2 partial = 0.98) but not of style (F[1, 21] = 0.022, p > 0.05, ŋ^2 partial = 0.001). There was a statistically significant interaction effect (F[1, 21] = 5.84, p < .05, ŋ^2 partial = .022) between fricative and style, although it was small. For the second ANOVA, duration was the dependent measure. Results revealed a significant main effect of both fricative (F[1, 21] = 38.08, p < 0.001, ŋ^{2} partial = 0.65) and style (F[1, 21] = 27.37, p < 0.001, ŋ^{2} partial = 0.57). There was not a statistically significant interaction effect for duration.

The third analysis examined the extent to which the acoustic characteristics of the fricative productions from the LRS task could be predicted by the VAS click location. These analyses were done separately for /s/ and /ʃ/ targets. In this analysis, the influence of lexical and phonetic-context effects on acoustic characteristics of fricatives was controlled by entering the values for the same words' productions in the conversational-speech baseline task in the first step of the regression. In the first set of regressions, the dependent measure was the M1 from the LRS task. The independent measure that was added in the second step of the regression was the VAS click location. For /s/, for the regression predicting M1 values from the LRS task, the M1s from the conversational-speech baseline task accounted for 50.9% of the variance. This was significant (F[1, 1697] = 1757.854, p < 0.001). The VAS click location accounted for an additional small (0.2%) but nonetheless significant (F[1,1696] = 7.339, p = 0.007) portion of variance in the dependent measure. Together, the variance accounted for by the entire regression was significantly different from 0% (F[2,1696] = 885.880, p < 0.001). The standardized β coefficients for both of the independent measures (0.713 and 0.046, respectively) were both significant in the full regression. For / ʃ /, M1 values from the conversational-speech task accounted for 44.0% of the variance of M1s in the LRS task, and this was significant (F[1,2648] = 2084.16, p < .001). The contribution of the VAS location was not significant, and was less than .01% (F[1, 2647] = 0.853, p > .05). The standardized beta coefficient for the first independent measure (M1 from the conversational task) was 0.663, and was significant in the entire regression. The second independent measure was not (β = 0.013, p > .05). Combined, the variance accounted for by the two explanatory variables was significantly different from 0% (F[2,2647] = 1042.45, p < .001).

In the second set of regressions, the dependent measure was the duration from the LRS task. In the first step of the regression the independent variable was the duration from the conversational speech task. In the second step of the regression, the VAS click location was added as an independent variable. For /s/, the duration from the conversational-speech task accounted for 19.4% of the variance in duration in the LRS task, and this contribution was significant (F[1, 1697] = 408.72, p < .001). The second independent variable's contribution of 1.4% of the variance was also significant (F[1, 1696] = 29.37, p < .001). When combined, the variance accounted for by the regression was significantly different from 0% (F[2, 1696] = 222.46, p < .001). The standardized beta coefficients for the independent variables were significant in the overall regression, and were 0.413 and 0.117 respectively. For / ʃ /, duration from the conversational-speech task accounted for 11.6% of the variance in duration in the LRS task (F[1, 2648] = 346.51, p < .001). Mouse-click location accounted for another 2.5% of the variance (F[1, 2647 = 76.90, p < .001). The variance accounted for by the two independent measures was significantly different from 0% (F[2, 2647] = 216.67, p < .001) The standardized beta coefficients (0.338 and -0.158, respectively for the two independent measures) were significant in the entire regression.

Regressions for individual participants are summarized in Table 1. One subject showed a significant relationship between mouse click and M1 for /s/, three subjects showed a significant relationship for / ʃ /. The remainder of the subjects had not significant relationship between the independent and dependent variables for either phoneme. Seven subjects had a significant relationship between mouse click and duration in the LRS task for both /s/ and / ʃ /. Six subjects had an effect of mouse click for / ʃ / but not /s/, one subject had a relationship for /s/ but not /ʃ/. The remainder of the subjects (eight) did not have significant results.

Discussion

This study focused on the relationship between adults' perceptions of children's speech, and their subsequent models. An important component of this study is the novelty of the task in which the participants engaged. We combined a method of measuring adults' ability to judge children's speech with a response task. The results show that participants were able to make fine-grained judgments of children's speech, and that knowledge of the target word did not affect this ability. Their judgments were significantly correlated with the fricative centroids of the children's productions.

We used a clear-conversational elicitation task to calibrate individual participants' speech within the context of the experiment. The clear-conversational speech tasks demonstrated that adults produced clear speech with a higher centroid for /s/, and longer duration for both /s/ and /ʃ/. Duration seems to be a variable that adult speakers manipulate when explicitly instructed to produce more clear speech.

From our analysis of the LRS task, we can note that for a subset of our adult participants, duration was also a variable which was mediated by their perception of the accuracy of a child’s speech. This finding was not for all participants, or as robustly manifest for /s/ and /ʃ/ in the same way. We know that duration is associated with a more clear speaking style. Therefore, considering the findings of the current study, duration seems to be a way in which adults respond to a child whom they perceive to be farther away from the intended target.

Analysis of the centroids demonstrated that while style did account for some of the variance in acoustics, it was not statistically significant. The results of the regressions demonstrated that variance in centroid in the LRS task could not be accounted for by the VAS rating. Perhaps models with a longer duration function to provide the learner with more acoustic information that is more consistent with the ‘targets’ that will be accepted (within their speaking community) as conversational speech. That is, instead of providing a learner with a model which is more contrasted in centroid (higher for /s/, lower for /ʃ/) in relation to what the child is currently producing, perhaps the adult manipulates duration such that the learner gets more information which is more representative of conversational speech. From a longer duration, the learner has a longer ‘window’ to encode relevant detail. Because speech is one way in which socio-indexical information is conveyed, perhaps caregivers/models alter duration instead of frequency. Clearly, further research is needed to parse apart the dynamics of perception and modeling to children as they acquire speech sounds.

