Published in final edited form as:
J Speech Lang Hear Res. 2012 December ; 55(6): 1836–1849. doi:10.1044/1092-4388(2012/11-0131).

Modifying Speech to Children based on their Perceived Phonetic Accuracy

Hannah M. Julien and **Benjamin Munson**
Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis

Abstract

Purpose—We examined the relationship between adults' perception of the accuracy of children's speech, and acoustic detail in their subsequent productions to children.

Methods—Twenty-two adults participated in a task in which they rated the accuracy of 2- and 3-year-old children's word-initial /s/and /ʃ/ using a visual analog scale (VAS), then produced a token of the same word as if they were responding to the child whose speech they had just rated.

Result—The duration of adults' fricatives varied as a function of their perception of the accuracy of children's speech: longer fricatives were produced following productions that they rated as inaccurate. This tendency to modify duration in response to perceived inaccurate tokens was mediated by measures of self-reported experience interacting with children. However, speakers did not increase the spectral distinctiveness of their fricatives following the perception of inaccurate tokens.

Conclusion—These results suggest that adults modify temporal features of their speech in response to perceiving children's inaccurate productions. These longer fricatives are potentially both enhanced input to children, and an error-corrective signal.

During language acquisition, learners must learn both to perceive and to produce the speech sounds in the language of their speech community. They must be able to do this within the context of immense variation in the phonetic forms that they are exposed to during acquisition, and which they are expected to produce as mature language user. Children hear speech produced from multiple talkers and in multiple contexts, and they must learn to control variation that is a socioindexical marker of talker identity and discourse register and also variation that is an allophonic marker of prosodic position and inter-word segmental coordination. This study attempts to further our understanding of this learning process by investigating the social dynamics of speech-sound acquisition. Specifically, it examines the relationship between adults' perception of children's speech and their subsequent productions to children. It is important to frame this relationship in the context of what we know about children's speech development, adults' ability to perceive children's productions accurately, and the interactions between adults' production and children's development.

Input

Early work on language acquisition in the generative tradition posited that input to children plays only a minimal role in language development, as the input cannot include the theoretically infinite variations in linguistic structure that adults might produce or comprehend (the so-called 'logical problem of language acquisition', Baker, 1979). More recent research has examined whether the input preverbal infants receive predispose them to

Address Correspondence to: Benjamin Munson Department of Speech-Language-Hearing Sciences University of Minnesota 115 Shevlin Hall 164 Pillsbury Drive Minneapolis, MN 55455 Munso005@umn.edu +1 612 612 0304 Fax: +1 612 624 7586.

learning linguistic categories. Much of this work has examined the acoustic characteristics of child-directed speech (DDS). Reviews of this work can be found in Soderstrom (2007) and in Munson, Edwards, and Beckman (in press). Work by Fernald and colleagues (Fernald, 2000; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Fernald & Simon, 1984) and by Kuhl and colleagues (e.g., Kuhl & Andruski, 1997; Liu, Kuhl, & Tsao, 2003) shows that, in addition to having higher and more modulated fundamental frequency, IDS is characterized by acoustically more distinct speech-sounds than adult directed speech (ADS). For example, Kuhl and Andruski found more-distinct vowels in IDS than in ADS in three languages, English, Swedish, and Russian. They noted that the vowel formants were not just higher (which would suggest that adults were merely imitating children's higher-frequency formants), 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 vowel contrasts more distinct perceptually, and also by providing a greater number of clear instances of the target vowels. The facilitative benefits of CDS are suggested by the fact that the expanded vowel spaces not observed in the prosodically similar speech style that is used when talking to pets (Burnham, Kitamura, & Volmmer-Conna, 2002). The distribution of vowels in CDS has been shown to support the learning of language-specific vowel contrasts, as shown both in observational (Werker, Ponsa, Dietrich, Kajikawa, Fais & Amano, 2007), and computational (Vallabha, McClelland, Pons, Werker, Amano, 2007) studies of English and Japanese vowel learning.

A smaller number of studies show that IDS includes exaggerated productions of consonant contrasts. Cristiá (2011) showed that American English IDS is characterized by more-distinct productions of the sibilant fricatives /s/ and /ʃ/. The degree of contrast was found to be dependent on the age of the child being spoken to. The contrast was exaggerated in the speech to 14- to 16-month old infants, but not to 6- to 8-month-old infants. A similar finding is reported by Sundberg and Lacerda (1999) for voice onset time (VOT), and suggests that IDS is maximally acoustically distinct when spoken to toddlers that are engaged in early word learning (i.e., around 14 months of age) than to toddlers that are not yet engaged in this process.

Output

A growing consensus in studies that examine speech-sound development is that children acquire contrasts among sounds gradually (for a review, see Hewlett & Waters, 2004). Two foundational studies on this topic examined gradual development of the voicing contrast in word-initial stop consonants. In word-initial position, English contrasts two voiceless stops, one with a short-lag voice-onset time (VOT), and one with a long-lag VOT. Kewley-Port and Preston (1974) examined children's productions longitudinally at the onset of meaningful words, and found that the distribution of VOTs in alveolar stops gradually moved from a relatively unimodal distribution to a more adult-like bimodal distribution. Macken and Barton (1980) also examined VOTs in the production of a small cohort of children longitudinally, and found that children progressed from a unimodal distribution of VOTs to a more adult-like, bimodal distribution. Critically, Macken and Barton showed that children produced distinct VOTs for target long-lag and short-lag VOTs prior to the point when they were perceptible to adults. The existence of these *covert contrasts—covert* because they were not perceptually salient to naïve listeners; *contrasts* because the target sounds were produced reliably differently—demonstrates that studies of language development that rely solely on phonetic transcription might not capture appropriately the continuous nature of speech-sound development.

Many studies subsequent to Macken and Barton have found evidence of covert contrast. Scobbie, Gibbon, Hardcastle, and Fletcher (2000) found covert contrasts in the production of

target /s/-stop and target stop sequences in children perceived to substitute stops for /s/-stop clusters. Baum and McNutt (1990) found covert contrasts between target /s/ and target /θ/ in children perceived to make [θ] for /s/ errors. Li, Edwards, and Beckman (2009) found evidence of covert contrast in English- and Japanese-acquiring children's /s/ and /ʃ/ productions. Tyler, Figurski, and Lansgdale (1993) demonstrated the clinical significance of covert contrasts by showing that children with speech-sound disorder who have covert contrasts progress through therapy more quickly and generalize correct production more readily than children who produce true neutralizations between target phonemes.

Interplay between Input and Output

The acoustic and articulatory characteristics of speech are highly variable, both across talkers and within talkers. At least one source of within-speaker variability is because talkers' speaking style is influenced by the audience, or the speech community within which they are communicating (e.g., Bell, 1984). Studies from adult laboratory phonology provide evidence that phonetic detail in productions can change depending on social dynamics. Two representative studies of this phenomenon are provided by Pardo (2006) and Babel (2009). These investigators examined the tendency for talkers to become progressively phonetically more similar to one another during an interaction. Pardo (2006) showed that phonetic convergence in a cooperative goal-oriented task varied as a function of talker sex and conversational role. Babel (2009) showed that phonetic convergence in a shadowing task was mediated by social factors: imitation occurred more when talkers had a positive perception of the talkers they were shadowing.

To our knowledge, no studies have examined the phonetic characteristics of adults' productions in adult-child interactions at a level of detail similar to that used by Pardo and Babel. Recent studies have shown, however, correspondences between individual differences in adults' productions and children's resulting perception abilities. Liu, Kuhl and Tsao (2003) found that Mandarin-speaking mothers produced an expanded vowel space when speaking to their infants, and that the size of the vowel space was positively correlated with the children's perception of a hard-to-perceive contrast between the fricative /s/ and the affricate /ts/. Recent research by Cristia studied the relationship between the extent to which caregivers contrasted /s/ and /ʃ/ and the ability of their children to perceive this contrast. Two different age groups of infants were studied: 4-6 month olds and 12-14 month olds. The adults' speech was recorded during natural interactions with their child and with the experimenter. On average, the caregivers to the older children more clearly differentiated /s/ from /ʃ/. Interestingly, an analysis of individual differences within the older aged group showed a relationship between the infants' perception abilities, and the difference between their caregiver's /s/ and /ʃ/.

