Running head: Perception-production links in children's speech

Perception-production links in children's speech

Joanna H. Lowenstein and Susan Nittrouer

Department of Speech, Language, and Hearing Sciences

University of Florida, Gainesville, FL

Conflict of Interest: The authors have no relevant conflicts of interest to report.

Funding: This work was supported by Grant No. R01 DC000633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

Correspondence should be addressed to: Joanna H. Lowenstein, Department of Speech, Language, and Hearing Sciences, PO Box 100174, The University of Florida, Gainesville, FL 32610.

Email: jlowenstein@phhp.ufl.edu

Tel: (352) 273-6170

Abstract

Purpose: Child phonologists have long been interested in how tightly speech input constrains the speech production capacities of young children, and the question acquires clinical significance when children with hearing loss are considered. Children with sensorineural hearing loss often show differences in the spectral and temporal structure of their speech production, compared to children with normal hearing. The current study was designed to investigate the extent to which this problem can be explained by signal degradation.

Method: Ten 5-year-olds with normal hearing were recorded imitating 120 three-syllable nonwords presented in unprocessed form, and as noise-vocoded signals. Target segments consisted of fricatives, stops, and vowels. Several measures were made: two duration measures (voice onset time and fricative length) and four spectral measures involving two segments (first and third moments of fricatives, and first and second formant frequencies for the point vowels).

Results: All spectral measures were affected by signal degradation, with vowel production showing the largest effects. Although a change in voice onset time was observed with vocoded signals for /d/, voicing category was not affected. Fricative duration remained constant.

Conclusion: Results support the hypothesis that quality of the input signal constrains the speech production capacities of young children. Consequently it can be concluded that the production problems of children with hearing loss – including those with cochlear implants – can be explained to some extent by the degradation in the signal they hear. However, experience with both speech perception and production likely plays a role, as well.

Introduction

A primary question addressed by speech research over the years has concerned the nature and extent of the relationship between speech perception and production during language acquisition. This is a tricky question to examine, precisely because perception and production develop in parallel, making it hard to determine if the development of one is driving development of the other. However, some data do exist to address the question.

Perception-production links in first speech

Evidence collected from the first year of life reveals that speech perception and production are related, even before children utter their first words. Particularly informative was a study by de Boysson-Bardies, Sagart, Halle, and Durand (1986), which examined the long-term average spectra of pre-word babble from 10-month-olds whose native languages were French, Cantonese, or Algerian. When these spectra were compared to those of adults in the infants' language communities, strong language specificity in the shapes of the spectra were observed, as well as strong similarities between spectra of infants and adults from the same language backgrounds. For example, long-term average spectra of French-speaking adults showed a definitive spectral peak below 500 Hz; infants long-term average spectra showed a definitive peak at slightly higher frequencies, reflecting their smaller vocal tracts. Thus, the structure of the speech children are hearing is shaping their earliest productions. However, these outcomes reflect the fact that long-term average spectra arise from general articulatory postures, such as degree of velo-pharyngeal closure or overall laryngeal height. Consequently, this finding that the broad spectral structure of infants' speech matches that of adults' speech cannot address the question of whether or not production affiliated with specific phonemic segments (i.e., tightly coordinated and rapidly occurring articulatory gestures) is affected by the structure of the speech infants and children hear.

In fact, when outcomes are examined across studies, some evidence actually seems to contradict the suggestion that there are strong and immediate perception-production links in the

speech processes of young children at the level of the phoneme. For example, one of the earliest phonetic contrasts that infants have been found to discriminate perceptually is that of voice-onset-time (VOT). This temporal structure is defined as the latency between release of a vocal-tract constriction and the onset of laryngeal vibration (Lisker & Abramson, 1964). This articulatory event has several acoustic consequences, one of which is a period of greatly reduced amplitude, or even silence. In English, this dip in amplitude is briefer for stops categorized as voiced than for those categorized as voiceless. No contrast is discriminated earlier in life than this one, with infants as young as two months of age demonstrating the ability to recognize the difference between voiced and voiceless stops (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Werker & Tees, 1999). Nonetheless, the development of the ability to produce voiced and voiceless stops with appropriate timing is protracted, with immature patterns being observed for children up to six years of age (e.g., Kewley-Port & Preston, 1974; Lowenstein & Nittrouer, 2008; Macken & Barton, 1980; Nittrouer, 1993; Zlatin & Koenigsknecht, 1976). The explanation attributed to this lengthy developmental period is that the coordination of vocal tract and laryngeal gestures is difficult to achieve. Thus it seems that the need for experience in producing speech might explain why tighter ties in age of acquisition for perceiving and producing specific phonemic segments have not always been found; but more evidence is needed to inform this question regarding timing.

One study explicitly investigated the relationship in time of acquisition for perception and production of specific phonemic contrasts. Edwards (1974) examined the abilities of 28 children between 1 year; 8 months and 3 years; 11 months to perceive and produce English stops, fricatives, and glides, and gleaned the order of acquisition from these data. Several trends were apparent, and differed from predictions. Overall it was found that perception of a specific minimal pair generally preceded production of a recognizable difference, but the timing of acquisition of perception and production was not tightly aligned. The relationship between perception and production varied across contrasts, suggesting that several factors account for

the age of acquisition of any specific phoneme, or phonemic contrast, in perception and production.

One factor that likely influences patterns of acquisition in speech perception and production is signal quality, which can be affected by both temporary and permanent hearing loss. McGowan, Nittrouer, and Chenausky (2008) analyzed speech samples from twenty 12-month-olds: ten with normal hearing (NH) and ten with hearing loss (HL). All of the children with HL had pure-tone thresholds poorer than 50 dB hearing level, and none of them had yet received cochlear implants. These authors examined syllable shape (i.e., numbers of syllables with consonantal constrictions on one side or the other), consonant types, and vowel formant frequencies. Results showed that the infants with HL produced fewer syllables with consonantal constrictions, fewer fricatives, and fewer stops with alveolar or velar places of closure. Lingual placement for vowels in the front-to-back dimension was less extreme. Thus it was concluded that children's earliest productions are influenced by what they can hear.

Studies with slightly older children (5 to 6 years of age) show a more nuanced pattern regarding speech production in children with HL, specifically those with cochlear implants. In general, these studies suggest that children with cochlear implants are able to produce voicing distinctions in stop consonants as well as their peers with normal hearing (Bunta, Goodin-Mayeda, Procter, & Hernandez, 2016), but fail to demonstrate appropriate spectral structure, specifically in voiceless sibilants (Li, Bunta, & Tomblin, 2017). These findings suggest that children with cochlear implants may have access to veridical temporal structure in the acoustic speech signal, but access only to degraded spectral structure. This discrepancy in quality of temporal and spectral structure for speech along with the production patterns of children with cochlear implants supports the hypothesis of a perception-production link in speech acquisition.

However, one constraint in experiments investigating whether the speech production difficulties of deaf infants and children can be attributed to degraded input is that these children have both degraded sensory input, as well as diminished experience hearing and producing speech. Children with HL start hearing speech to any meaningful extent only after receiving

amplification. The numbers of opportunities they have to hear speech are further constrained by the disproportionately greater influence of noise on speech recognition for listeners with HL. Finally, children with HL start talking later than children without HL, so the amount of experience in speech production is diminished at any given age. Consequently, the general finding that children with HL are poorer at speech production than their peers with NH is inconclusive evidence regarding potential perception-production links in speech acquisition, because their production delays cannot be attributed directly to their degraded inputs. Experimental methods are needed that more tightly link the input signal and the productions.

Perception-production links in adult speech

In addition to the work with young children examining perception-production links in speech processes, studies involving adults have revealed some relationship between perception and production. Many of these studies are able to do just what is suggested above: more tightly link the signal input and the speech production. For example, Perkell et al. (2004) examined perception and production of back vowel contrasts, and found that speakers who were more accurate at discriminating these close contrasts were also more likely to produce those contrasts with great specificity of lingual placement. A study of American English-speaking learners of French by Levy and Law (2010) similarly found that learners who were more accurate at categorizing tokens from the French /y-œ/ contrast (which is not present in English) also produced those vowels with formant frequencies matching those of native talkers more closely than did their peers who were not good at categorizing vowel tokens. In another study, adults heard the formant frequencies of their own vowel productions altered in real time (Houde and Jordan, 1998). For half of the eight participants, formant frequencies were shifted upwards; for the other half, formant frequencies were shifted downwards. After brief trainings the participants were found to adapt their productions so that they were matching intended targets with the altered signals. These studies demonstrate that there can be a strong and

immediate relationship between what is heard and what is produced, at least where vowels are concerned.

Listening and speaking with a cochlear implant

The primary motivation for the current study involved outcomes for children with cochlear implants (CIs). The signal processing of these devices provides only a degraded spectral representation to the auditory system, which is further degraded by the spread of excitation along the basilar membrane. Consequently, the frequency structure of speech is not well represented. Other aspects of acoustic structure, such as amplitude modulations over time and duration of syllables and segments, are not as deleteriously affected by the signal processing of CIs. Consequently, it could be predicted that features of speech perception more dependent on spectral structure, such as fricative and vowel identity, would be more affected than features dependent on temporal structure, such as VOT or segment durations.