One limitation of this study is that it considered a relatively limited range of dependent measures. The centroid measure, in particular, may not capture fully the acoustic detail in /ʃ/. Ongoing work with this data-set should consider other dependent measures, including the psychophysically motivated measures of fricatives described in Munson, Edwards, Schellinger, Beckman, and Meyer (2010).

Regarding the research questions posed by this study: do adults produce speech differently in response to productions they judge to be inaccurate compared to ones they judge to be accurate, and how continuous is the nature of perceived accuracy and modification of their response, the results were not strong, although they do provide a basis for future investigations.

While the scope of this project limited analysis to the M1s and durations, further analysis should be conducted to further parse apart the acoustics of the adults' productions. One of the results of this study is therefore, the opportunity for future investigation/analysis of the great amount of data which was collected. Analyses which focus on vowel formants should be one focus of future investigations using this data.

From an experimental design standpoint, it would be interesting to examine the relationship of perception and production within caregiver-child dyads. That is, perhaps parents' productions from natural interactions could be acoustically analyzed and these measurements could be compared to the same parents' judgments of their children's speech. This could be done using the same VAS-scale and the children's productions taken from the interactions with the parent. Perhaps familiarity within a caregiver-child interaction motivates different speech responses on the part of the parent. The work by

Liu, Kuhl & Tsao (2003) and Cristiá (2009) utilized participants from natural dyads. It could be said that pairs such as these have organic interactions nearly continuously and tacitly negotiate the production and modeling process.

References

Babel, M. (2009). *Phonetic and Social Selectivity in Speech Accommodation*. Doctoral dissertation, University of California, Berkeley.

Barton, D., & Macken, M. (1980). An instrumental analysis of the voicing contrast inword-initial stops in the speech of four-year-old English-speaking children. *Language and Speech,* 23, 159-169.

Baum, S. R., & McNutt, J. C. (1990). An acoustic analysis of frontal misarticulation of /s/in children. *Journal of Phonetics*, 18, 51-63.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (version 4601) [Computer Program]. Retrieved 10-2010, from http://www.praat.org/.

Cristià, A. (2009). *Individual variation in infant speech processing: Implications for language acquisition theories.* Doctoral dissertation, Purdue University.

Docherty, G., Foulkes, P., Tilloston, J., & Watt, D. (2006) On the scope of phonological learning: Issues arising from socially structured variation. *Laboratory Phonology*, 8, 393-422.

Dietrich, C., Swingley, D., & Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of National Academy of Science*, 104, 16027-31.

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

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *Journal of the Acoustical Society of America*, 84, 115-123.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.

Harnsberger, J. D., Wright, R., & Pisoni, D. (2008) A new method for eliciting three speaking styles in the laboratory. *Speech Communication*, 50, 323-336.

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

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

Kewley-Port, D. & Preston, M. S. (1974) Early apical stop productions: a voice onset time analysis. *Journal of Phonetics*, 2, 195-210.

Kuhl, P. K., & Andruski, J. E. (1997). Cross-language analysis of phonetic units in language address to infants. *Science*, 227, 684-687.

Kraljic, T., Brennan, S. E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects and speech processing. *Cognition*, 107, 54-81.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141-178.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple talkers. *Journal of Memory and Language*, 56, 1-15.

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

Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K.C. (submitted) *Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development*. Manuscript under consideration.

Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

Liu, H. M., Kuhl, P., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6, F1-F10.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, 11, 122-134.

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

Munson, B., Edwards, J., & Beckman, M.E. (in press). *Phonological representations in language acquisition: Climbing the ladder of abstraction.* To appear in slightly different form in *Handbook of Laboratory Phonology* (A.C. Cohn, C. Fougeron, & M. K. Huffman, Eds.), Oxford: Oxford University Press. Downloaded on January 26, 2010 from http://www.tc.umn.edu/~munso005/.

Munson, B. (2001). A method for studying fricatives using dynamic measures of spectral mean. *Journal of the Acoustical Society of America*, 110, 1203-1206.

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of the spectral mean. *Journal of Speech, Language, and Hearing Research*, 47, 58-69.

Munson, B., Kaiser, E., & Urberg Carlson, K. (2008). *Assessment of children's speech production 3: Fidelity of responses under different levels of task delay*. Poster presented at the 2008 ASHA Convention, Chicago, 20-22. Downloaded on January 29, 2010 from http://www.tc.umn.edu/~munso005/.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21, 422-432.

Narayan, C. (2006). *Acoustic-perceptual salience and developmental speech perception.* Unpublished doctoral dissertation, University of Michigan.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. (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.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 2382-2393.

Schellinger, S., Edwards, J., Munson, B., & Beckman, M. E. (2008). *Assessment of phonetic skills in children 1: Transcription categories and listener expectations.* Poster presented at the 2008 ASHA Convention, Chicago 20-22, November 2008.

Scobbie, J.E., Gibbon, F., Hardcastle, W.J., & Fletcher,P. (2000) Covert contrast as a stage in the acquisition of phonetics and phonology. *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon.* Cambridge: Cambridge University Press, 194–203.

Sharf, D., Ohde, R., & Lehman, M. (1988). Relationship between the discrimination of /w-r/ and /t-d/ continua and the identification of distorted /r/. *Journal of Speech and Hearing Research*, 31, 193-206.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27, 501-532.

Stoel-Gammon, C., Williams, K., & Buder, E. (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica*, 51, 146-158.

Urberg-Carlson, K., Kaiser, E., & Munson, B. (2008). *Assessment of children's speech production 2: Testing gradient measures of children's productions.* Poster presented at the 2008 ASHA Convention, Chicago, 20-22. Downloaded on February 2, 2010 from http://www.tc.umn.edu/~munso005/.

Urberg-Carlson, K., Munson, B., & Kaiser, E. (2009). *Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness*. Poster presented at the spring 2009 meeting of the Acoustical Society of America. Also in *Journal of the Acoustical Society of America, 125*, 2529. Downloaded on January 26, 2010 from http://www.tc.umn.edu/~munso005/.