Studies of covert contrast imply that adults' perception of children's speech is limited to perception of variation that crosses adult phonemic boundaries. A number of recent studies have challenged that notion. Munson, Edwards, Schellinger, Beckman, and Meyer (2010) showed that naïve listeners discern fine differences in children's speech. Munson et al. used a visual analog scaling task (VAS) in which listeners were presented with CV sequences produced by children in a picture-prompted word-repetition task. The picture prompts were all of words or nonwords beginning with /s/ or /θ/. Their productions had been transcribed by experienced clinicians as /s/, /θ/ or as a fricative intermediate between /s/ and /θ/. Listeners heard these syllables and clicked on a line anchored by the text "The 's' sound" and "The 'th' sound" where they perceived the sound to be relative to the endpoints. Munson et al. found that listeners rated correct productions of /s/ and /θ/ (i.e., [s] productions for target /s/ and [θ] productions for target /θ/) as more /s/- or /θ/-like, than substitutions (i.e., [s] productions for target /θ/, [θ] productions for target /s/). Moreover,

VAS click locations for individual tokens were found to correlate well with the acoustic characteristics of the fricatives, and in particular with the acoustic characteristics that differentiate /s/ from /θ/. Hence, performance on the VAS task showed that listeners can perceive more phonetic detail in children's speech than is typically denoted by phonetic transcriptions. Other research has extended the VAS scale for use with other contrasts using natural productions from the same corpus from which Schellinger et al selected their stimuli. Urberg-Carlson, Kaiser and Munson (2008) investigated the /s/-/ʃ/ contrast, Arbisi-Kelm, Edwards, Munson, and Kong (2010) examined the /d/-/g/ contrast, Johnson, Munson, and Edwards (2010) examined the /t/-/k/ contrast, and Kong (2009) examined the /t/-/d/ contrast. Johnson et al. showed that, while both laypeople and speech-language clinicians are able to perceive fine detail in children's speech, speech-language clinicians' ratings are even more-strongly correlated with the acoustic characteristics than are laypeople's.

Purpose of the Current Study

The research reviewed thus far in this article has argued that phonological acquisition is gradual, that the phonetic characteristics of the speech adults produce to language learners facilitates phonological category learning, and that adults can perceive more fine phonetic detail in children's productions than is typically coded in phonetic transcriptions. One question that remains is the extent to which the acoustic characteristics of adults' speech are affected by the specific child utterances that they are responding to in a particular interaction. In short, we know relatively little about the link between adults' perception of children's speech and their subsequent productions to children. Perhaps adults talk differently when they perceive a child's word productions to be nearly adult-like as compared to when they respond to a child whose word productions are perceived to be further away from the intended targets. That is, adults might produce hyperarticulated speech in interactions with children whom they perceive to be less accurate as compared to interactions with children whom they perceive to be accurate.

This study examines this possibility with an experiment in which adults engaged in simulated interactions with children whose speech they rated. The experiment targeted the English voiceless sibilant fricative contrast, using tasks that elicited adults' perceptual ratings of children's fricatives, and immediately subsequent productions of the same fricatives in simulated interactions to the child. This contrast was chosen because it has been studied extensively, and because it develops gradually. Work by Nittrouer and colleagues (e.g., Nittrouer, Stuttert-Kennedy, & McGowan, 1989) has shown that the development of sibilant fricatives involves gradual decreases in the coarticulatory effects of following vowels on fricative acoustics. Work by Li et al. (2009) showed that development involves gradual increases in the robustness of the acoustic difference between target /s/ and /ʃ/ productions.

Our first research question is whether adults produce speech differently in response to productions they judge to be inaccurate compared to ones they judge to be accurate. Specifically, we hypothesize that adults will hyperarticulate fricatives following productions that they had rated to be less accurate as compared to those they perceived to be more accurate.

Our second research question is whether there are systematic differences in the extent to which different adults modify their speech in response to hearing inaccurate productions by children. We address this question by examining whether any other measures of the talkers or of their productions that predict the extent to which they modify their speech in response to hearing inaccurate productions. Specifically, we examine two potential mediating variables: the extent to which adults modify the clarity of their speech in tasks in which they

are asked to do so explicitly, and self-reported experience in interacting with young children. We predict that individuals who make larger clear-speech modifications would modify their speech more in response to inaccurate tokens than those who make smaller modifications. We also predicted that people with more experience interacting with children would make larger modifications than those with relatively less experience.

Methods

Participants

Twenty four talkers (19 female, five male) participated in the study. They ranged in age from 19 to 49 years (M=25.8 years, SD=8.4). 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 participant produced too many unintelligible utterances to provide a full set of data. The final set of subjects included four men and 18 women. As part of their participation, individuals filled out a questionnaire regarding, among other things, the amount of time spent interacting with children in a typical week, ranging from 1 (none) to 10 (20 hours a week or more) on a 10-point equally appearing interval scale. The range of responses across the 22 subjects was 1 to 5. We refer to this measure as self-reported experience perceiving children's speech.

Stimuli

The 200 stimuli were from 24 (12 girls, 10 boys) different 2- and 3-year-old children's productions of the voiceless alveolar and voiceless palatoalveolar fricatives /s/ and /ʃ/. These recordings are described in Li et al. (2009). The productions were elicited during a scripted imitative naming task in which the children saw a picture (an item representing the target word) and were played a recorded verbal model of the target word. The words had a high frequency of occurrence, and were taken from lists of words expected to be in these children's expressive and receptive vocabularies. The fricatives and the first 150 ms of the following vowel were excised from the target word. Therefore, the stimuli in this experiment were not entire target words, but rather the initial CVs from the target words. Further details about this procedure can be found in Edwards and Beckman (2008). They were taken from a total of 30 different words.

The fricatives were selected to represent a range of the acoustic parameters that differentiate /s/ from /ʃ/ in adults' productions. These include the first four spectral moments, the second-formant frequency of the following vowel (*onset F2*), the fricatives' intensity, and the fricatives' durations. The fricatives were distributed quasi-normally along most of these parameters, with the exception of onset F2, which was relatively bimodal. These fricatives were also used in a perception study by Li, Munson, Edwards, Yoneyama, and Hall (2011); readers are referred to that study for more detail on how these stimuli were selected, their acoustic characteristics, and for data on a different group of adults' perception of these stimuli.

Procedures

Listen-Rate-Say Task—The primary task in this experiment is referred to henceforth as the *listen-rate-say* task (LRS) task. This is shown schematically in Figure 1. As this task's name implies, it involved three steps. First, the participant was presented with a stimulus item over headphones at a comfortable listening level. Each stimulus item was paired with the same picture which had been used to elicit the child's production in the original task described by Edwards and Beckman (2008). The auditory stimuli were presented over *Sennheiser HD280 Pro* Headphones. The visual component was presented on a 17″

computer monitor positioned directly in front of the participant. 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 at the other end. 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. The click location in pixels was logged automatically. Immediately following their judgment, participants were prompted to respond to the stimulus item. The picture and word were not played 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 five practice items, followed by 200 actual recorded trials. Please see Appendix B for the actual instructions provided to the participants during the experiment. The most critical part of these instructions was that talkers were asked to "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."

Clear-Speech Task—In this task, talkers produced clear- and conversational-style productions of sentences containing the 30 words that were used in the LRS task. See Appendix A for the list of sentences. The participants read the sentences from the computer screen. In the conversational-style condition, the participants were instructed to 'read the sentence.' In the clear-speech condition, 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 to an individual with a learning disability.

The entire experiment 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 clear-speech and the LRS tasks initially analyzed using a script which marked deviations from baseline amplitude. Each fricative boundary was then manually checked and marked by the first author, and by an undergraduate student who had specialized coursework in speech acoustics and extensive experience using Praat. The fricative-initial boundary was defined as the start of the aperiodic frication noise visible on the spectrogram, and the fricative-final boundary was defined the cessation of the aperiodic signal except in cases where there was formant-like structure in the aperiodic signal. An interval with such formant-like structure was interpreted as overlap between the glottal-opening gestures of the fricative and the lingual posture of following vowel, and was included as part of the vowel.

A variety of acoustic measures were extracted from the participants' productions. The first spectral moment of each fricative were extracted from a 40 ms window centered on the midpoint of the fricative. Spectral moments analysis (Forrest, Weismer, Milenkovic, & Dougall, 1988) treats the power spectrum as a random distribution, and summarizes it using a variety of measures, the most important of which is the mean (m1, also called the *centroid*) of the fricative. M1 has 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), and to differ between normal and intentionally hyperarticulated /s/ (Maniwa, Jongman, & Wade, 2009). The duration of the frication interval was also extracted from the data. Maniwa et al. showed

that hyperarticulated fricatives are produced with longer durations than those produced in a typical speech style. The duration, first formant (F1) and second formant (F2) frequencies of the following vowel at midpoint in Bark were also measured automatically from LPC formant tracks in Praat, and were hand-corrected for outlying values. The dispersion of vowels in the F1/F2 space was measured using the method described by Bradlow, Toretta, and Pisoni (1996). The degree of vowel dispersion is associated with more-intelligible speech (Bradlow et al., 1996), and is associated with child-directed speech (Kuhl et al., 1997).

An additional measure was collected from the LRS task, specifically, the location of the mouse click on the visual analog scale. This indicates the adults' perception of how accurate the children's productions were. We used these data in two ways. First, we examined whether the ratings in the LRS task were similar to ratings of the same tokens in previous studies using VAS described in Urberg-Carlson et al. (2008). The listeners in the previous studies 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 examined whether the acoustic characteristics of adults' productions were statistically related to their perception of the children's speech, as indexed by their mouse-click location.