A primary spectral property that has been examined in the speech production of children with CIs is the frequency content of the sibilant fricatives /s/ and /ʃ/. In numerous studies it has been observed that children with CIs produce less distinction between /s/ and /ʃ/ than do children with NH. For example, Uchanski and Geers (2003) found that English-speaking children with CIs (ages 8-9 years) tended to have lower spectral means for /s/, which were not well differentiated from /ʃ/. Similar results were found for children with CIs (ages 9-15 years) who were speakers of Croatian: they produced /s/ and /ʃ/ with overlapping frequency ranges, with /s/ produced more similarly to /ʃ/ (Liker, Mildner, & Šindija, 2007; Mildner & Liker, 2008). Todd, Edwards, and Litovsky (2011) investigated sibilant production in children with CIs (ages 4-9 years), taking care to include correct productions only. They found that the children with CIs produced /s/ and /ʃ/ with closer spectral peaks and more overlap than did children with NH, even for productions that were judged as correct. A follow-up study found that children with CIs produced less acoustic contrast between /s/ and /ʃ/ than children with NH, with /ʃ/-initial words being judged more intelligible than /s/-initial words (Reidy, Kristensen, Winn, Litovsky, &

Edwards, 2017). Studies analyzing the speech of children with CIs using narrow transcription also provide evidence for fricative production confusions, particularly /ʃ/ substituting for /s/ (Baudonck, Dhooge, D'haeseleer, & Van Lierde, 2010; Mahshie, Core, & Larsen, 2015).

Vowel production in children with CIs has also been examined acoustically. Liker et al. (2007) defined the vowel spaces of children with CIs and NH peers using the frequencies of the first and second formants (F1 and F2) in the "point" vowels, /i/, /a/, and /u/. These investigators found that the children with CIs had smaller vowel spaces than the children with NH, and their vowel spaces tended to be fronted. Another study by the same research group, which was longitudinal in design, showed that the vowel spaces of children with CIs tended to become less fronted over time (Mildner & Liker, 2008). That outcome is important because it suggests that speech production is not entirely constrained by input; even children with degraded input can learn to produce speech more accurately.

A study examining vowel production in children with CIs who spoke Persian reported that those children had much smaller and more centralized vowel spaces than the children with NH in the study (Jafari et al., 2016). Similarly, a study looking at vowel production in Mandarin-speaking children found reduced vowel spaces for children with CIs, compared to children with NH (Chuang, Yang, Chi, Weismer, & Wang, 2012). However, Baudonck, Van Lierde, Dhooge, and Corthals (2010) observed that Dutch-speaking children who used CIs had slightly larger vowel spaces than the NH children in their study. Kant, Patadia, Govale, Rangasayee, and Kirtane (2012) examined speech production in children with CIs (ages 5-11 years) who were native speakers of Hindi. They found that the vowel /e/ for these children tended to have lower F1 and F2 compared to those of age-matched peers with NH, suggesting raised and backed tongue placements. No effects for /i/ or /u/ were observed. Salas-Provance, Spencer, Nicholas, and Tobey (2014) found that young children (age 42 months) with CIs had more trouble producing central vowel targets, compared to NH peers. Similarly, Yang, Brown, Fox, and Xu (2015) reported that children with CIs learning Mandarin had more variable F1 and F2 frequencies for central vowels than children with NH, suggesting that a consequence of

degraded input could be greater variability in production. Overall, these studies show that children with CIs tend to have a reduced vowel space, compared to children with NH, and more variable productions for central vowels.

Fewer studies have examined temporal, or duration, properties of speech produced by children with CIs. Perhaps that paucity of studies arises because CIs degrade the spectral representation of speech more than the temporal structure. That suggestion is borne out by the finding that children with CIs demonstrate similarly shaped labeling functions for synthetic /d/-to-/t/ continua as do children with NH, and phoneme boundaries are in the same place (Caldwell & Nittrouer, 2013). As a result, it can be concluded that CIs preserve temporal structure, at least fairly well. Nonetheless, several investigators have reported errors in VOT production by children with CIs (e.g., Bharadwaj & Graves, 2008; Horga & Liker, 2006; Kant et al., 2012), but that may be due to a lack of experience in production. Comparable VOTs for children with CIs and children with NH were observed by Uchanski and Geers (2003), for children with CIs receiving spoken-language intervention. These children produced /t/ and /d/ with the same VOTs as children with NH, while children with CIs in sign-supported intervention programs exhibited shorter VOTs for /t/, with more variability than either the children with NH or those with CIs in spoken-language programs. Bunta et al. (2016) replicated the finding of similar VOT values for children with NH and those with CIs who use spoken language. Similarly to younger children with NH, the 8- to 9-year-olds in sign-supported programs in the Uchanski and Geers study exhibited difficulty and inconsistency in coordinating the vocal-tract opening and laryngeal abduction gestures. More generally, both Chuang et al. (2012) and Yang et al. (2015) reported that children with CIs had longer segment and syllable durations than their peers with NH. Yang et al. concluded that this trend arose specifically from difficulty on the part of the children with CIs in articulatory movements affiliated with transitioning from one vocal-tract constriction to another. Outcomes of developmental studies lend strong support for the suggestion that experience – in this case, especially with speech production – may explain these outcomes.

Current Study

This study examined potential links between perception and production in 5-year-old children with NH and typical language development by having them repeat three-syllable nonwords presented in both natural form, and after having been spectrally smeared using noise-vocoding techniques (e.g., Eisenberg, Shannon, Schaefer Martinez, Wygonski, & Boothroyd, 2000; Nittrouer, Lowenstein, & Packer, 2009; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). A total of 120 three-syllable nonwords were created, based loosely on the design of the CVCVCVC stimuli of Dollaghan and Campbell (1998). Children with NH listening to noise-vocoded models were used as speakers, rather than children with CIs, for reasons already stated: Any degradation in the speech of children with CIs is likely attributable to both the degraded signal as well as decreased experience. Furthermore, the nature and extent of signal degradation undoubtedly varies across children with CIs.

In the current study, two kinds of consonants were of interest. First, voiceless sibilant fricatives were included, and both temporal and spectral properties were measured. Second, voiced and voiceless stops were included, and the temporal property of voice-onset-time was measured. Tense vowels were included so that formant frequencies could readily be compared in the vocoded and unprocessed conditions; these vowels are preferable to lax vowels, which tend to have less stable formant patterns. Thus, segments were included that readily supported the analyses of temporal and spectral properties, as other investigators have done (Bunta et al., 2016; Li et al., 2017). Nonwords were used as stimuli, because the imitation of real words might be based more on stored representations than on immediate input.

Predictions regarding outcomes of the current study arose from both the quality of the signal input, and the experience of the children participating. Outcomes could shed light on the relationship between the nature of the signal for speech perception and speech production. First, the spectral structure of the speech produced by the children was predicted to be affected by vocoding. Specifically, it was predicted that /s/ spectra would be more similar to /ʃ/ spectra in frequency location and spectral shape, and that there would be a reduction in vowel space.

Those predictions follow from the simple fact that noise vocoding strongly degrades the spectral structure of speech. If the prediction were supported, it would reveal a strong and direct connection between the quality of the signal heard and the speech produced.

It was more difficult to make specific predictions for temporal aspects of the children's speech production. The temporal structure of the targets presented was not affected by the noise vocoding. Thus it was predicted that none of these children would show negative effects of the noise vocoding on temporal properties, but that was not for certain. Perhaps there is an effect of listening to degraded signals on speech production, not explained specifically by the acoustic structure of that signal. It may be that simply hearing a degraded speech signal renders all acoustic properties less salient, making imitation difficult.

Finally, variability among multiple imitations of the same spoken models was examined, to see if hearing degraded signals makes it more difficult to extract precise structure, which is needed for close imitation. If the degraded signal just provides an unreliable signal, then we could predict greater variability in speech production.

Method

Participants

Ten children participated in this study, ranging in age from 5 years; 0 months to 5 years; 8 months. The mean age was 5 years; 3 months. Four were boys, and six were girls. Five children were recorded in Columbus, Ohio and five in Gainesville, Florida. The children in this study all came from households where at least one parent had a bachelor's degree or higher. All children were native speakers of American English, and none of their parents reported any history of hearing or speech disorder in the children. All children passed hearing screenings consisting of the pure tones of 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB hearing level to each ear separately. Parents reported that their children were free from significant histories of otitis media, defined as six or more episodes during the first three years of life. Children were given the Goldman Fristoe 2 Test of Articulation (Goldman & Fristoe, 2000) and all scored

above the 30th percentile for their age. In particular, all children were able to produce without error the stop and fricative target sounds in the nonwords serving as stimuli in this study.

Equipment

Stimuli for imitation were recorded in a sound booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Laboratories Soundblaster soundcard using a 44.1-kHz sampling rate and 16-bit digitization. All testing took place in a soundproof booth, with the computer that controlled stimulus presentation in an adjacent room. Hearing was screened with a Welch Allyn TM262 audiometer using TDH-39 headphones. Stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. This system has a flat frequency response and low noise. Custom-written software controlled the presentation of the stimuli. Children were recorded via a Shure MX185 lavalier microphone and a Marantz PMD661 solid state audio recorder using a 44.1-kHz sampling rate and 24-bit digitization. Acoustic analysis of obtained samples was performed using TF32 software (Milenkovic, 2005).

Stimuli

Stimuli consisted of 120 three-syllable CVCVCVC nonwords. The first syllable consisted of one of the consonants /ɹ/, /l/, /m/, /n/, /w/, /k/, or /g/ followed by one of the vowels /ɪ/, /ɛ/, /æ/, or /ʌ/. The second and third syllables consisted of syllable-initial /s/, /ʃ/, /b/, /p/, /d/, or /t/ followed by vowels /i/, /e/, /ɑ/, /o/, or /u/. The final syllable ended with /b/, /p/, /g/, or /k/. The full set of stimuli is listed in the Appendix. With the exception of first-syllable /k/ and /g/, all acoustic analyses were performed on consonants and vowels in the second and third syllables. Each vowel in those syllables appeared 48 times in the list (24 times in each syllable), and each syllable-initial consonant appeared 40 times in the list (20 times in each syllable). First-syllable /k/ and /g/ each appeared 19 times in the list. Stimuli were recorded by a male native speaker of

English who is a trained phonetician. The words were pronounced with a natural stress pattern, with equal stress on the first two syllables and a slight drop in intensity and fundamental frequency in the third syllable. The waveform of the word /mʌdipɑp/ in unprocessed form at the top of Figure 1 illustrates this stress pattern.