Werker, J. F., Shi, R., Desjardins, R. N., Polka, L., & Patterson, M. (1998). *Three methods for testing infant speech perception*. In A. M. Slater (Ed.), Perceptual development: Visual, auditory, and speech perception in infancy (p. 389-420). London:UCL Press.

Appendix A: Clear-Conversational Speech Style Elicitation Sentences

1. The boat ride felt safe
2. They lifted the sail
3. It was the same color
4. I saw that new movie last night
5. The girl put her favorite shell back on the beach
6. The child stopped saying her alphabet
7. The sheep were white and black
9. The costume included a shield
10. We learned about a mutiny on the ship
11. The boy lost one shoe
12. The kids shoot baskets after school
13. They had to shop for new clothes
14. She was too short for the rollercoaster
15. I got my flu shot this winter
16. My favorite show was cancelled
17. The child was home sick
18. Her sister looked older
19. We played soccer in the park
20. The mother always lost a sock in the dryer
21. We drank soda at the movie
22. The family bought a new sofa
23. The tomato soup tasted good on the cold day
24. He added sugar and cream to the coffee
25. The students won a super prize
26. I packed my suitcase before my vacation
27. Wash your hands with soap and water
28. He fell and hurt his shoulder
29. The knife had a sharp edge
30. He had to shave before the interview
31. It was just the right shape

Appendix B: Instructions

This experiment investigates adults' perceptions of children's speech, and how adults respond to children when they communicate with them. There are three steps that you will have to follow throughout the experiment.

First you will see a picture of an item and see the word written down that the picture was supposed to represent. For example, you might see a picture of a bowl of soup and the word "SOUP." You will hear a child naming that item but will only hear the first few sounds that they say. For example, you might see a picture of soup but just hear the child's production of the "s" and "ou" sounds.

The children in this experiment are of different ages and are at different stages of speech-sound development. Sometimes you will hear "s" and "sh" productions that are very accurate, and sometimes they will sound inaccurate.

You should listen carefully to how the child says the sounds, because we will ask you to rate the child's production. This rating will be made on a line. On one end is "the 's' sound" and the other end of the line there is "the 'sh' sound"

Click the mouse to see this line. It will stay on the screen for three seconds.

You will be clicking on the line to make your response. Click the mouse to have some practice just clicking on the line.

When you hear what you think is a PERFECT "s" sound, click on the line close to where it says "The 's' sound". When you hear what you think is a PERFECT "sh" sound, click on the line close to where it says "the 'sh' sound."

Click the mouse to continue the directions.

Sometimes, you won't be sure the syllable began with an "s" sound or an "sh" sound. In those cases, you should click the place on the line to show whether you thought it sounded more like "s" or more like "sh". If the sound wasn't really "s" or "sh" but sounded more like "s", then click somewhere on the line closer to the text that says "the 's' sound." If it sounds more like "sh," then click closer to the text that says "the 'sh' sound."

Click on the mouse to continue the directions.

We hope that you will use the whole line when rating these sounds. We don't have any specific instructions for what to listen for when making these ratings. We want you to go with your 'gut' feeling about what you hear at the beginning of the syllables.

Remember, after you rate the child's production, we want you to say the word the child was trying to say. Say the word as if you were responding to the child whose production you just rated. It might help if you think of each trial as a different interaction with a different child.

Click the mouse to do some practice items.

Now that you have done the practice items, you're ready to do the full experiment. Click the mouse to begin. It will begin as soon as you click the mouse, don't click it until you are sure that you are ready.