The data from the clear-speech task were used for two different purposes. First, they were used as reference data in the analysis of responses on the LRS task. As described below, our analysis of each token from the LRS task compares a subject's production in that task to their production of the same word in the conversational-speech task, to control for effects of vowel context and potential word-specific articulations. Hence, the baseline clear/conversational task provides us with reference data to help understand some of the variability in the LRS task that arises from the phonetic and lexical characteristics of the words on that task, rather than on articulatory modifications made in response to hearing accurate or inaccurate productions by children.

Second, these baseline data were analyzed in their own right. A series of within-subjects ANOVAs were used to examine the extent to which the talkers modified the acoustic characteristics of fricatives and vowels when asked to produce clear speech. We derived measures of the difference in the various acoustic measures between the clear and conversational speech conditions. These were Cohen's d measures, and were used as factors in the analyses of the LRS data to determine whether they could predict the extent to which a given acoustic parameter changed in response to hearing accurate versus inaccurate productions by children.

Results

Clear-Speech Task

The first analysis examined differences between the clear and conversational speech tokens. A series of two-way analyses of variance (ANOVAs) were conducted with style (clear vs. conversational) and fricative (/s/ vs. /ʃ/) as the within-subject variables and m1, fricative duration, and vowel-space dispersion as the dependent measures. There were significant main effects of fricative for all three measures ($F[1, 21] = 852.8$, $p < 0.001$, $\eta^2_{partial} = 0.98$ for M1, ($F[1, 21] = 30.1$, $p = <0.001$, $\eta^2_{partial} = 0.65$ for duration, and $F[1, 21] = 12.1$, $p = 0.002$, $\eta^2_{partial} = 0.37$, for vowel dispersion. As in previous research (Jongman et al., 2000), /s/ was associated with a higher centroid, and a shorter duration than /ʃ/. There was a significant main effect of speaking style for duration ($F[1, 21] = 27.369$, $p < 0.001$, $\eta^2_{partial} = 0.566$) and vowel dispersion ($F[1, 21] = 53.0$, $p < 0.001$, $\eta^2_{partial} = 0.72$). Fricatives produced in clear speech styles were longer than those in conversational speech styles, and

vowels in clear-speech styles were produced with greater dispersion than those in conversational-speech styles. Finally, there was a significant interaction between style and fricative for m1 ($F[1, 21] = 5.844$, $p = 0.025$, $\eta^2_{partial} = 0.218$). This interaction occurred because the /s/ and /ʃ/ in the clear speech condition had higher and lower m1s, respectively, than in the conversational-speech condition. Two single-factor ANOVAs examined the effect of speech style (clear vs. conversational) on vowel duration and vowels-space dispersion. The latter measure was calculated as the mean Euclidean distance of each vowel from the center of the vowel space (i.e., the average F1 and F2) (Bradlow, Toretta, & Pisoni, 1995). The effect of style was significant in both of these ANOVAs ($F[1,21] = 31.885$, $p < 0.001$, η^2 partial = 0.603 for duration, $F[1,21] = 53.332$, $p < 0.001$, $\eta^2_{partial} = 0.717$ for dispersion). As in previous research, vowels in clear-speech styles were longer and more-dispersed than those in conversational speech styles.

To assess individual differences in the magnitude of the acoustic differences between the clear- and conversational-speech styles, a series of Cohen's d measures were calculated for six measures: the m1 of /s/ and /ʃ/, the duration of /s/ and /ʃ/, the duration of vowels, and the Euclidean distances of vowels from the center of the F1/F2 space. Cohen's D is calculated by taking the difference between means between two sets of data (in this case, measures in the conversational- and clear-speech conditions), and dividing them by the average of the variance of the two data sets. These are shown in Figure 2. As these figures show, there was a wide distribution of these Cohen's D measures. Overall, the largest values—indicating the strongest differences between conditions—were found for the fricative duration measures. The smallest differences were found for the fricative m1 measures. Vowel duration and Euclidean distance measures were intermediate.

LRS Task

Accuracy Measures—The first analysis examined the VAS ratings of the accuracy of children's productions from the LRS task. The first analysis of these data examined correlations between individual subjects' ratings and the mean ratings of the same stimuli made by subjects in Urberg-Carlson et al. (2008). Recall that participants in Urberg-Carlson et al. simply provided VAS judgments after hearing the same CV sequences that were used in this study. They were not aware of the words that the children were attempting, nor were they required to produce the fricative 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. Spearman's r values for correlations between individual participants' ratings in this study and average ratings in Urberg-Carlson et al. ranged from 0.66 to 0.91; the average was 0.84 (SD = 0.07). This illustrates that the ratings of fricative accuracy in this study were not substantially influenced by the knowledge of the target production.

The next analysis of the VAS ratings examined their relationship with acoustic characteristics of the stimuli. A series of hierarchical multiple regressions were performed with VAS click location as the dependent measure, and fricative m1 and duration as the independent measures. When the ratings averaged over the entire group of listeners were used as the dependent measure, a regression found that m1 and stimulus duration accounted for 53% of the variance in the dependent measure. The relationship between the m1 and average VAS ratings is shown in Figure 3. The R^2 for individual subjects' regressions ranged from 0.15 to 0.54, with an average of 0.39. All regressions were significant. For all 22 subjects, the β coefficient for m1 was significant. For seven of the 22 subjects, the β coefficient for stimulus duration was also significant. This finding establishes that listeners were attending to phonetic detail in the fricatives when judging the accuracy of children's productions.

ANOVAs—The final set of analyses examined whether the acoustic detail in adults' productions varied as a function of their perception of the accuracy of the child's speech that they had just heard. The first set of analyses used ANOVA to examine fricative productions. Each subject's 200 production were divided into two bins, those following the 50% of the stimuli whose rating were higher than the median rating for that target (i.e., those that were perceived to be more accurate), and those following the 50% whose ratings were lower than the median for that target (i.e., were perceived to be less accurate). The following illustrates how these values were calculated: for each subject, the median VAS rating for the /s/ targets and the /ʃ/ targets were determined separately, and each rating was coded as above or below these medians. The ratings that were closer to the /s/ end of the VAS scale were labeled as "more accurate" if the target was /s/ and "less accurate" if the target was /ʃ/. The average duration and m1 of the /s/ and /ʃ/ tokens that adults produced were averaged separately for words following less-accurate and more-accurate judgments. The 22 listeners' ratings were not perfectly rank-order correlated; hence, each subject had a unique set of stimuli in the 'more accurate' and 'less accurate' categories[1].

The first statistical test was a two-factor ANOVA with fricative (/s/, /ʃ/) and perceived accuracy (more accurate vs. less accurate) as within-subjects factors. When adults' m1 values were examined, there was, as expected, a strongly significant effect of fricative, $F[1,21] = 872.96$, $p < 0.001$, $\eta^2_{partial} = 0.977$. However, there was no effect of perceived accuracy, nor did perceived accuracy interact with fricative. In contrast, when fricative duration was examined, there were significant effects for both fricative, $F[1,21] = 5.545$, $p = 0.028$, $\eta^2_{partial} = 0.209$, and perceived accuracy, $F[1,21] = 12.787$, $p = 0.002$, $\eta^2_{partial} = 0.378$, and an interaction between them, $F[1,21] = 7.301$, $p = 0.013$, $\eta^2_{partial} = 0.258$. This interaction arose because there was not a difference between the duration of /s/ following less-accurate and more-accurate judgments (M = 246 ms, M = 243 ms, respectively), but there was a difference in the duration of /ʃ/ following less-accurate and more-accurate judgments (M = 264 ms, M = 240 ms, respectively). Specifically, participants produced significantly longer tokens of /ʃ/ in response to tokens of /s/ that they perceived to be misarticulated.

The next set of ANOVAs examined m1 and fricative duration measures from the LRS relative to the measures from the same subjects' productions of the same words in the conversational-speech condition of the baseline clear-conversational speech task. This was done as an additional control to ensure that the subject means for the less- and more-accurate tokens were not an artifact of the lexical composition of the tokens in those two bins. These relative measures were taken by simply subtracting the value from the LRS task from the value for the same word in the conversational-speech baseline task. These relative values were averaged separately for productions following tokens rated to be more and less accurate.