To create the vocoded stimuli, the same MATLAB routine was used as in previous experiments (e.g., Nittrouer & Lowenstein, 2014; Nittrouer & Lowenstein, 2010; Nittrouer et al. 2009). All words were first band-pass filtered with a low-frequency cut-off of 50 Hz and a high-frequency cut-off of 8,000 Hz. The stimuli were vocoded using eight channels, a decision made based on earlier work by Friesen, Shannon, Başkent, & Wang (2001) demonstrating that CI users actually have about seven or eight perceptual channels, even though the devices have more physical channels. In that study, the best CI users – who were adults – performed similarly to adults with normal hearing listening to speech vocoded using eight channels. For stimuli in this current study, cutoff frequencies between channels were at 0.4, 0.8, 1.2, 1.8, 2.4, 3.0, and 4.5 kHz. Each channel was half-wave rectified using a 160-Hz high-frequency cutoff, and results used to modulate white noise limited by the same bandpass filters as those used to divide the speech signal into channels. Figure 1 shows the spectrograms of the word /mʌdipɑp/ in unprocessed form (middle) and after vocoding (bottom).

Procedures

All procedures were approved by the Institutional Review Boards of the Ohio State University and the University of Florida. After the parent signed the consent form and the child assented, the hearing screening and Goldman-Fristoe were administered.

Stimuli were presented under headphones at 68 dB sound pressure level. There were two presentation blocks in the experiment. In the first block, the participants heard half of the nonwords in unprocessed form and the other half in vocoded form. Stimuli presented in unprocessed or vocoded form were randomly assigned by the software for each child. In the second block of stimuli, each child heard each nonword in the alternate form. Stimuli were

presented in random order, with the rule that no more than two unprocessed or two vocoded stimuli could be presented in a row.

During testing, the child was seated across from the tester. The lavalier microphone was attached to a vest that the child wore. This vest kept the microphone at an appropriate distance from the mouth. Before testing started, the child was asked to repeat the vowels /i/, /ɑ/, and /u/ three times in random order, so that recording levels could be adjusted. Levels were set so that the child's speech productions fell between -12 dB and -3 dB recording level. This procedure kept recordings generally at an appropriate level, but if the level drifted up or down over the course of recording it was adjusted.

Children were told that they would hear a man or a robot say made-up words, and that they should repeat them. Children were presented with each nonword once, unless there was interference with that presentation, such as the child coughing. Only in those rare cases were stimuli replayed.

Children moved a game piece on a 10-space game board after presentation with every 12 nonwords to help keep track of where they were in the task. After the first block of stimuli were presented (120 nonwords), children were given a short break before the second block was started.

Measurements

Utterances were separated into their own files and then down-sampled to a 22.05-kHz sampling rate with 16-bit digitization. For this study, only word-initial stops /g/ and /k/, syllable-initial stops /b/, /p/, /d/, and /t/, syllable-initial fricatives /s/ and /ʃ/, and vowels /i/, /ɑ/ and /u/ were analyzed. Several measurements were made.

For word-initial stops /g/ and /k/ and syllable-initial stops /b/, /p/, /d/, and /t/, VOT was measured. To do this, one cursor was placed at the start of the broadband aperiodic burst in the wave form (correlate of the oral release) and the other was placed at the onset of a regular periodic signal in the wave form (correlate of voicing onset). Cursor placement was confirmed in

the spectrographic display. VOT was computed as the interval between the two cursors. For syllable-initial voiced stops, if voicing was continuous through the closure, VOT was designated as 0 ms. Voiceless stops should have longer VOTs than voiced stops, and stops produced in the back of the mouth, such as /g/ and /k/, should have longer VOTs than stops produced in the front of the mouth, such as /b/ and /p/ (Byrd, 1993; Cho & Ladefoged, 1999). The total number of stops measured per child was 40 each for /b/, /p/, /d/, and /t/, and 19 for /g/ and /k/.

For the fricatives /s/ and /ʃ/, spectral moments were measured over a 46-ms window centered at the temporal midpoint of the fricative noise. This window length was chosen because it was the longest analysis window available in the TF32 software. A cursor was placed at the onset of fricative noise, and a second cursor at the offset. Duration was measured between the two cursors. The temporal midpoint was calculated and cursors were placed at 23 ms before and 23 ms after that midpoint to establish the 46-ms window. The first and third spectral moments are presented in this report because they are the ones that best describe spectral shape. The first spectral moment (M1) describes the mean frequency of the noise, indicating general spectral weight. The third spectral moment (M3) describes skewness. Mean frequency is typically higher for /s/ than for /ʃ/, and spectra for /s/ tend to be more negatively skewed than those for /ʃ/. Thus, M1 tends to be higher (more positive), and M3 tends to be lower (more negative) for /s/ than for /ʃ/. Second spectral moments (M2), which describe variance, were not analyzed, because previous studies have found that they do not differentiate sibilant fricatives (Forrest, Weismer, Milenkovic, & Dougall, 1988; Jongman, Wayland, & Wong, 2000; Nittrouer, 1995). Fourth spectral moments (M4), which describe kurtosis, were analyzed, but were omitted from this report because they correlated so strongly with M3 that they did not provide any additional information about sibilant production. The total number of fricatives measured per child was 40 for /s/ and 40 for /ʃ/.

Finally, for vowels /i/, /ɑ/, and /u/, F1 and F2 were measured in Hz over three pitch periods in the most stable vowel region using 26-pole LPC analysis. F1 is lower for close, or high vowels (/i/ and /u/) than for open, or low vowels (/ɑ/). F2 should be highest for the front

vowel /i/ and lowest for the back vowel /u/.The total number of vowels measured per child was 48 each for /i/, /ɑ/, and /u/.

For all measurements, within-child standard deviations (within-child SDs) were also computed, in order to index variability in productions.

Measurements of samples from seven children were made by the first author and measurements of samples from the other three children were made by an independent consultant. Both individuals have extensive experience in acoustic analysis of children's speech. The speech samples were coded so that the consultant was blinded as to whether each sample was produced in response to vocoded or unprocessed speech. The first author independently measured 10% of the samples of the first child that the consultant measured, to ensure that samples were measured consistently.

Analyses

Data for each child were analyzed separately, in order to derive means for each measurement made, across tokens. In addition, the standard deviation (within-child SD) was obtained for each measurement. Analyses were then conducted on the derived means, with the focus on examining whether differences existed for productions that were imitations of the unprocessed versus vocoded stimuli. These analyses involved repeated-measures Analyses of Variance (ANOVAs), with planned contrasts for condition (unprocessed or vocoded).

Multiple analyses are reported here, so concern may be raised about increased risk of Type I error. Accordingly, a Bonferroni adjustment could be applied experiment-wide. Based on the number of analyses, this adjustment would suggest that the observed *p* would need to be equal to or less than .003 in order to meet the specified alpha level of .05. However, caution should be exercised in considering this adjustment, both because the Bonferroni adjustment is highly conservative, and the sample size was relatively low. These factors raise the risk of Type II error.

Results

The correlation coefficient between all acoustic measures made by the first author and the consultant was .999. The correlation coefficient specifically for duration measures was 1.000; for vowel formant frequencies it was .999; and for spectral moments it was .999. This level of agreement was judged to be extremely reliable.

A significance level of .05 was used, although precise *p* values are reported for $p < .10$; for $p > .10$, outcomes are reported simply as *not significant*. Values for *p* which meet the .003 level for multiple analyses are indicated with an asterisk.

Spectral moments

First, the spectral structure of the fricatives was examined, and compared across presentation conditions. Figure 2 shows mean M1s for /s/ and for /ʃ/, for each presentation condition separately. It is apparent that the children differentiated M1 for /s/ and /ʃ/ in both the unprocessed and vocoded conditions. A repeated-measures ANOVA was conducted on these M1 values, with condition and fricative place as the repeated measures, and planned contrasts for condition. The main effect of fricative place was significant, $F(1,9) = 19.82$, $p = .002$*, $\eta^2 = .69$. This confirms the observation that children were producing /s/ and /ʃ/ with different overall spectral weight. Regarding the effect of presentation condition, it appeared that the children produced /s/ with a slightly lower M1 and /ʃ/ with a slightly higher M1 in the vocoded compared to the unprocessed condition, but neither the condition main effect nor the Condition x Fricative Place interaction was significant. The planned contrast of M1 in the unprocessed versus the vocoded condition was not significant for /s/, with $p > .10$. Neither was this contrast significant for /ʃ/, although $F(1, 9) = -4.811$, $p = .056$, $\eta^2 = .35$.

Mean within-child SDs for M1 for each presentation condition separately are presented on the top line of Table 1. A repeated-measures ANOVA was conducted on these values, with condition and fricative place as the repeated measures, along with planned contrasts. No

significant effects were found. Thus, children were no more variable in their imitations of fricative place for the vocoded condition than for the unprocessed condition.

Figure 3 shows mean M3s for /s/ and for /ʃ/, for each presentation condition separately. When imitating both unprocessed and vocoded stimuli, the children differentiated /s/ and /ʃ/ in terms of skewness, with /s/ being more negatively skewed. A repeated-measures ANOVA was conducted on these values, with condition and fricative place as the repeated measures, and planned contrasts for condition. The main effect of fricative place was significant, $F(1,9) = 17.46$, $p = .002$*, $\eta^2 = .66$, confirming the observation that the children were producing /s/ and /ʃ/ with different degrees of skewness. The imitated productions of vocoded stimuli showed a similar pattern for M3 as they did for M1, with the difference between these values slightly reduced in the vocoded condition, compared to the unprocessed condition. But in the case of M3, this effect was almost entirely due to a shift in the /s/ M3 for the vocoded condition. Returning to the ANOVA outcomes, no main effect of condition was found, but the Condition x Fricative Place interaction was significant, $F(1, 9) = 6.03$, $p = .036$, $\eta^2 = .40$. Planned contrasts for condition were significant for /s/, $F(1,9) = 7.47$, $p = .023$, $\eta^2 = .45$. These findings provide some evidence that the spectral smearing of the vocoded signals resulted in children producing less-differentiated /s/ and /ʃ/ tokens.