group	Centroids						Durations					
	s			S			s			S		
	β	ΔR²	*p*	β	ΔR²	*p*	β	ΔR²	*p*	β	ΔR²	*p*
	0.71	0.51	<0.001	0.66	0.44	<0.001	0.43	0.19	<0.001	0.34	0.12	<0.001
	0.46	0.002	<0.01	0.01	0.0001	ns	0.12	0.14	<0.001	-0.16	0.03	<0.001
s121	0.49	0.24	<0.001	0.16	0.027	ns	0.18	0.05	<0.05	0.08	0.02	ns
	0.06	0.004	ns	0.07	0.005	ns	0.30	0.08	0.01	-0.23	0.05	<0.01
s151	0.43	0.18	<0.001	0.27	0.08	<0.01	0.001	0.00	ns	0.13	0.02	ns
	0.09	0.01	ns	-0.07	0.01	ns	-0.06	0.003	ns	-0.11	0.01	ns
s169	0.33	0.11	<0.01	-0.20	0.04	<0.05	0.20	0.04	ns	0.19	0.04	<0.05
	-0.10	0.01	ns	-0.02	0.000	ns	-0.01	0.000	ns	0.04	0.002	ns
s172	-0.11	0.01	ns	-0.04	0.003	ns	0.02	0.000	ns	-0.10	0.01	ns
	0.19	0.04	ns	0.27	0.07	<0.01	-0.02	0.000	ns	-0.41	0.17	<0.001
s257	0.34	0.11	<0.01	0.06	0.004	ns	0.31	0.10	<0.01	0.49	0.002	ns
	0.09	0.01	ns	0.05	0.003	ns	-0.01	0.000	ns	0.003	0.000	ns
s279	0.38	0.13	<0.01	0.27	0.08	<0.01	0.12	0.02	ns	-0.04	0.001	ns
	-0.14	0.02	ns	-0.03	0.001	ns	0.07	0.01	ns	0.17	0.03	ns
s288	0.11	0.01	ns	0.07	0.01	ns	-0.02	0.001	ns	0.06	0.000	ns
	0.05	0.003	ns	0.13	0.02	ns	0.52	0.28	<0.001	-0.62	0.38	<0.001
s290	-0.12	0.01	ns	0.01	0.000	ns	-0.05	0.003	ns	0.10	0.01	ns
	-0.05	0.002	ns	0.01	0.000	ns	-0.05	0.003	ns	0.01	0.000	ns
s292	0.10	0.01	ns	0.14	0.02	ns	0.27	0.05	ns	0.60	0.01	ns
	0.06	0.003	ns	0.003	0.000	ns	0.32	0.10	<0.01	-0.50	0.25	<0.001
s294	0.28	0.10	<0.05	0.002	0.000	ns	-0.224	0.05	ns	0.11	0.01	ns
	-0.10	0.01	ns	-0.15	0.02	ns	0.11	0.01	ns	-0.32	0.10	<0.001
s296	0.12	0.01	ns	-0.07	0.01	ns	-0.01	0.003	ns	0.10	0.004	ns
	0.04	0.001	ns	-0.10	0.01	ns	0.20	0.04	ns	-0.59	0.34	<0.001
s297	0.06	0.003	ns	-0.04	0.002	ns	0.15	0.02	ns	-0.003	0.000	ns
	0.08	0.01	ns	-0.001	0.000	ns	-0.03	0.001	ns	0.04	0.002	ns
s298	0.67	0.45	<0.001	0.16	0.03	ns	0.01	0.001	ns	0.40	0.16	<0.001
	0.07	0.01	ns	0.12	0.02	ns	0.25	0.06	<0.05	-0.15	0.02	ns
s299	0.15	0.03	ns	0.04	0.002	ns	-0.04	0.002	ns	0.16	0.02	ns
	0.30	0.10	<0.05	-0.10	0.01	ns	0.12	0.014	ns	0.08	0.006	ns
s300	0.42	0.17	<0.001	0.06	0.003	ns	0.26	0.07	<0.05	0.06	0.01	ns
	0.13	0.16	ns	-0.03	0.001	ns	0.13	0.02	ns	-0.16	0.02	<0.01
s301	0.14	0.02	ns	-0.33	0.11	ns	0.09	0.10	ns	-0.11	0.02	ns
	0.07	0.004	ns	-0.01	0.000	ns	0.37	0.13	<0.01	-0.46	0.21	<0.001
s302	-0.12	0.01	ns	-0.04	0.001	ns	0.32	0.12	<0.01	-0.10	0.01	ns
	0.01	0.000	ns	0.20	0.04	<0.05	0.35	0.12	<0.01	-0.49	0.24	<0.001
s303	0.08	0.01	ns	-0.04	0.002	ns	0.09	0.01	ns	0.12	0.02	ns
	0.13	0.02	ns	0.10	0.01	ns	-0.15	0.02	ns	-0.25	0.06	<0.05
s304	0.19	0.04	ns	0.08	0.01	ns	0.33	0.11	<0.01	0.23	0.70	<0.01
	-0.05	0.003	ns	-0.27	0.07	<0.05	0.22	0.05	<0.05	-0.40	0.16	<0.001
s305	-0.02	0.001	ns	0.10	0.01	ns	0.23	0.05	<0.05	0.14	0.02	ns
	0.15	0.12	ns	-0.14	0.02	ns	0.12	0.02	ns	-0.13	0.02	ns
s306	0.44	0.20	<0.001	0.07	0.01	ns	0.16	0.03	ns	0.12	0.001	ns
	0.16	0.03	ns	0.07	0.004	ns	0.16	0.03	ns	-0.34	0.14	<0.001
s307	0.08	0.01	ns		0.000	ns	0.23	0.06	<0.05	-0.04	0.001	ns
	0.01	0.000	ns		0.01	ns	0.04	0.002	ns	0.15	0.02	ns

Table 1. Results of regressions predicting fricative centroid and duration from mouse-click location.

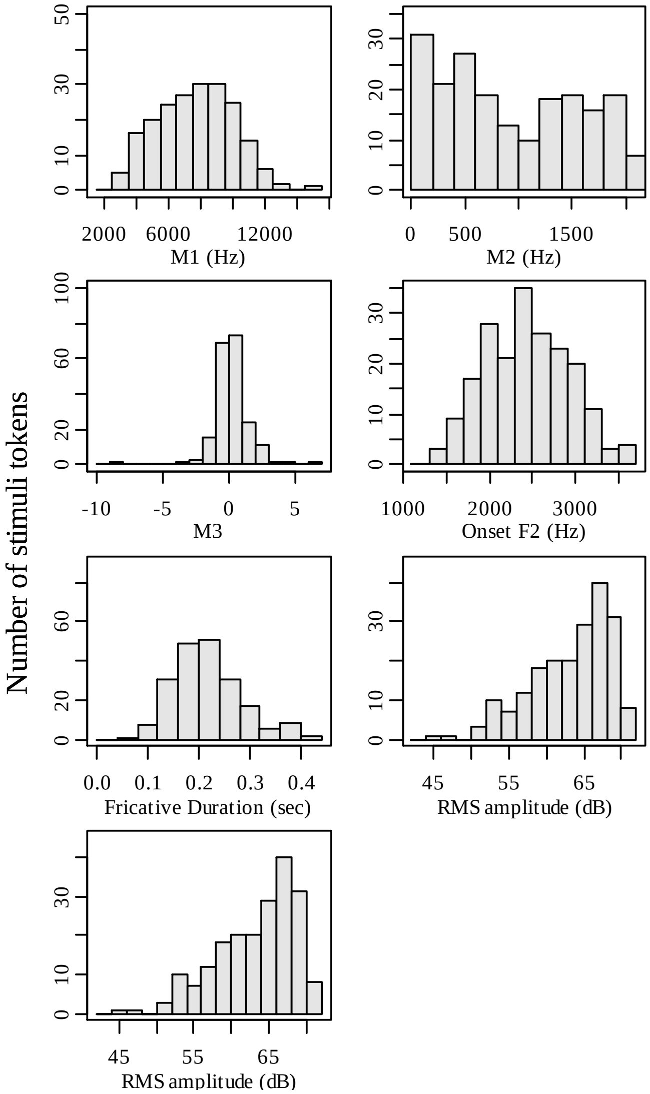


Figure 1. Histograms showing distribution of acoustics (M1, M2, M3, onset F2, Fricative Duration and RMS amplitude) of 200 stimulus items.