For the ANOVA examining relative m1 values, there were no main effects of either fricative, or of perceived accuracy, nor was there an interaction between them. This is consistent with the analysis of the untransformed values. For the relative duration values, there was not a significant main effect of fricative, but there was a significant main effect of accuracy, $F[1,21] = 7.742$, $p = 0.011$, $\eta^2_{partial} = 0.269$, and a significant interaction between these two factors, $F[1,21] = 6.482$, $p = 0.019$, $\eta^2_{partial} = 0.236$. The productions of /s/

[1]Recall that the children's fricatives were in a variety of vowel contexts. It is known that vowel context affects fricatives' acoustic characteristics. Hence, we did a series of 44 chi-squared contingency tests to examine whether the distribution of vowel contexts was significantly different for the more- and less-accurate tokens. Only one of these achieved statistical significance at the $\alpha < 0.05$ level. This was for one subjects' perception of /ʃ/ accuracy. That subject judged more /ʃ/ tokens in /a/ and /o/ contexts to be accurate than was expected by chance, and more tokens to be inaccurate in /u/ contexts.

following tokens to be less accurate were 96 ms longer than the baseline conversational-speech productions, while those following more-accurate tokens were only 86 ms longer than baseline productions. This difference did not achieve significance in a post-hoc paired-samples t-test at the conventional $\alpha < 0.05$ level, but did approach this value, $t[21] = -2.056$, $p = 0.052$. In contrast, the productions of /ʃ/ following tokens to be less accurate were 105 ms longer than the baseline conversational-speech productions, while those following more-accurate productions were only 81 ms longer than baseline productions. This difference was significant in a paired-samples t-test, $t[21] = -2.953$, $p = 0.008$.

The next set of ANOVAs examined the duration of the vocalic portions of the words produced in the LRS task. Recall that subjects modulated vowel durations in the baseline clear-conversational speech task. Hence, it is possible that they would also produce longer vowels in the LRS task following words rated to be less accurate as compared to words rated to be more accurate. When raw vowel durations were examined, there was no a significant main effect of perceived accuracy. As was done with the fricative measures, a set of relative vowel duration measures was calculated by taking the difference between the vowel durations in the LRS task and the durations of the vowels in the same words in the conversational-speech task. When these relative durations were examined, a significant effect of accuracy category was found, $F[1,21] = 29.507$, $p < 0.001$, $\eta^2_{partial} = 0.584$. Vowels in words produced after tokens rated to be less accurate were 45 ms longer than baseline tokens, while those after more-accurate tokens were 29 ms longer than baseline tokens.

Linear Mixed-Effects Regressions—The next set of analyses used Linear mixed effects regression with crossed random effects for subjects and items (henceforth LMER) to examine the LRS data. LMER is described by Baayen, Bates, and Davidson (2008), and allows for a number of analyses that cannot be conducted with ANOVA. LMER models the variance associated with both subjects and items as random effects, thus sidestepping the "language as fixed effects" fallacy (Clark, 1973). These analyses are meant to complement the ANVOAs reported in the previous section, rather than to duplicate them. LMER allows us to examine whether the relationship between the perceived accuracy of children's speech and characteristics of adults' productions are mediated by other factors. We examined whether the relationship between three dependent measures—fricative duration, fricative m1s, and vowel duration of the talkers' productions in the LRS task—and the perceived accuracy of children's speech was mediated by measures of the extent to which these talkers modified their speech in the clear-speech baseline task, by their self-reported experience interacting with children, and the relationship between their accuracy ratings in the LRS task and the acoustic characteristics of the stimuli that they were rating. We predicted that there would be stronger associations between perceived accuracy and acoustic measures in individuals who had bigger differences between clear and conservational speech characteristics, those whose ratings in the LRS task were more strongly associated with the acoustic characteristics of the speech being rated, and those who had more experience perceiving children's speech.

The first set of LMERs examined predictors of the relationship between the perception of accuracy in children's fricatives, and the duration of fricatives in adults' subsequent productions. The dependent measure in this model was the difference between the fricative duration in the LRS task and the duration of the fricative in the same word in the conversational-speech condition of the clear-speech task. For illustration, consider one participant's production of the word *safe*. Five tokens of this work were included as stimuli in the LRS task. The durations of the /s/ in this participant's five productions of this word were 166 ms, 128 ms, 164 ms, 203 ms, and 212 ms. The duration of /s/ in this participant's production of *safe* in the conversational-speech condition of the clear-speech task was 134 ms. The difference between these values was -32 ms, 6 ms, −30 ms, −69 ms, and −78 ms,

respectively, with smaller values indicating greater slowing relative to the durations in the baseline task. The measure was chosen to mitigate any word-specific effects on duration. The primary predictor of interest was perceived accuracy. For this analysis, the VAS click locations for /s/ and /ʃ/ targets from the LRS task were converted to a single scale, so that the analyses could examine both /s/ and /ʃ/ productions simultaneously. To do this, the VAS ratings for the /ʃ/ and /s/ targets were re-scaled so that they reflected the distance from the /s/ and /ʃ/ endpoints of the scale, thus creating a single dimension of perceived accuracy. A second predictor was the Cohen's d measure for the difference in fricative duration between the clear and conversational speech conditions of the baseline clear-speech task. Recall that this is an index of the extent to which individuals lengthened fricatives in the clear-speech condition of that task. A third predictor was the self-report measure of experience perceiving children's speech. In this and all analyses, random intercepts for subjects and items were included in the model.

In the model in which only perceived accuracy was included, its effect was statistically significant, $\beta = 0.0697$, $SE(\beta) = 0.0072$, $t = 9.729$, $p(>|t|) < 0.0001$. Adults' fricatives were longer following productions that they perceived to be less inaccurate. A second model was evaluated that included an interaction term between perceived accuracy and self-reported experience perceiving children's speech. The coefficient for this interaction was significant, $\beta = 0.0191$, $SE(\beta) = 0.0057$, $t = 3.377$, $p(>|t|) = 0.0007$. This interaction is illustrated by comparing the relationship between perceived accuracy and fricative duration for individuals with different levels of self-reported experience interacting with children, shown in Figure 4. As this Figure shows, there was a stronger relationship between perceived accuracy and fricative duration for the individuals who self-reported greater experience perceiving children's speech than for those who reported less experience.

The next set of LMERs examined predictors of the m1 values of /s/ and /ʃ/ in the LRS task. Perceived accuracy was predicted to have different effects on the m1 values of /s/ and /ʃ/: talkers were predicted to raise m1 for /s/ but lower m1 for /ʃ/ in response to hearing inaccurate tokens. Hence, separate models were run for /s/ and /ʃ/. Again, the dependent measure was the difference between m1 for a token in the LRS task and the m1 for the same word in the conversational speech condition of the baseline clear-speech task. There was no effect of perceived accuracy on relative m1 values for /s/ , nor did this factor interact with any of the other variables examined: experience perceiving children's speech, or the Cohen's d for the difference between clear- and conversational-speech m1 for /s/. For relative m1 of /ʃ/, the coefficient for the interaction between experience and relative m1 did not achieve statistical significance at the conventional $\alpha < 0.05$ level, but did approach that level, $\beta = 0.116$, $SE(\beta) = 0.066$, $t = 1.757$, $p(<|t|) = 0.079$. This interaction is shown in Figure 5. As this figure shows, individuals with more self-reported experience perceiving children's speech produced lower m1 values for /ʃ/ in response to hearing tokens they rated as less-accurate sounding. Given that low m1 values exaggerate the acoustic difference between /s/ and /ʃ/, this presumably had the effect making these tokens clearer-sounding. Those who reported less experience perceiving children's speech modified their /ʃ/ m1 values less, and in the opposite-than-predicted direction.

The final set of LMERs examined vowel duration. Again, the dependent measure in these analyses was the difference in vowel duration between the individual productions on the LRS task, and the duration of the vowel in the same word in the conversational speech portion of the baseline clear-speech task. In a model including only perceived accuracy as a predictor, its effect was statistically significant, $\beta = 0.04363$, $SE(\beta) = 0.01146$, $t = 3.806$, $p(<|t|) = 0.0001$. This effect did not interact with self-reported experience perceiving children's speech, or with variance accounted for in the regression predicting VAS ratings from the acoustic characteristics of the stimuli. It did, however, interact with a measure of

the difference in vowel duration between the clear- and conversational-speech conditions of the baseline clear-speech task, namely, the Cohen's d measure comparing the mean vowel durations between these conditions, $\beta = -0.07583$, $SE(\beta) = 0.03491$, $t = -2.173$, $p(<|t|) = 0.0299$. This interaction is shown in Figure 6. As this figure shows, the participants whose productions showed strongest relationships between perceived accuracy and vowel duration were those whose vowel durations in the clear- and conversational-speech conditions were the smallest. As this Figure also shows, the participants whose vowel durations between the two conditions of the baseline task differed the most produced uniformly longer durations in the LRS task; hence, this interaction likely reflects the use of a clearer-speech style overall in the LRS task by those who made the biggest clear-speech modulations in the baseline task.