Variability was also examined for M3. Within-child SDs for each presentation condition separately are presented on the bottom line of Table 1. A repeated-measures ANOVA was conducted on within-child SDs for M3, with condition and fricative place as the repeated measures, and planned contrasts. Fricative place was significant, $F(1,9) = 10.94$, $p = .009$, $\eta^2 = .55$, reflecting that for these children, productions of /s/ were more variable in M3 than productions of /ʃ/. No other significant effects were observed.

Vowel Formant Frequencies

Figure 4 shows vowel areas across participants, specifically plotting average F1 and F2 for /i/, /ɑ/, and /u/ for imitations of unprocessed stimuli (solid black lines) and imitations of

vocoded stimuli (dashed red lines). The black square represents the centroid (geometric middle) for the unprocessed condition and the red circle represents the centroid for the vocoded condition. There appears to be a trend towards centralizing vowels when repeating the vocoded stimuli. High vowels were generally lowered (reflecting changes in F1 for /i/ and /u/), and back vowels were fronted (reflecting changes in F2 for /ɑ/ and /u/). The front vowel /i/ was also backed (reflecting changes in F2).

To examine these results, repeated-measures ANOVAs were conducted on F1 and F2 means separately, with condition and vowel place as the repeated measures, and planned contrasts for condition. Results for F1 are presented in Table 2. The main effect of vowel place was significant, which confirms the observation that the children differentiated these vowels in terms of F1. The main effect of condition failed to reach significance but the Condition x Vowel Place interaction was significant. This likely resulted from larger changes in F1 for /i/ and /u/ and only small changes in /ɑ/. F1 for /i/ was higher in the vocoded condition, compared to the unprocessed condition, and the planned contrast was significant. F1 for /u/ appeared to be higher in the vocoded than the unprocessed condition, but the planned contrast failed to reach significance. Nonetheless, the significant interaction and significant planned contrast for /i/ F1 provide broad evidence of jaw lowering when vocoded stimuli were heard, relative to when unprocessed stimuli were heard.

The ANOVA results for F2 are presented in Table 3. The main effect of vowel place was again highly significant, which indicates that children differentiated these vowels in terms of F2. The main effect of condition was significant, as was the Condition x Vowel Place interaction. These results reflect the changes in F2 across the three vowels based on condition; the back vowels /ɑ/ and /u/ were more fronted in the vocoded than in the unprocessed condition, while the front vowel /i/ was backed, demonstrating a tendency towards vowel centralization. Planned contrasts were significant for all three vowels.

Table 4 presents mean within-child SDs for F1 and F2 for each vowel and presentation condition separately. For F1, it appears that variability was greatest for /ɑ/, followed by /u/, and

finally by /i/. For F2, the order appears to be that variability was greatest for /u/, followed by /ɑ/, and finally by /i/. The within-child SDs appear to be larger for the vocoded than for the unprocessed condition, for both formants in all vowels. To examine these apparent effects, separate repeated-measures ANOVAs for F1 and F2 within-child SDs were conducted with condition and vowel place as the repeated measures, and planned contrasts for condition. Table 5 shows results of the ANOVA for F1 within-child SDs. The main effects of condition and vowel place were significant, but the Condition x Vowel Place interaction was not significant. The planned contrast reached significance for all three vowels, with /ɑ/ showing the largest effect size. Table 6 shows results of the ANOVA for F2 within-child SDs. Again, the main effects of condition and vowel place were significant, as well as the Condition x Vowel Place interaction. This interaction reflects the fact that the back vowels /u/ and /ɑ/ were more variable in the vocoded condition compared to the unprocessed condition, as can be seen in the planned contrast results.

Table 7 shows vowel areas for each child, calculated as the geometric area of the triangle defined by the /i/, /ɑ/, and /u/ F1 and F2 points (Bradlow, Torretta, & Pisoni, 1996; Yang et al., 2015). The reduction in vowel area for individual children ranged from 6% to 64%, with an average reduction of 33%. A *t* test comparing mean vowel area (bottom line of Table 7) between the unprocessed condition and the vocoded condition was highly significant, $t(9) = 6.405$, $p < .001$*, Cohen's $d = 1.08$.

Temporal measures

Fricative duration and VOT were the two temporal measures examined in this study. Figure 5 presents mean duration measures for the fricative tokens, for each presentation condition separately. The fricatives produced when imitating vocoded stimuli appear to be slightly shorter than those produced when imitating unprocessed stimuli, but variability was high. A repeated-measures ANOVA with condition and fricative place as the repeated measures and

planned contrasts for condition resulted in no significant findings. Thus fricative duration did not differ across presentation conditions.

VOT measures are shown on Figure 6. As expected, VOTs for voiceless stops were longer than VOTs for voiced stops. In addition, VOT differed across stop place, with labial stops /p/ and /b/ having the shortest VOTs and velar stops /g/ and /k/ having the longest VOTs. It appears that there was an overall tendency to produce voiceless stops with slightly shorter VOTs when repeating vocoded stimuli, particularly for /k/, and to produce voiced stops with slightly longer VOTs when repeating vocoded stimuli, particularly for /d/. A repeated-measures ANOVA with condition, voicing category, and stop place as the repeated measures and planned contrasts for condition was conducted, and results are presented in Table 8. The main effect of voicing category was highly significant, confirming the observation that children produced voiceless stops with longer VOTs than voiced stops. The main effect of stop place was also significant, confirming the observation that the children produced stops with different VOTs based on place of constriction. The main effect of condition, however, failed to reach significance. The two-way interactions of Condition x Place, Condition x Voicing, and Place x Voicing all failed to reach statistical significance, as well, and the three-way interaction of Condition x Place x Voicing was not significant. Out of the six planned contrasts, the only one that was significant was for /d/, $F(1,9) = 12.81$, $p = .006$, $\eta^2 = .59$. However, this change did not result in a shift of voicing category.

Variability was also examined for VOT. Mean within-child SDs are presented in Table 9, for each stop and presentation condition separately. It appears that voiceless stops were produced with more variability in VOT than voiced stops, and VOT productions were slightly more variable in the vocoded condition. A repeated-measures ANOVA with condition, voicing category, and stop place as the repeated measures and planned contrasts for condition was conducted to further examine these observations. The main effect of voicing was significant, $F(1,9) = 61.87$, $p < .001$*, $\eta^2 = .87$, confirming that voiceless stops were produced with more variable VOTs than voiced stops. The main effect of condition was also significant, $F(1,9) =$

6.52, p = .031, η^2 =.42, confirming that stops produced when repeating vocoded stimuli had more variable VOTs than stops produced when repeating unprocessed stimuli. The main effect of stop place, all of the two-way interactions, and the three-way interaction failed to reach significance. The only significant planned contrast of the six conducted was again for /d/, $F(1,9)$ = 13.03, p = .006, η^2 =.59.

Discussion

The current study was undertaken to examine the influence of having a degraded spectral input on the immediate speech production of young children. The motivation for this investigation stemmed largely from reports on the speech production of children with CIs. Earlier evidence has robustly shown that these children with CIs have speech production patterns that are less precise than those of their peers with NH, especially in terms of spectral properties. The question has long existed regarding whether those deficits in production are due primarily to the poor signal quality children with CIs receive, or due to their diminished experience in either perception or production of speech.

The first prediction of this study was that children with NH listening to vocoded speech would demonstrate effects of degraded spectral inputs by producing /s/ more similarly to /ʃ/. That prediction was met, to some extent. The children showed some changes in production of fricatives when imitating vocoded speech. In particular, M3 for /s/ became more /ʃ/-like, indicating a difference in production for /s/ that led to a less-skewed spectrum. This is similar to the findings for analyses of speech produced by children with CIs which have demonstrated that they produce /s/ with lower spectral means (Liker et al., 2007; Mildner & Liker, 2008; Uchanski & Geers, 2003). Though the changes in fricative production seen here were small, they do provide some evidence that spectral smearing resulted in changes in production for these children.

The second prediction in this study was that children would demonstrate changes in vowel formant frequencies when imitating vocoded speech. This prediction was supported robustly. In particular, all children in the study showed reduced vowel spaces, although one

child showed only a minor vowel-space reduction of 6%. The other nine children's vowel spaces were reduced between 24% and 64%. This reduction in vowel space was the result of the high-vowel /i/ being lowered (F1 increased for /i/), back vowels being fronted (F2 increased for /u/ and /ɑ/), and the front vowel being backed (F2 decreased for /i/). These results parallel those of children with CIs, who have reduced vowel spaces compared to their peers with NH (Chuang et al., 2012; Jafari et al., 2016; Liker et al., 2007; Mildner & Liker, 2008).

The third prediction of this study was that temporal measures of speech production (fricative duration and VOT) would not be influenced by the quality of the input signal. A review of the data reveals that this prediction was clearly supported for fricative duration: these children produced fricatives with consistent durations, whether they were imitating unprocessed or vocoded speech. However, there was a small, but significant shift in VOT for /d/. Nonetheless, this change in VOT was not large, and all values for /d/ remained solidly within range for voiced stops. It could be that the degraded signal introduced enough processing demands that this small shift in VOT was the consequence.

Finally, the idea was considered that all acoustic measurements made in this study – both spectral and temporal – might be more variable across productions of individual children in the vocoded condition, if degraded speech simply provides a less reliable input signal. Evidence was found to support this suggestion, but only for acoustic properties that showed effects of signal degradation. Specifically, increased variability was observed for vowel formant frequencies in the vocoded condition, compared to the unprocessed condition. In particular, those measures showing the most change from unprocessed to vocoded condition also showed the greatest increases in within-child SDs. That means that F1 for the high vowels showed increased variability in the vocoded condition, compared to the unprocessed condition; /i/ was lowered significantly when vocoded speech served as the input signal. In addition, F2 for back vowels showed increased variability in the vocoded condition, compared to the unprocessed condition; these vowels were fronted when the children were listening to vocoded speech.