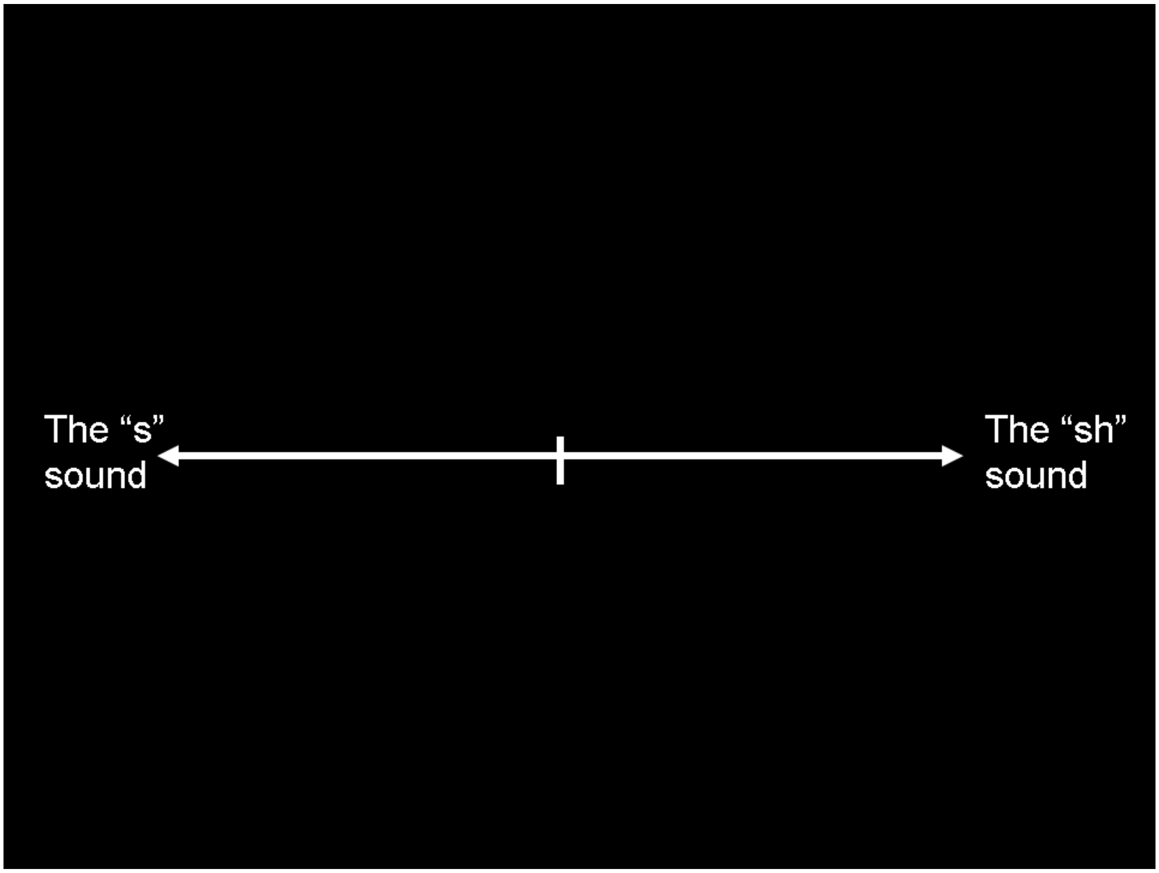
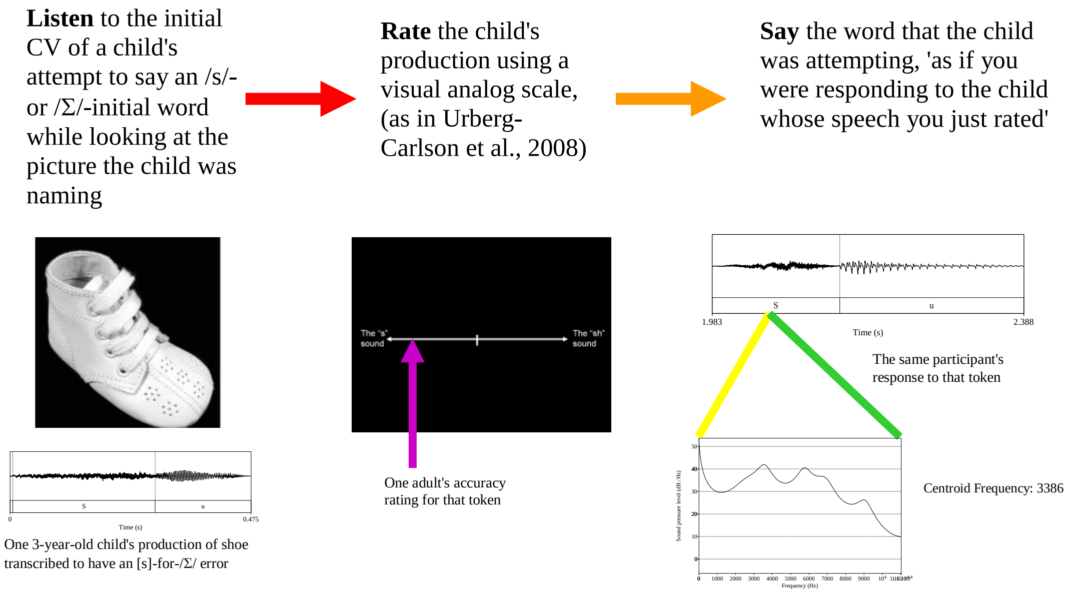


Figure 2. Picture of screen on which participants made their judgment (VAS rating)

The General Tactic



Listen to the initial CV of a child's attempt to say an /s/- or /Σ/-initial word while looking at the picture the child was naming

Rate the child's production using a visual analog scale, (as in Urberg-Carlson et al., 2008)

Say the word that the child was attempting, 'as if you were responding to the child whose speech you just rated'

One 3-year-old child's production of shoe transcribed to have an [s]-for-/Σ/ error