Discussion

This study examined whether adults modify the acoustic characteristics of their speech in response to hearing children's productions that they perceive to be inaccurate, relative to ones they perceive to be accurate. This was examined using a laboratory task—the listen-rate-say (LRS) task—intended to mimic the social dynamics of adult-child communication dyads. The stimuli were fricative-vowel sequences excised from real productions of English-acquiring children. Consistent with our hypothesis, we found that the adults did modulate aspects of their own speech after hearing children's speech that they perceived to be inaccurate. Specifically, they produced longer fricatives and vowels after hearing children's productions that they perceive to be inaccurate. These findings were observed both when ANVOA was used and when linear mixed-effects regression (LMER) was used to analyze the data. LMERs showed that the extent to which adults modified their fricative and vowel durations was mediated by their performance on other speech-production tasks, and by their experience with children. The talkers who made the largest modifications in vowel duration in response to hearing speech that they perceived to be inaccurate were those who modified the duration of their vowels in a baseline clear-speech task *least*. This unexpected finding may be due to the fact that the adults who made the largest modulations in vowel durations on the baseline clear-speech task produced uniformly long vowels on the LRS task. Adults who reported more experience interacting with children modified their fricative durations more than those who reported less experience. One interpretation of this is that spending time with children makes someone more willing to accommodate their production to children's specific communicative needs. The LMER analysis also found a weak relationship between perceived accuracy and the spectral characteristics of /ʃ/, though this was true only for the participants who reported more experience perceiving children's speech. The influence of experience on the association between perceived accuracy and subsequent phonetic modification is reminiscent of Rowe (2008), who found that caregivers' knowledge of child development was a better predictor of the use of different CDS morphosyntactic and lexical forms than was their socioeconomic status, or their own verbal fluency.

It is noteworthy that there was a mismatch between the acoustic characteristics that adults used in rating children's speech, and the parameters that they modified in their own speech. The VAS ratings of the accuracy of children's speech were predicted most strongly by a spectral property of their fricatives, the first spectral moment (m1). M1 is the acoustic measure that most-effectively differentiates between /s/ and /ʃ/. In contrast, the adults modified the duration of sounds, producing longer speech sounds—both fricatives and vowels—in response to hearing inaccurate fricatives, rather modifying their own m1. This finding is consistent with the findings of the baseline clear-speech task. It is also consistent with the findings Maniwa, Jongman, and Wade (2009), who showed that talkers produced much larger differences in fricative duration than m1 in intentionally clear speech than in conversational speech. In general, this tendency may stem from the fact that articulatory-

acoustic relationships for fricatives are highly nonlinear. That is, small changes in articulatory postures for fricatives do not always result in similar small changes in the resulting acoustic output (e.g. Stevens, 1989, cf. Tabain, 2001). Given this, the most effective production strategy for improving the clarity of fricatives is to simply lengthen them.

There are two important implications of these findings. The first concerns the proper description of child-directed speech (CDS) styles. These are typically described in research studies as being categorically different from other speech styles. The finding in this studies suggest that CDS may be a gradient phenomenon, and that individuals modulate their use of hyperarticulated forms in response to changing percepts of the communicative needs of their child interlocutors. This hypothesis is supported by findings from cross-sectional studies, such as those by Cristia (2011) and by Sundberg and Lacerda (1999), which show that the consonant hyperarticulations in CDS are strongest when the children who are being addressed are at an age when they are strongly engaged in learning new words. It is also consistent with many models of speech production. Both the H&H model (Lindblom, 1990) and the Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004) propose that listeners modulate the degree of hyperarticulation and reduction in their speech to maintain a consistent level of intelligibility. These modifications can occur when a word is predicted to be difficult to perceive, based on its low frequency of occurrence or its similarity to many other real words in the lexicon (e.g., Munson & Solomon, 2004), or when the listening environment is unfavorable, or when the speaker is presumed to need require an especially clear signal because of a hearing impairment or low language proficiency.

The second implication concerns the status of error-correcting feedback in language acquisition. There is a long-standing claim that adults' responses to children's incorrect utterances (e.g., Bohannon & Stanowicz, 1988) are not sufficiently robust to serve as negative evidence, at least for syntactic acquisition (e.g, Marcus, 1993). It is possible that hyperarticulations in response to incorrect productions may serve as robust error-correcting feedback for the acquisition of phonemes. This would be consistent with the work of Saxton (1997), who argued that the contrast between children's productions and adults' responses to them serve as implicit negative evidence during language acquisition. This possibility remains to be tested. The data described in this project were collected as part of a larger project that uses unsupervised and semi-supervised computational learning algorithms to model phoneme acquisition (Plummer, Beckman, Belkin, Fosler-Lussier, & Munson, 2010). A finding that this type of hyperarticulated feedback to incorrect productions does facilitate phoneme acquisition would show that there is sufficient negative evidence in phoneme acquisition.

This study had at least two limitations that should be dealt with in future research. The first concerns the measures that we used. In this study, as in many others, acoustic measures were used as indices of a perceptual phenomenon, speech clarity. The assertion that longer fricatives are clearer is based on previous research (i.e., Maniwa et al., 2009); however, we have not tested directly whether the modifications that adults made in this study would indeed be perceived as better exemplars by the children whose productions were used as stimuli. The second limitation concerns the ecological validity of this study. A clear next step in this research program is to examine the effect of perceived accuracy on acoustic detail in natural caregiver-child dyads. Parents' productions from natural interactions could be acoustically analyzed and compared to their off-line judgments of the accuracy of their children speech, using the same visual analog scaling method used in the LRS task. A failure to find a similar relationship in natural dyads would suggest that the findings in this paper are a consequence of the somewhat artificial character of the LRS task. Such work could also examine the next logical step in the caregiver-child dyad, children's subsequent

productions. A finding that children's productions more closely approximate adult forms following a hyperarticulated response would be particularly powerful evidence that these serve an error-correcting function. More generally, future research on this topic should examine whether the modifications that adults make in response to hearing inaccurate speech do in fact facilitate children's acquisition of spoken language.

Acknowledgments

This work could not have been accomplished without input and support from Mary E. Beckman and Jan Edwards. Portions of this research were conducted as part of the first author's 2010 MA thesis in the Department of Speech-Language-Hearing Sciences at the University of Minnesota. We generously thank Gerald Burke for many hours of volunteer work event-marking fricatives. We thank Kari Urberg-Carlson, Marie Meyer, and Eden Kaiser for assistance with subject testing. We thank Peter Watson and Mary E. Beckman for serving as committee members, and for providing many useful comments on this work when it was in progress. This research was supported by NSF grant BCS 0729277 to Benjamin Munson and NIH Grant R01 DC02932 to Jan Edwards.

Appendix A: Clear-Conversational Speech Style Elicitation Sentences. Target words are in bold

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
8. The costume included a ***shield***
9. We learned about a mutiny on the ***ship***
10. The boy lost one ***shoe***
11. The kids ***shoot*** baskets after school
12. They had to ***shop*** for new clothes
13. She was too ***short*** for the rollercoaster
14. I got my flu ***shot*** this winter
15. My favorite ***show*** was cancelled
16. The child was home ***sick***
17. Her ***sister*** looked older
18. We played ***soccer*** in the park
19. The mother always lost a ***sock*** in the dryer
20. We drank ***soda*** at the movie
21. The family bought a new ***sofa***
22. The tomato ***soup*** tasted good on the cold day
23. He added ***sugar*** and cream to the coffee

24. The students won a ***super*** prize
25. I packed my ***suitcase*** before my vacation
26. Wash your hands with ***soap*** and water
27. He fell and hurt his ***shoulder***
28. The knife had a ***sharp*** edge
29. He had to ***shave*** before the interview
30. 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” 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 “s” sound, click on the line close to where it says “the ‘sh’ sound.” 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.” 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.

References

Arbisi-Kelm T, Edwards J, Munson B, Kong E. Cross-linguistic perception of velar and alveolar obstruents: A perceptual and psychoacoustic study. Poster presentation at the Acoustical Society of America, also in Journal of the Acoustical Society of America. 2010; 127:1957.

Aylett M, Turk A. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Language and Speech. 2004; 47:31–56. [PubMed: 15298329]

Baayen H, Bates D, Davidson D. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Babel, M. Doctoral dissertation. University of California; Berkeley: 2009. Phonetic and Social Selectivity in Speech Accommodation..

Baker CL. Syntactic theory and the projection problem. Linguistic Inquiry. 1990; 10:533–81.

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

Baum SR, McNutt JC. An acoustic analysis of frontal misarticulation of /s/in children. Journal of Phonetics. 1990; 18:51–63.

Bell A. Language style as audience design. Language in Society. 1984; 13:145–204.

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer, Version 4.6.02 [Computer Program]. 2005.

Bohannon J, Stanowicz L. The issue of negative evidence: adult responses to children's language errors. Developmental Psychology. 1988; 24:684–689.