Limitations and Future Directions

The primary limitation of the current study was that data were examined for only ten children. However, the large number of samples collected from each child mitigates concern that these outcomes may not be representative of perception-production links in the broader population. In general, the same effects of signal degradation were observed across children. In particular, the robust finding of rather large vowel space reductions for nine of the ten children indicates that similar results would likely be obtained with a larger sample size.

Regarding future directions for this work, studies using children with NH who are presented with more varieties of signal degradation should be conducted. Although noise vocoding, as done here, can simulate the spectral smearing of CIs, vocoding alone does not simulate the frequency shifting inherent in CI stimulation or the holes in the spectrum that can arise due to damage in spiral ganglion cells (Culling, Jelfs, Talbert, Grange, & Backhouse, 2012; Dunn, Tyler, Oakley, Gantz, & Noble, 2008). Subsequently, speech production of children with CIs could be compared to that of children with NH listening to these degraded signals, to assess the extent to which the problems of children with CIs are directly attributable to the signal degradation they endure, as opposed to diminished experience. Regarding children with NH, the speech production of children with speech and language disorders should be examined, as part of a plan to examine whether their problems might be due to perceptual anomalies.

Conclusion

The current study was undertaken to investigate the strength and the nature of the perception-production link in children's speech acquisition. Specifically, the question was asked if the links that have been described in the literature could reasonably be attributed primarily to the nature of the signal children are hearing, or if other factors are at work. The speech production capacities of children with HL who receive CIs served as the bases for predictions in this study. Results showed that when the young children in this study were asked to imitate speech models spectrally degraded by noise vocoding, there were some changes in the

spectral structure of the speech they produced. There was even one small change observed for one temporal property, VOT. However, the deficits in speech production observed for these children with NH and typical language experience were not as extreme as what has been reported for children with CIs. Consequently the conclusion can be reached that some of the problems observed in the speech of children with CIs may result from impoverished experience. Thus, enhanced intervention consisting of practice producing speech should help to ameliorate the production problems of children with CIs.

Acknowledgements

The authors thank Richard McGowan for making acoustic measurements, Robert Fox for producing the stimuli, Eric Tarr for help with programming, and Ellen Hambley, Demarcus Williams, Lauren Linker, and Kierstyn Tietgens for their help with sound file processing.

References

Baudonck, N., Dhooge, I., D'haeseleer, E., & Van Lierde, K. (2010). A comparison of the consonant production between Dutch children using cochlear implants and children using hearing aids. *International Journal of Pediatric Otorhinolaryngology, 74*, 416-421.

Baudonck, N., Van Lierde, K., Dhooge, I., & Corthals, P. (2011). A comparison of vowel productions in prelingually deaf children using cochlear implants, severe hearing-impaired children using conventional hearing aids and normal-hearing children. *Folia Phoniatrica et Logopaedica, 63*, 154-160.

Bharadwaj, S. V., & Graves, A. G. (2008). Efficacy of the discreteness of voicing category (DOVC) measure for characterizing voicing errors in children with cochlear implants: a report. *Journal of Speech Language and Hearing Research, 51*, 629-635.

de Boysson-Bardies, B., Sagart, L., Halle, P., & Durand, C. (1986). Acoustic investigations of cross-linguistic variability in babbling. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 113-126). Wenner-Gren Int. Symp. Series 44. New York: Stockton Press.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

Bunta, F., Goodin-Mayeda, C. E., Procter, A., & Hernandez, A. (2016). Initial stop voicing in bilingual children with cochlear implants and their typically developing peers with normal hearing. *Journal of Speech, Language, and Hearing Research, 59,* 686-698.

Byrd, D. (1993). 54,000 American Stops. *UCLA Working Papers in Phonetics, 83,* 97-116.

Caldwell, A., & Nittrouer, S. (2013). Speech perception in noise by children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 56*, 13-30.

Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*, 207-229.

Chuang, H. F., Yang, C. C., Chi, L. Y., Weismer, G., & Wang, Y. T. (2012). Speech intelligibility, speaking rate, and vowel formant characteristics in Mandarin-speaking children with cochlear implant. *International Journal of Speech-Language Pathology, 14*, 119-129.

Culling, J.F., Jelfs, S., Talbert, A., Grange, J.A. & Backhouse, S.S. (2012). The benefit of bilateral vs. unilateral cochlear implantation to speech intelligibility in noise . *Ear & Hearing, 33*, 673-682 .

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1136-1146.

Dunn, C.C., Tyler, R.S., Oakley, S., Gantz, B.J., & Noble, W. (2008). Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation . *Ear & Hearing, 29,* 352-359.

Edwards, M. L. (1974). Perception and production in child phonology: The testing of four hypotheses. *Journal of Child Language, 1,* 205-219.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America, 107*, 2704-2710.

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

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110*, 1150-1163.

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe 2: Test of Articulation*. Circle Pines, MN: American Guidance Service, Inc.

Horga, D., & Liker, M. (2006). Voice and pronunciation of cochlear implant speakers. *Clinical Linguistics & Phonetics, 20*, 211-217.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*, 1213-1216.

Jafari, N., Drinnan, M., Mohamadi, R., Yadegari, F., Nourbakhsh, M., & Torabinezhad, F. (2016). A comparison of Persian vowel production in hearing-impaired children using a cochlear implant and normal-hearing children. *Journal of Voice, 30,* 340-344.

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

Kant, A. R., Patadia, R., Govale, P., Rangasayee, R., & Kirtane, M. (2012). Acoustic analysis of speech of cochlear implantees and its implications. *Clinical and Experimental Otorhinolaryngology, 5 Suppl 1*, S14-S18.

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

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America, 128*, 1290-1305.

Li, F., Bunta, F., & Tomblin, J. B. (2017). Alveolar and postalveolar voiceless fricative and affricate productions of Spanish-English bilingual children with cochlear implants. *Journal of Speech Language and Hearing Research, 60*, 2427-2441.

Liker, M., Mildner, V., & Šindija, B. (2007). Acoustic analysis of the speech of children with cochlear implants: a longitudinal study. *Clinical Linguistics & Phonetics, 21*, 1-11.

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

Lowenstein, J. H., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. *Journal of the Acoustical Society of America, 124*, 1180-1191.

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

Mahshie, J., Core, C., & Larsen, M. D. (2015). Auditory perception and production of speech feature contrasts by pediatric implant users. *Ear and Hearing, 36*, 653-663.

McGowan, R. S., Nittrouer, S., & Chenausky, K. (2008). Speech production in 12-month-old children with and without hearing loss. *Journal of Speech, Language, and Hearing Research, 51*, 879-888.

Mildner, V., & Liker, M. (2008). Fricatives, affricates, and vowels in Croatian children with cochlear implants. *Clinical Linguistics & Phonetics, 22*, 845-856.

Milenkovic, P. (2005). TF32 [Computer program]. Madison, WI: University of Wisconsin-Madison. Retrieved September 6, 2012. Available from http://userpages.chorus.net/cspeech/.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research, 36*, 959-972.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America, 97*, 520-530.

Nittrouer, S., & Lowenstein, J. H. (2010). Learning to perceptually organize speech signals in native fashion. *Journal of the Acoustical Society of America, 127*, 1624-1635.

Nittrouer, S., & Lowenstein, J. H. (2014). Separating the effects of acoustic and phonetic factors in linguistic processing with impoverished signals by adults and children. *Applied Psycholinguistics, 35*, 333-370.

Nittrouer, S., Lowenstein, J. H., & Packer, R. (2009). Children discover the spectral skeletons in their native language before the amplitude envelopes. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1245-1253.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M. et al. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America, 116*, 2338-2344.

Reidy, P. F., Kristensen, K., Winn, M. B., Litovsky, R. Y., & Edwards, J. R. (2017). The acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral cochlear implants. *Ear and Hearing, 38,* 42-56.

Salas-Provance, M. B., Spencer, L., Nicholas, J. G., & Tobey, E. (2014). Emergence of speech sounds between 7 and 24 months of cochlear implant use. *Cochlear Implants International, 15*, 222-229.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.

Todd, A. E., Edwards, J. R., & Litovsky, R. Y. (2011). Production of contrast between sibilant fricatives by children with cochlear implants. *Journal of the Acoustical Society of America, 130*, 3969-3979.

Uchanski, R. M., & Geers, A. E. (2003). Acoustic characteristics of the speech of young cochlear implant users: a comparison with normal-hearing age-mates. *Ear and Hearing, 24*, 90S-105S.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50*, 509-535.

Yang, J., Brown, E., Fox, R. A., & Xu, L. (2015). Acoustic properties of vowel production in prelingually deafened Mandarin-speaking children with cochlear implants. *Journal of the Acoustical Society of America, 138*, 2791-2799.

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research, 19*, 93-111.

Supplemental File

The Appendix presents the full set of 120 non-words used in the experiment.

Figure Captions

FIGURE 1. Spectrogram of /mʌdipɑp/ in unprocessed form (top) and after vocoding (bottom).

FIGURE 2. Mean first spectral moments for /s/ and for /ʃ/, for each presentation condition separately. Error bars are standard errors of the mean.

FIGURE 3. Mean third spectral moments for /s/ and for /ʃ/, for each presentation condition separately. Error bars are standard errors of the mean.

FIGURE 4. Mean F1 and F2 of the vowels /i/, /u/, and /ɑ/. Solid black lines represent speech produced as imitations to the unprocessed condition and dashed red lines represent speech produced as imitations to the vocoded condition. The black square represents the centroid of the unprocessed condition and the red circle represents the centroid of the vocoded condition. Error bars are standard errors of the mean.

FIGURE 5. Mean duration for /s/ and for /ʃ/, for each presentation condition separately. Error bars are standard errors of the mean..