One adult's accuracy rating for that token

The same participant's response to that token

Centroid Frequency: 3386

Figure 3. Schematic representation of each Listen-Rate-Say (LRS) trial

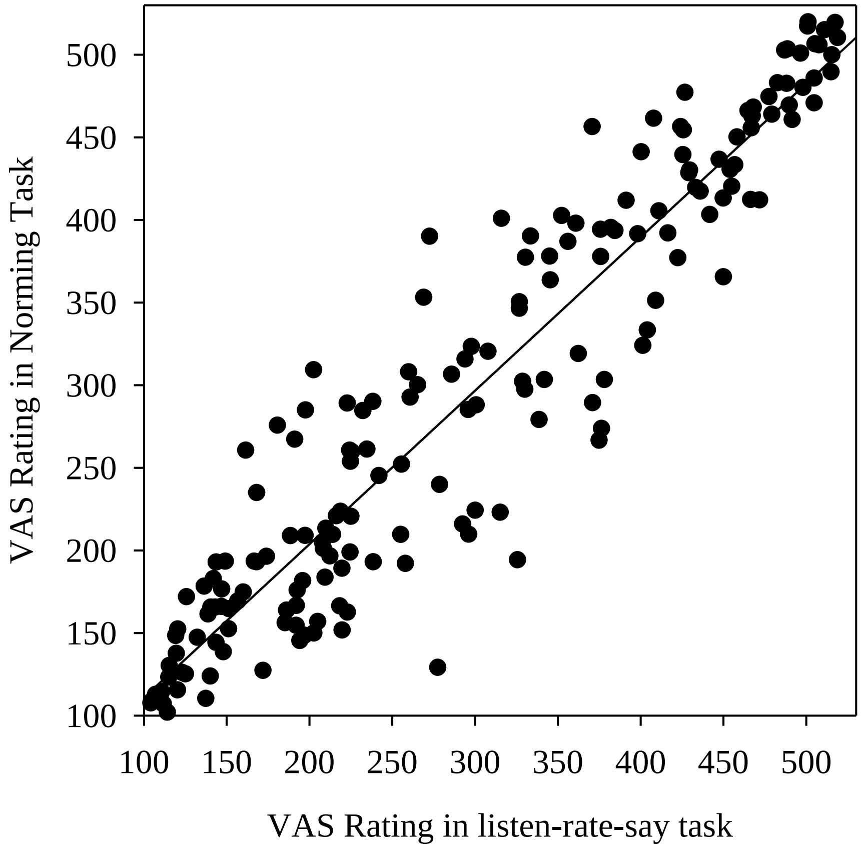


Figure 4. Scatterplot showing the relationship between the VAS ratings from the Urberg et al. (2008) and the VAS ratings from the present study. This suggests the perception of children's speech does not differ significantly when the target word is known by the rater nor when they provide a response following their judgment.

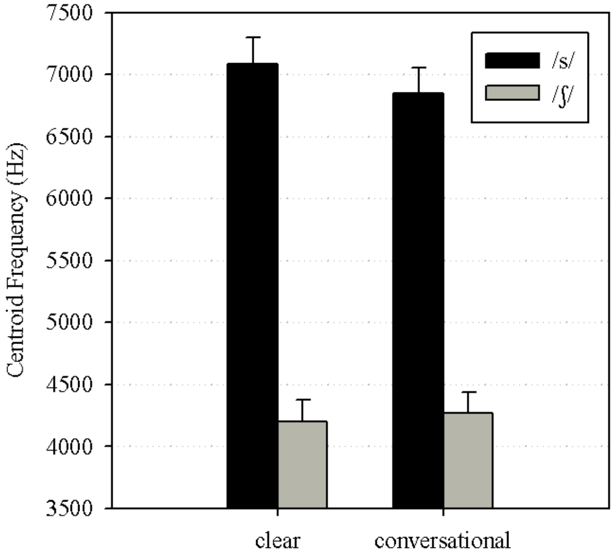


Figure 5. Centroids for /s/ and /ʃ/ from Clear-Conversational Speech Elicitation Task

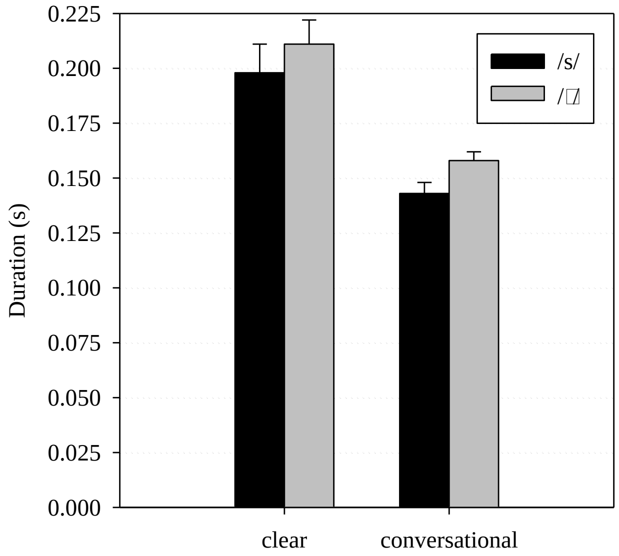


Figure 6. Durations for /s/ and /ʃ/ from Clear-Conversational Speech Elicitation Task