Bradlow AR, Torretta T, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication. 1996; 20:255–272. [PubMed: 21461127]

Burnham D, Kitamura C, Vollmer-Conna U. What's New Pussycat? On talking to babies and animals. Science. 2002; 296:1435. [PubMed: 12029126]

Cristià A. Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. Journal of the Acoustical Society of America. 2011; 129:3271–3280. [PubMed: 21568428]

Fernald A. Speech to infants as hyperspeech: Knowledge-driven process in early word recognitions. Phonetica. 2000; 57:242–254. [PubMed: 10992144]

Fernald A, Simon T. Expanded intonation contours in mothers' speech to newborns. Developmental Psychology. 1984; 20:104–113.

Fernald, Anne; Pinto, JP.; Swingley, D.; Weinberg, A.; McRoberts, GW. Rapid gains in speed of verbal processing by infants in the 2nd year. Psychological Science. 1998; 9:228–231.

Forrest K, Weismer G, Milenkovic P, Dougall R. Statistical analysis of word-initial voiceless obstruents: preliminary data. Journal of the Acoustical Society of America. 1988; 84:115–123. [PubMed: 3411039]

Hewlett N, Waters D. Gradient change in the acquisition of phonology. Clinical Linguistics & Phonetics. 2004; 18:523–533. [PubMed: 15573488]

Johnson, JM.; Munson, B.; Edwards, J. The role of clinical experience in listening for phonetic detail in children's speech.. Poster presentation at the Symposium for Research in Child Language Disorders.; Madison, WI. June 3-5; 2010.

Jongman A, Wayland R, Wong S. Acoustic characteristics of English fricatives. Journal of the Acoustical Society of America. 2000; 108:1252–1263. [PubMed: 11008825]

Kewley-Port D, Preston MS. Early apical stop productions: a voice onset time analysis. Journal of Phonetics. 1974; 2:195–210.

Kong, E. The Development of Phonation-type Contrasts in Plosives: Cross-linguistic Perspectives. Unpublished Ph.D. Dissertation. Department of Linguistics.; Ohio State University; Columbus, OH: 2009.

Kuhl PK, Andruski JE, Chistovich L, Chistovich I, Kozhevnikova E, Sundberg U, Lacerda F. Cross-language analysis of phonetic units in language address to infants. Science. 1997; 227:684–687. [PubMed: 9235890]

Li F, Edwards J, Beckman M. Contrast and covert contrast: The phonetic development of the voiceless sibilant fricatives in English and Japanese toddlers. Journal of Phonetics. 2009; 37:111–124. [PubMed: 19672472]

Li F, Munson B, Edwards J, Yoneyama K, Hall KC. Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. In publication in Journal of the Acoustical Society of America. 2010

Lindblom, B. Explaining phonetic variation: a sketch of the H&H theory.. In: Hardcastle, W.; Marchal, A., editors. Speech production and speech modeling. Kluwer; Dordrecht: 1990. p. 403-439.

Liu HM, Kuhl P, Tsao FM. An association between mothers' speech clarity and infants' speech discrimination skills. Developmental Science. 2003; 6:F1–F10.

Maniwa K, Jongman A, Wade T. Acoustic characteristics of clearly spoken English fricatives. Journal of the Acoustical Society of America. 2009; 125:3962–3973. [PubMed: 19507978]

Marcus G. Negative evidence in language acquisition. Cognition. 1993; 46:53–85. [PubMed: 8432090]

Munson, B.; Edwards, J.; Beckman, ME. Phonological representations in language acquisition: Climbing the ladder of abstraction.. In: Cohn, AC.; Fougeron, C.; Huffman, MK., editors. To appear in slightly different form in Handbook of Laboratory Phonology. Oxford University Press; Oxford: in press Downloaded on January 26, 2010 from http://www.tc.umn.edu/~munso005/

Nissen S, Fox RA. Acoustic and spectral characteristics of young children's fricative productions: a developmental perspective. Journal of the Acoustical Society of America. 2005; 118:2570–2578. [PubMed: 16266177]

Nittrouer S, Studdert-Kennedy M, McGowan R. 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. 1989; 32:120–132. [PubMed: 2704187]

Pardo J. On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America. 2006:2382–2393. [PubMed: 16642851]

Plummer, A.; Beckman, ME.; Belkin, M.; Fosler-Lussier, E.; Munson, B. In the Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010). Makuhari, Japan: 2010. Learning speaker normalization using semisupervised manifold alignment.; p. 2918-2921.ISSN 1990-9772

Rowe M. Child-directed speech: relation to socioeconomic status, knowedge of child development and child vocabulary skill. Journal of Child Language. 2008; 35:185–205. [PubMed: 18300434]

Saxton M. The contrast theory of negative input. Journal of Child Language. 1997; 24:139–161. [PubMed: 9154012]

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

Scobbie, JE.; Gibbon, F.; Hardcastle, WJ.; Fletcher, P. Papers in Laboratory Phonology V: Language Acquisition and the Lexicon. Cambridge University Press; Cambridge: 2000. Covert contrast as a stage in the acquisition of phonetics and phonology.; p. 194-203.

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

Stevens K. On the quantal nature of speech. Journal of Phonetics. 1989; 17:3–46.

Sundberg U, Lacerda F. Voice onset time in speech to infants and adults. Phonetica. 1999; 56:186–199.

Tabain M. Variability in fricative production and spectra: implications for the Hyper- & Hypo- and Quantal theories of speech production. Language and Speech. 2001; 44:58–93.

Tyler AA, Figurski GR, Langdale T. Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. Journal of Speech and Hearing Research. 1993; 36:746–759. [PubMed: 8377487]

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

Urberg-Carlson K, Munson B, Kaiser E. 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. 2009; 125:2529. Downloaded on January 26, 2010 from http://www.tc.umn.edu/~munso005/.

Vallabha G, McClelland J, Pons F, Werker J, Amano S. Unsupervised learning of vowel categories from infant-directed speech. Proceedings of the National Academy of Sciences. 2007; 104:13273–13278.

Werker J, Pons F, Dietrich C, Kajikawa S, Fais L, Amano S. Infant-directed speech supports phonetic category learning in English and Japanese. Cognition. 2007; 103:147–162. [PubMed: 16707119]