FIGURE 6. Mean VOT (voice onset time) for voiceless and for voiced stops, for each presentation condition separately. Error bars are standard errors of the mean.

Table 1. Mean within-child SDs for first moments (M1, in kHz) and third moments (M3) for each condition separately. UP: Unprocessed. VC: Vocoded. Standard deviations are in parentheses.

	/s/		/ʃ/	
	UP	**VC**	**UP**	**VC**
M1	1.14 (0.30)	1.15 (0.34)	1.06 (0.32)	1.10 (0.28)
M3	0.88 (0.21)	0.87 (0.22)	0.66 (0.22)	0.73 (0.22)

Table 2. Outcomes of a two-way, repeated-measures ANOVA performed on F1 means, across processing condition and vowel place, with planned contrasts for condition. NS = not significant.

	F	*df*	*p*	η^2
Main Effects				
Condition	3.84	1,9	.082	.30
Vowel Place	186.36	2,18	<.001*	.95
Two-way Interaction				
Condition x Vowel Place	6.24	2,18	.009	.41
Planned Contrasts				
/ɑ/	NS	NS	NS	
/i/	16.65	1,9	.003*	.65
/u/	3.73	1,9	.086	.29

*Significant with Bonferroni adjustment for multiple analyses

Table 3. Outcomes of a two-way, repeated-measures ANOVA performed on F2 means, across processing condition and vowel place, with planned contrasts for condition.

	F	*df*	*p*	η^2
Main Effects				
Condition	16.79	1,9	.003*	365
Vowel Place	240.26	2,18	<.001*	.96
Two-way Interaction				
Condition x Vowel Place	25.35	2,18	<.001*	.82
Planned Contrasts				
/ɑ/	19.86	1,9	.002*	.69
/i/	42.14	1,9	<.001*	.82
/u/	21.45	1,9	.001*	.70

*Significant with Bonferroni adjustment for multiple analyses

Table 4. Mean within-child SDs (in Hz) for F1 and F2 for each vowel and presentation condition separately. Standard deviations are in parentheses.

	F1		**F2**	
	UP	**VOC**	**UP**	**VOC**
/ɑ/	160.7 (57.8)	187.9 (63.8)	286.0 (72.9)	406.2 (109.6)
/i/	53.8 (22.2)	84.6 (40.0)	254.0 (95.9)	336.8 (103.0)
/u/	74.8 (18.6)	123.0 (53.5)	390.1 (94.7)	662.9 (162.1)

Table 5. Outcomes of a two-way, repeated-measures ANOVA performed on mean within-child SDs for F1, across processing condition and vowel place, with planned contrasts for condition. NS = not significant.

	F	*df*	*p*	η^2
Main Effects				
Condition	8.93	1,9	.015	.50
Vowel Place	25.09	2,18	<.001*	.74
Two-way Interaction				
Condition x Vowel Place	NS	NS	NS	
Planned Contrasts				
/ɑ/	8.90	1,9	.015	.50
/i/	5.16	1,9	.049	.36
/u/	6.12	1,9	.035	.41

*Significant with Bonferroni adjustment for multiple analyses

Table 6. Outcomes of a two-way, repeated-measures ANOVA performed on mean within-child SDs for F2, across processing condition and vowel place, with planned contrasts for condition.

	F	*df*	*p*	η^2
Main Effects				
Condition	24.90	1,9	.001*	.74
Vowel Place	20.76	2,18	<.001*	.70
Two-way Interaction				
Condition x Vowel Place	9.25	2,18	.002*	.51
Planned Contrasts				
/ɑ/	13.13	1,9	.006	.59
/i/	4.683	1,9	.059	.34
/u/	27.99	1,9	.001*	.76

*Significant with Bonferroni adjustment for multiple analyses

Table 7. Vowel space area calculations (in Hz ^{2}) across both syllables, for each presentation condition separately. UP: Unprocessed. VC: Vocoded. The percentage of change in vowel area is in the fourth column. Means and standard deviations at the bottom are for the ten individual means shown.

	UP	**VC**	**Percent reduction**
B1	507,557	475,926	6%
G1	451,378	330,694	27%
B2	302,914	204,312	33%
G2	252,058	125,764	50%
B3	531,271	349,845	34%
G3	357,218	270,696	24%
B4	308,756	211,367	32%
G4	429,574	269,423	37%
G5	195,722	130,323	33%
G6	370,143	133,995	64%
Mean	370,659	250,189	33%
SD	109,440	113,075	

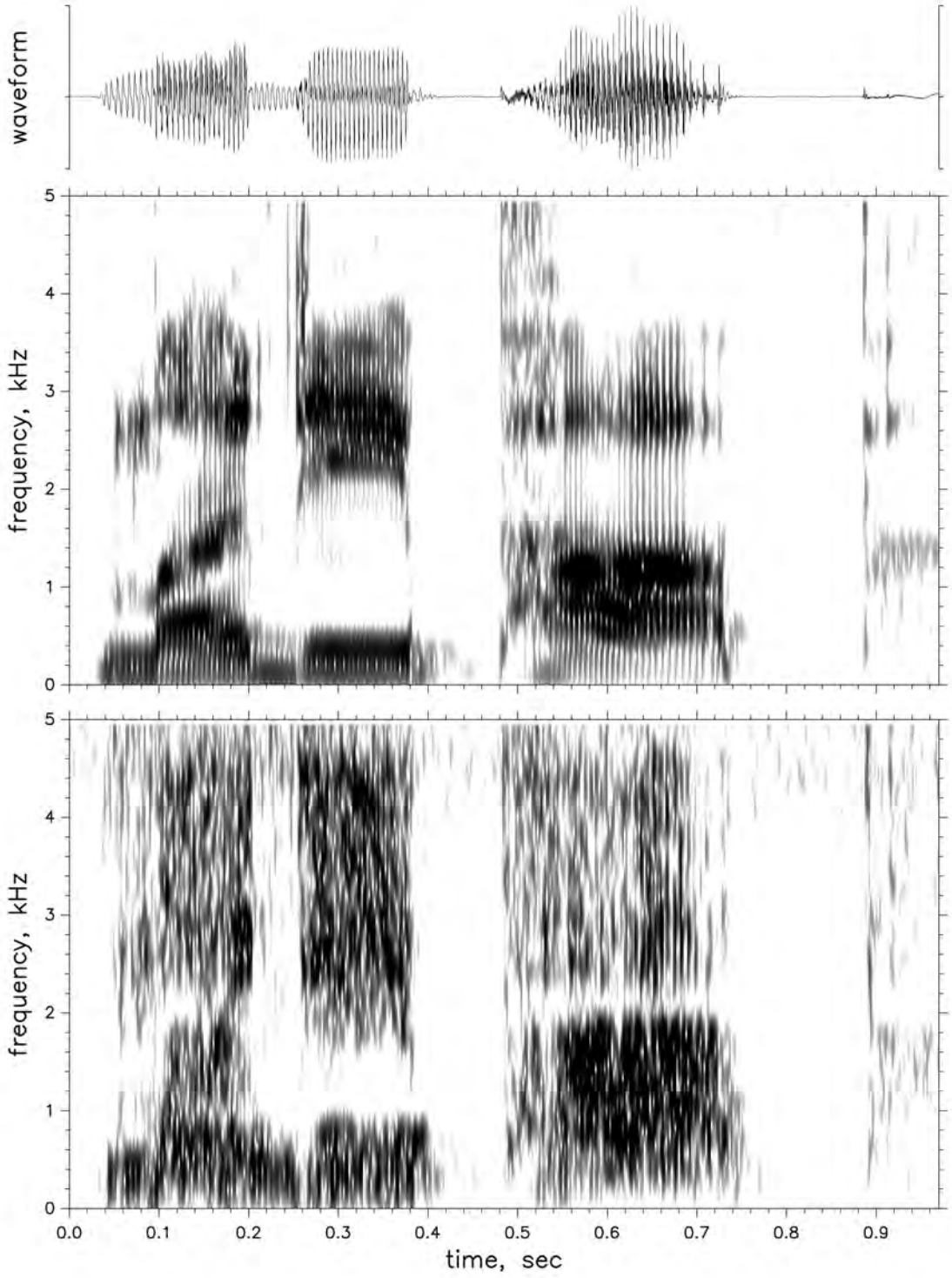
Table 8. Outcomes of a three-way, repeated-measures ANOVA performed on VOT means, across processing condition, voicing category and stop place. NS = not significant.