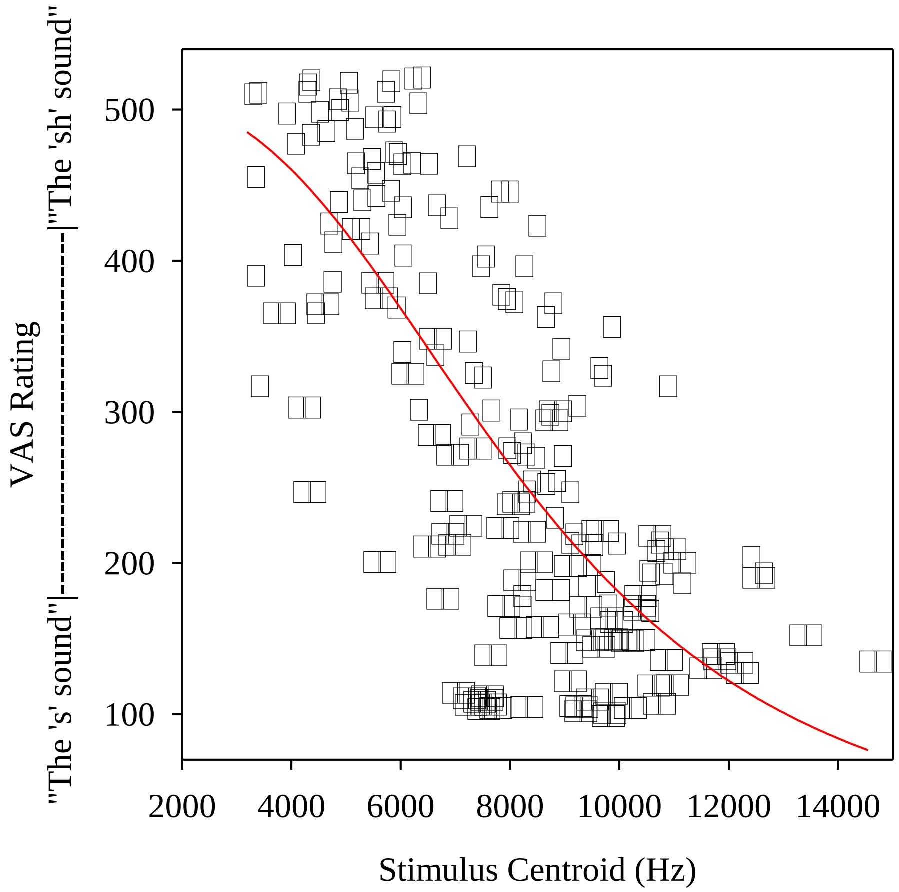


Figure 7. A scatterplot relating click location in the LRS task with the centroid frequencies of the tokens being rated. A nonlinear regression found that the sigmoidal function in the figure fit the data significantly, accounting for 56% of the variance. This was statistically significantly different from chance, F[2,197] = 125, p < 0.001. The IPA symbol refers to the target that the child was attempting, not to the token's perceived accuracy.

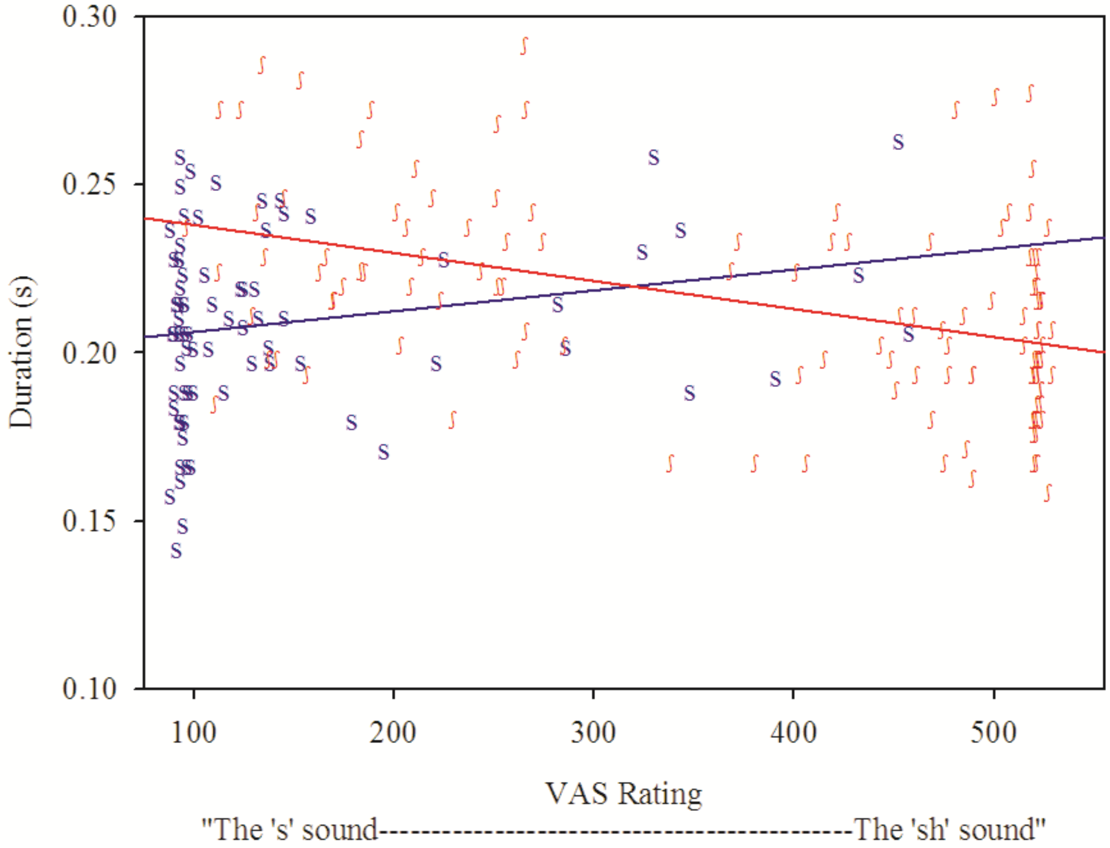


Figure 8 Individual data for subject 304. Showing a significant relationship between VAS rating and duration for both /s/ targets and /ʃ/ targets