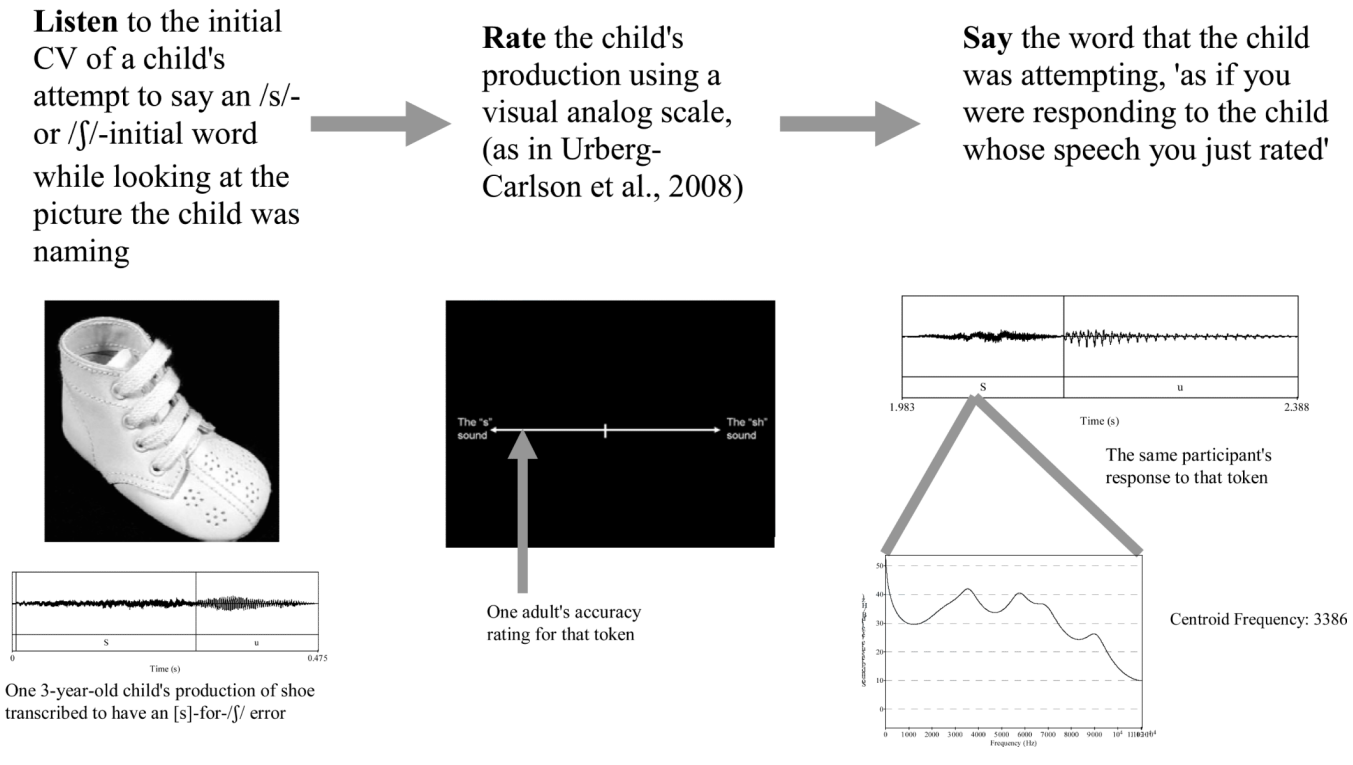

Figure 1.
Schematic representation of the Listen-Rate-Say (LRS) task

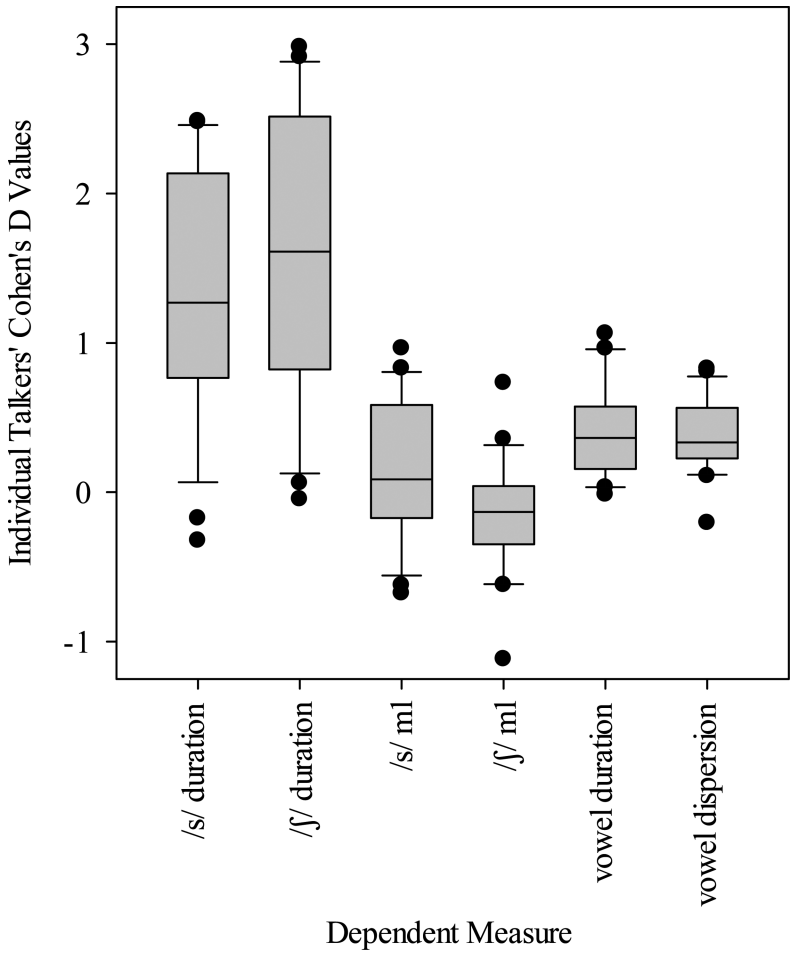

Figure 2.
Cohen's d values for the difference between clear and conversational speech conditions for six acoustic measures from the baseline clear-speech task.

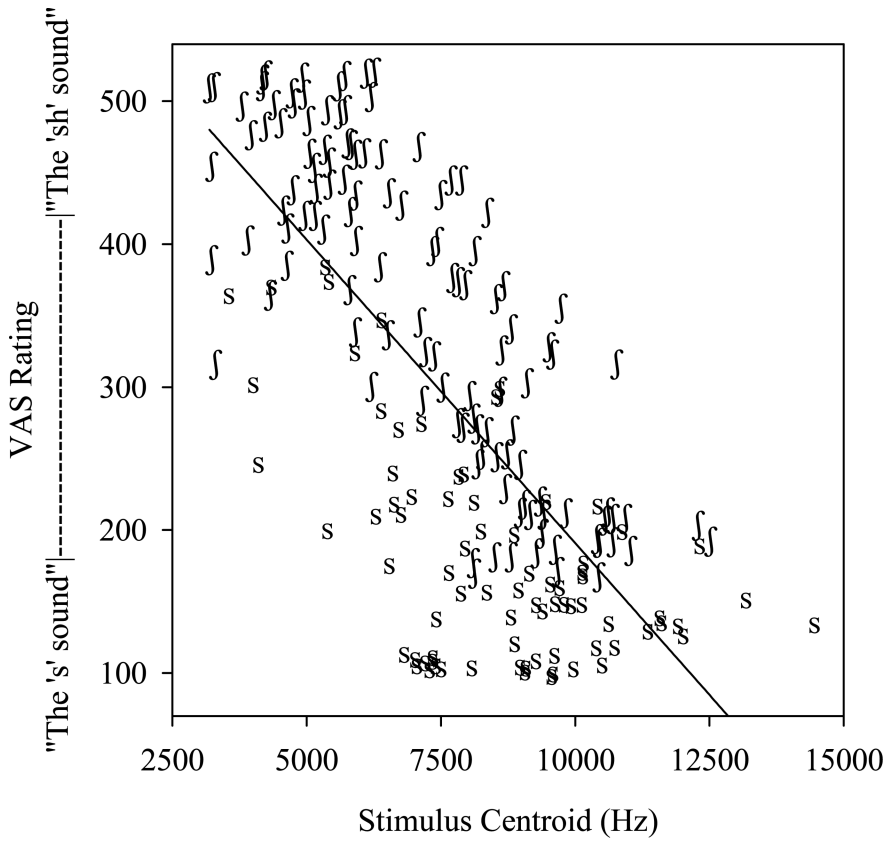

Figure 3.
Scatterplot showing relationship between VAS rating and m1s of stimuli.

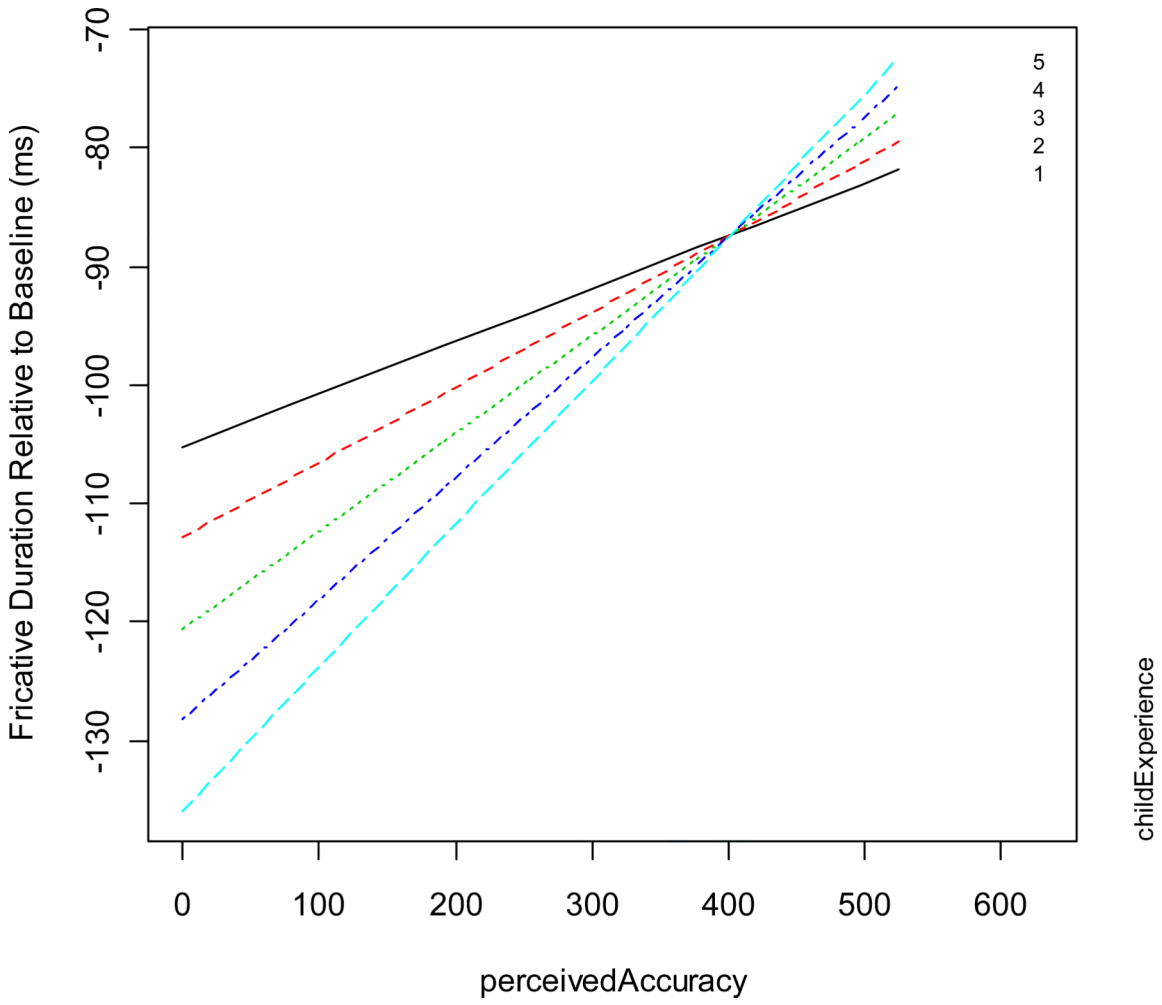

Figure 4.
Line-plot showing the mediating effect of self-reported experience perceiving children's speech (from 1 [less] to 5 [more]) on the relationship between perceived accuracy and relative fricative duration in the LRS task. See text for details.