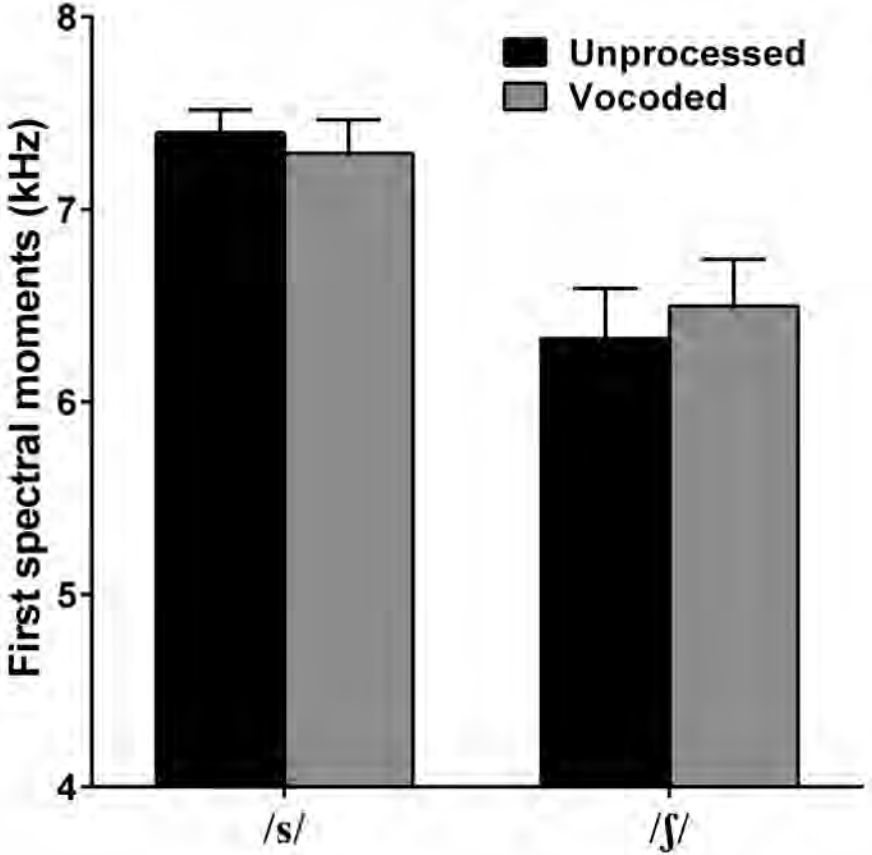
	F	*df*	*p*	η^2
Main Effects				
Condition	NS	NS	NS	
Place	29.42	2,18	<.001*	.77
Voicing	278.51	1,9	<.001*	.97
Two-way Interactions				
Condition x Place	3.09	2,18	.070	.26
Condition x Voicing	4.22	1,9	.070	.32
Place x Voicing	3.33	2,18	.059	.27
Three-way Interaction				
Condition x Place x Voicing	NS	NS	NS	

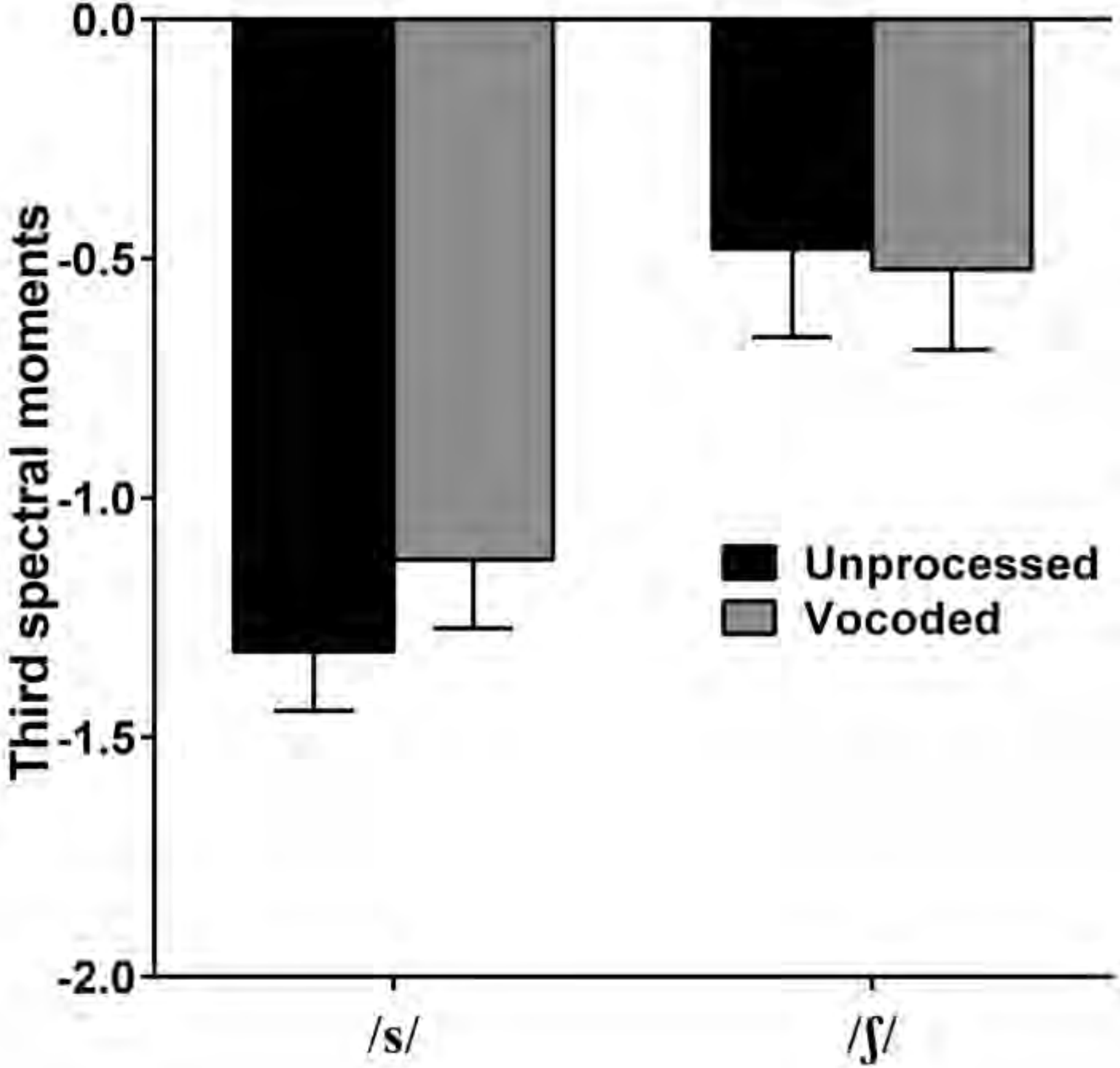
*Significant with Bonferroni adjustment for multiple analyses

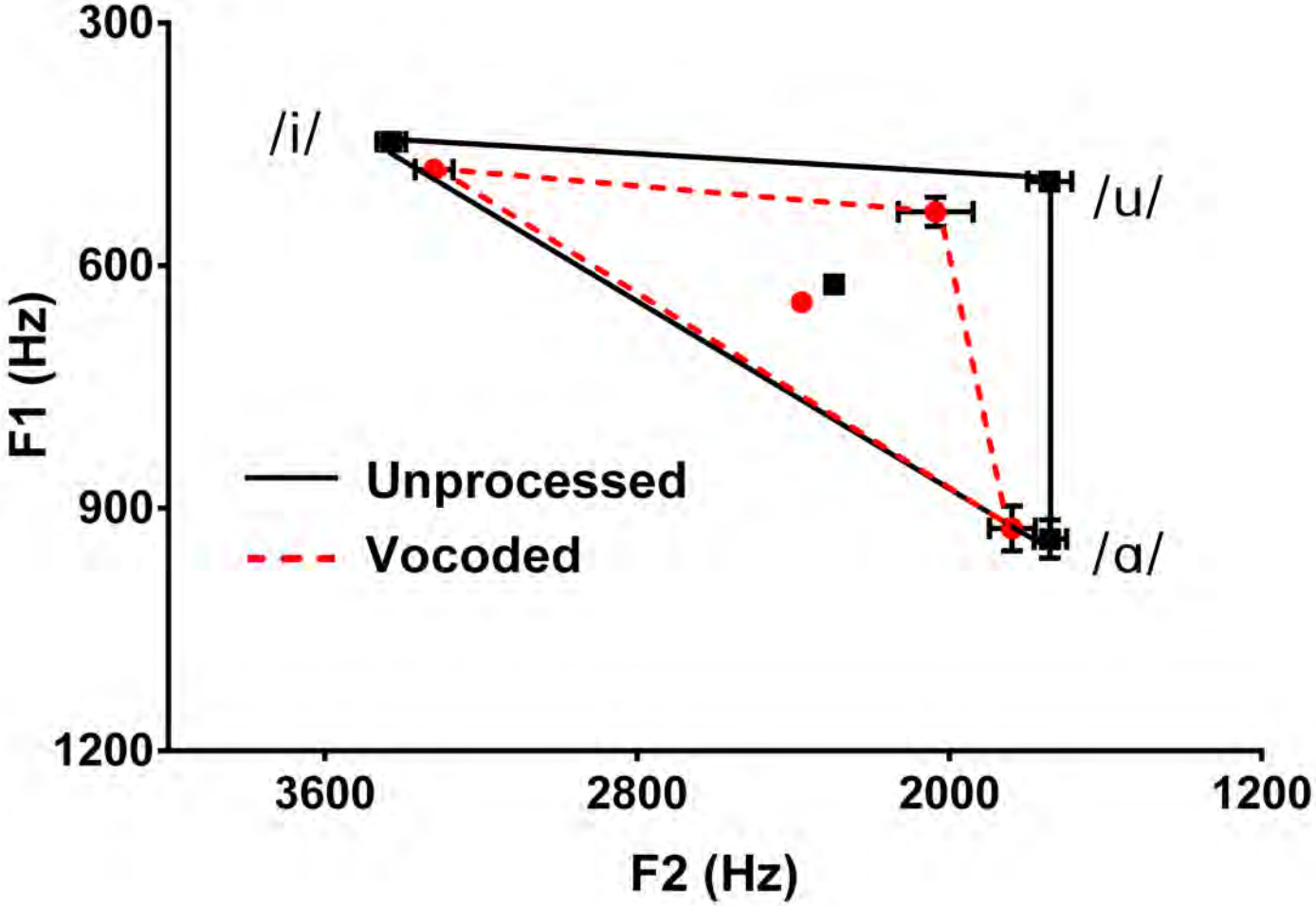
Table 9. Mean within-child SDs (in msec) for voiceless and voiced stop voice onset time, for each condition separately. UP: Unprocessed. VC: Vocoded. Standard deviations are in parentheses.