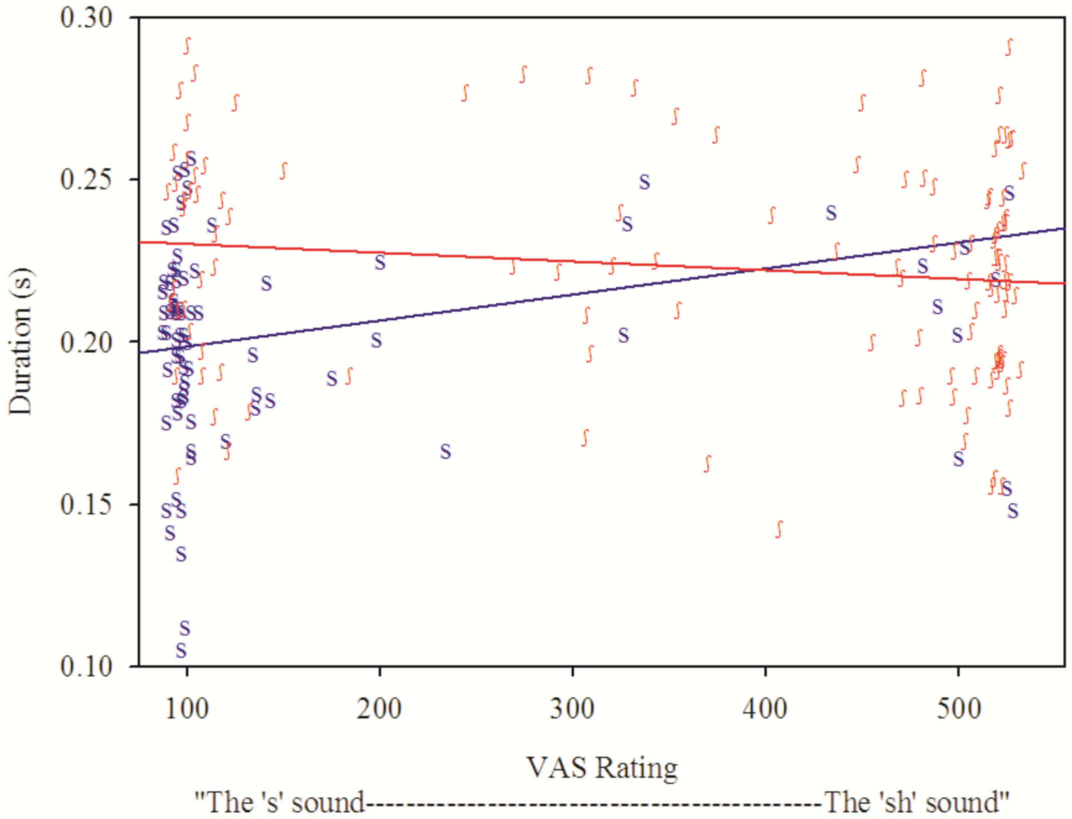


Figure 9. Individual data for subject 298. Showing a significant relationship between VAS rating and duration for /s/ targets but not for /ʃ/ targets

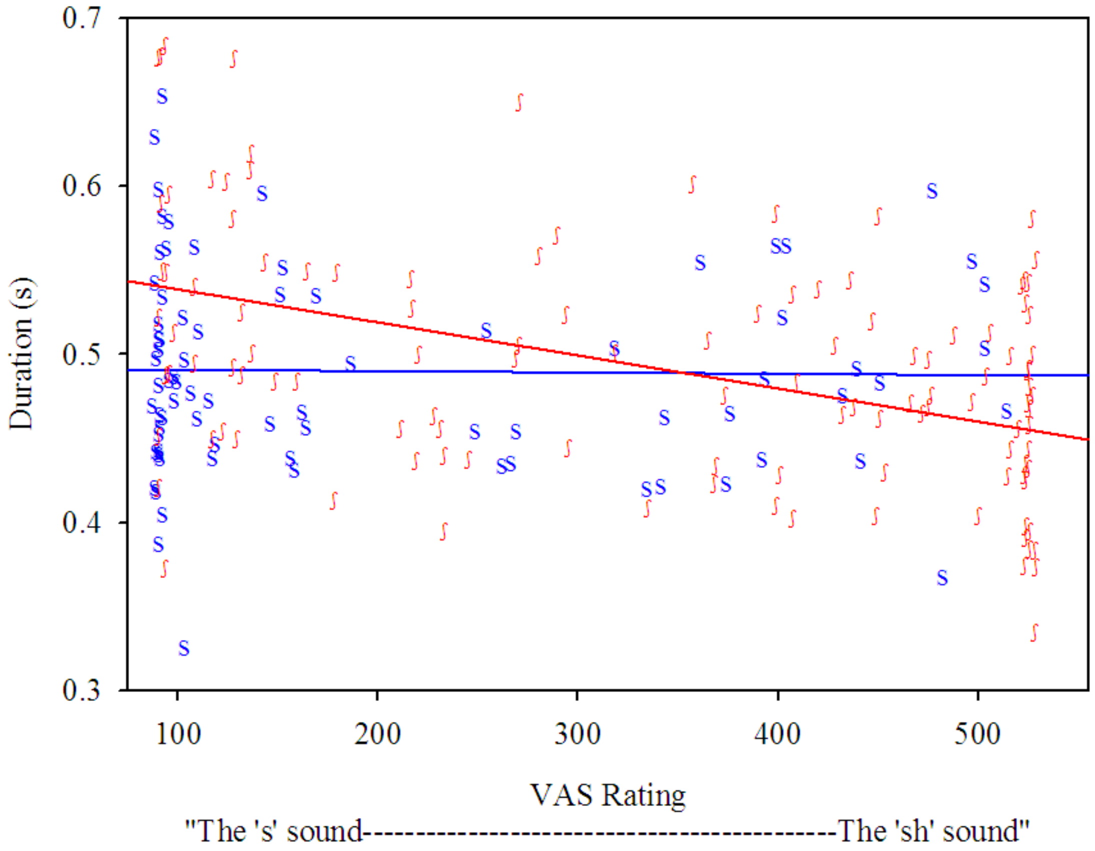


Figure 10. Individual data for subject 172. Showing a significant relationship between VAS rating and duration for /ʃ/ targets but not for /s/ targets