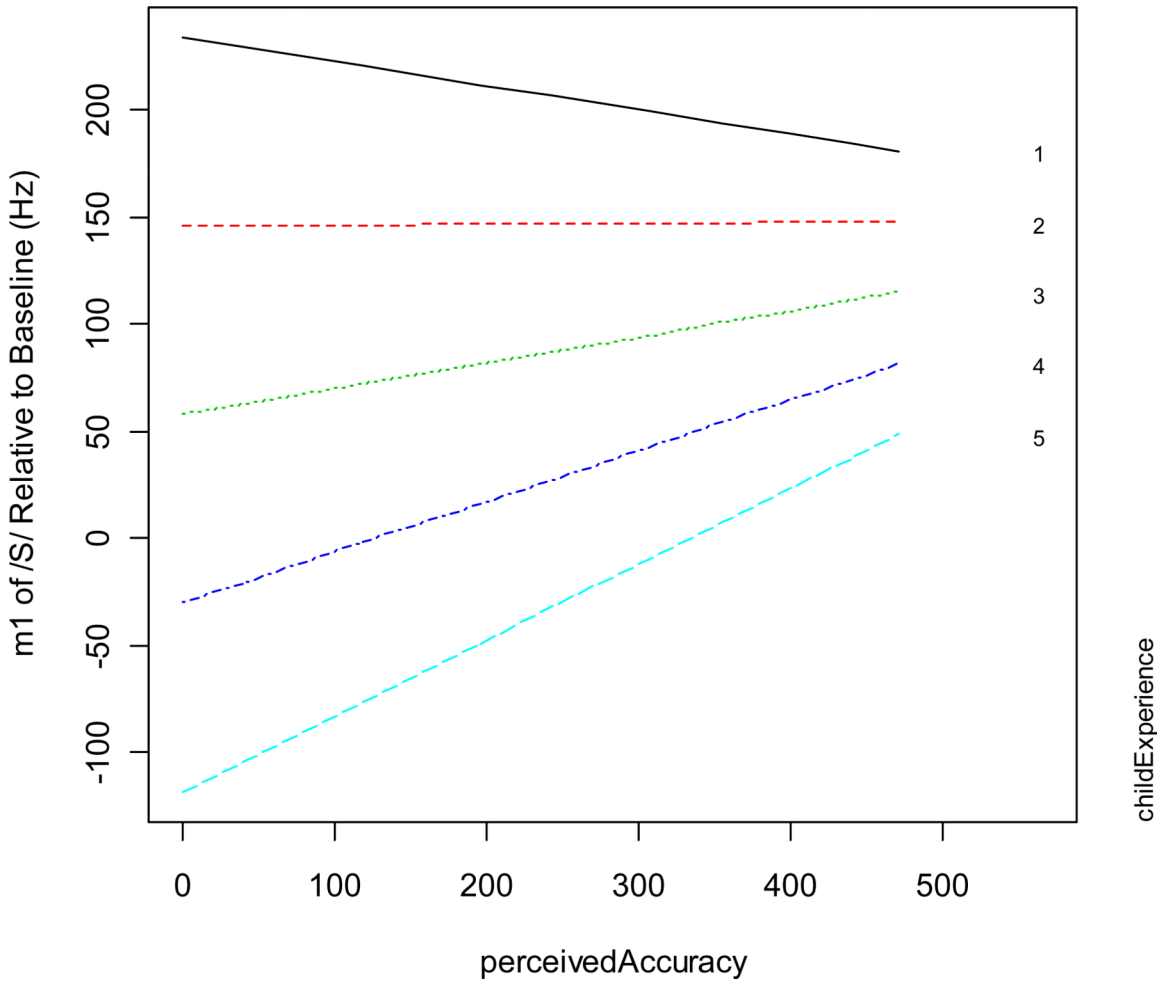

Figure 5.
Line-plot showing the mediating effect of self-reported experience perceiving children's speech (from 1 [less] to 5 [more]) on the relationship between perceived accuracy and the m1 of /ʃ/ in the LRS task. See text for details.

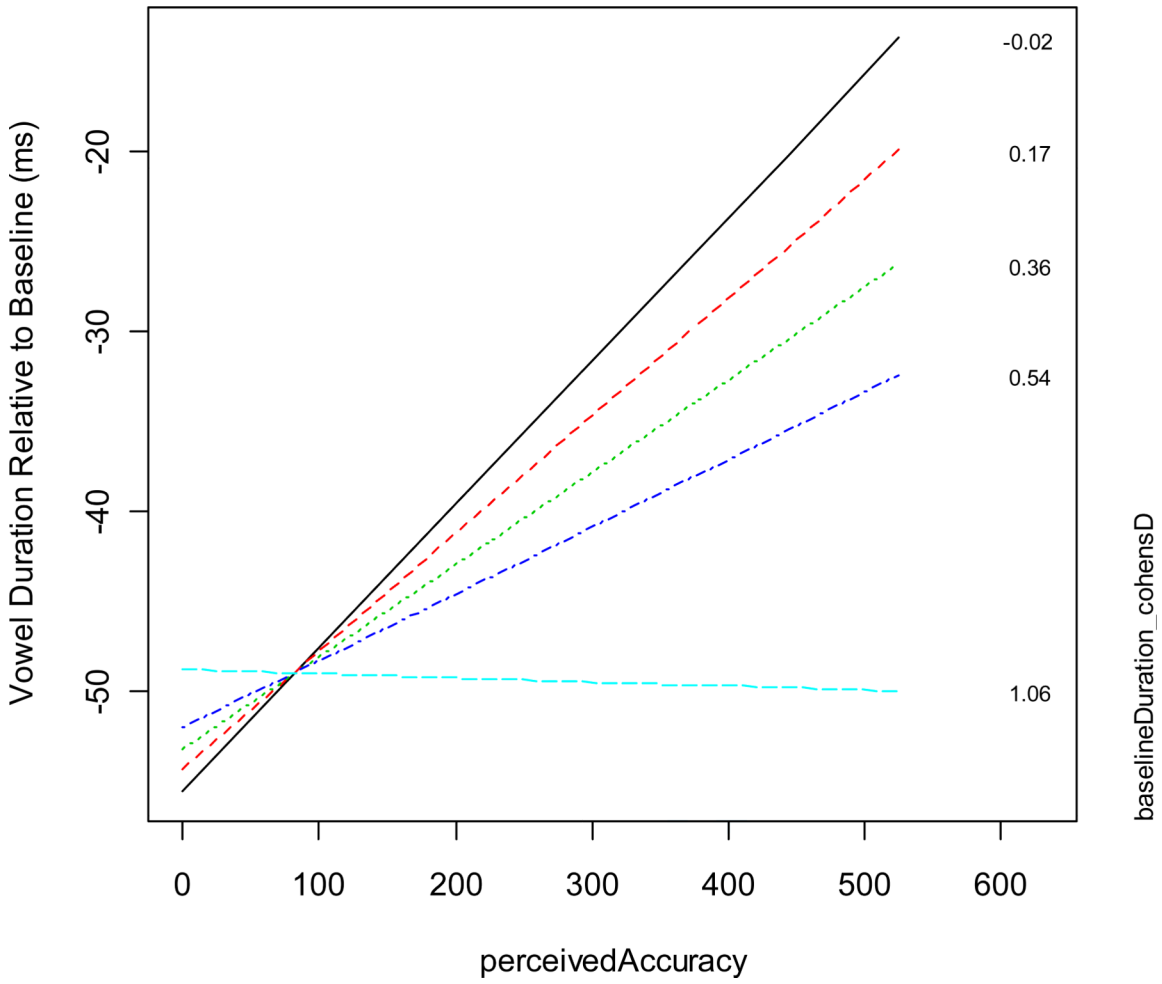

Figure 6.
Line-plot showing the mediating effect of the extent to which talkers modulated vowel duration in the baseline clear-speech task (larger values indicate greater modulation) on the relationship between perceived accuracy vowel duration in the LRS task. See text for details.