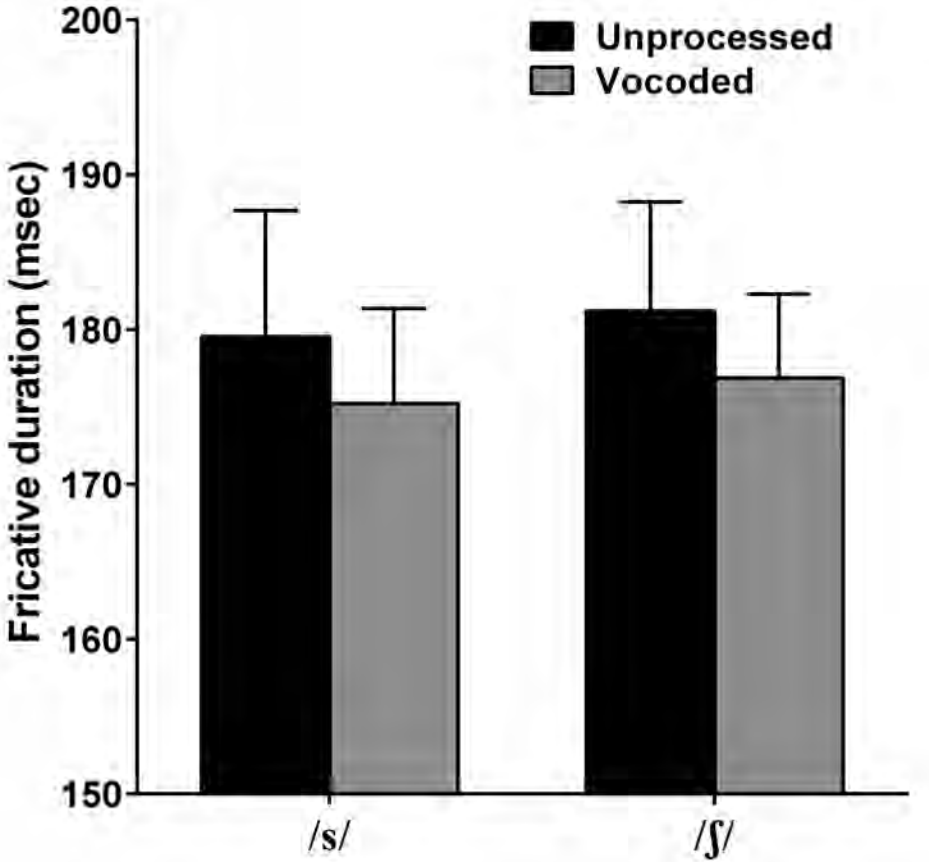
	UP	**VC**
/p/	32.93 (8.37)	32.90 (8.60)
/t/	29.74 (9.27)	33.13 (9.02)
/k/	27.30 (7.33)	32.08 (11.35)
/b/	13.79 (6.99)	17.41 (5.32)
/d/	15.21 (3.63)	19.47 (3.42)
/g/	14.00 (7.57)	14.51 (8.90)

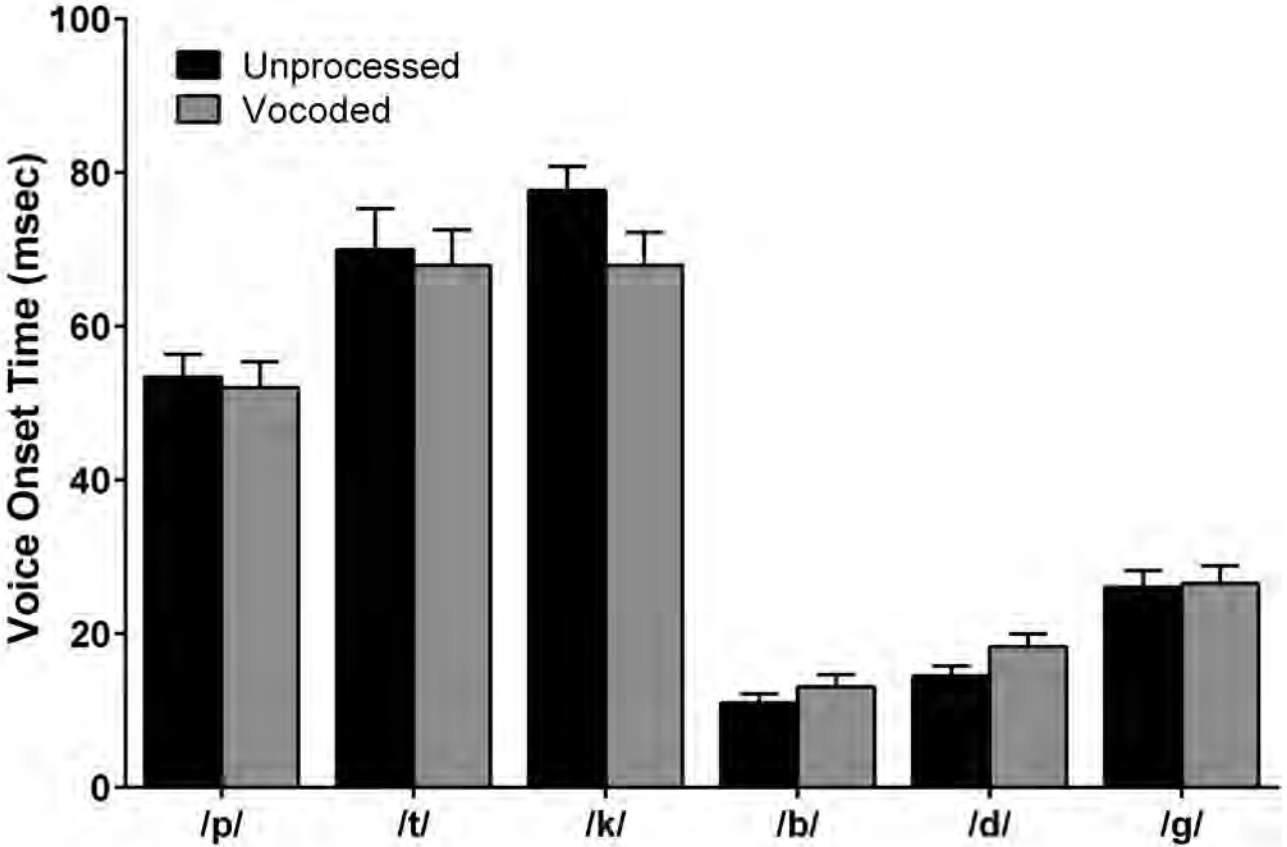

waveform
frequency, kHz
frequency, kHz
0
1
2
3
4
5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
time, sec


8
7
6
5
4
First spectral moments (kHz)
Unprocessed
Vocoded
/s/
/ʃ/


0.0
-0.5
-1.0
-1.5
-2.0
Third spectral moments
Unprocessed
Vocoded
/s/
/ʃ/


300
600
900
1200
F1 (Hz)
3600
2800
2000
1200
F2 (Hz)
/i/
/u/
/ɑ/
Unprocessed
Vocoded


Unprocessed
Vocoded
Fricative duration (msec)
200
190
180
170
160
150
/s/
/ʃ/


Unprocessed
Vocoded
Voice Onset Time (msec)
100
80
60
40
20
0
/p/
/t/
/k/
/b/
/d/
/g/

Appendix.

1	ɹɛbɑdob	33	nɪdobog
2	kæbɑpɑk	34	wɛdodep
3	ɹʌbɑtib	35	nædotip
4	kæbɑtok	36	mɛdotog
5	mɪbebɑb	37	ɹædubig
6	wæbedik	38	gɛdudip
7	mæbepɑg	39	nædudog
8	læbepog	40	gʌduʃeb
9	ɹɛbipep	41	mɪpɑʃɑb
10	næbipok	42	gʌpɑsug
11	kæbiʃig	43	nɛpɑtik
12	kɪbiʃop	44	wæpɑtug
13	lɪbobeb	45	mʌpedeg
14	ɹɪbodub	46	wɛpedig
15	gæbopik	47	næpesig
16	gɪbosib	48	nɪpesok
17	gɛbudup	49	læpibik
18	wʌbuʃɑk	50	gɪpidɑk
19	læbutig	51	gʌpiʃup
20	mɪbutob	52	gʌpituk
21	gɛdɑbok	53	kɛpobɑk
22	wɪdɑduk	54	mæpobep
23	ɹɛdɑʃep	55	nɛpobip
24	kʌdɑtop	56	gæpobub
25	gɛdebug	57	mæpuseb
26	mædepob	58	wɪpusek
27	kɪdeʃɑg	59	mʌpuʃek
28	nɛdeʃɑp	60	wɪpusop
29	ɹædibeg	61	mɛsɑdop
30	mʌdipop	62	gɛsɑpig
31	ɹʌdipeb	63	wʌsɑʃok
32	mædipuk	64	gæsɑteg

65	wɪsebɑg	93	mɪsopɑb
66	ɹɛsebob	94	næsoʃug
67	kɪsetɑb	95	lɪsosip
68	lɛsesub	96	wɪsotɑk
69	wæʃɑpip	97	ɹɪsubib
70	ɹɪʃɑsep	98	kʌsudeb
71	wɛʃɑtɑg	99	nɛsusog
72	gɪʃɑtek	100	mɛsudug
73	kɪʃedɑg	101	lɛtɑdib
74	lɛʃepop	102	lʌtɑsuk
75	nɪʃesɑb	103	ɹʌtɑpib
76	læʃeʃog	104	lɛtɑʃeg
77	wɛʃibup	105	wʌtebop
78	ɹɪʃidek	106	gætepup
79	nɪʃidok	107	kʌteseg
80	kɛʃisɑp	108	lɪteʃob
81	ɹæʃodɑp	109	wætisɑk
82	ɹɛʃopek	110	ɹʌtisup
83	kʌʃosɑg	111	kɛtiteb
84	lɪʃoʃik	112	kɛtitub
85	lʌʃupub	113	ɹætobek
86	gæʃupug	114	wʌtoʃip
87	mɪʃusik	115	mʌtoʃuk
88	gɪʃutep	116	nɪtotup
89	gʌsibɑp	117	kʌtudɑb
90	lʌsibuk	118	lʌtupeg
91	nɛsiʃub	119	kætuʃib
92	kæsitɑp	120	mɛtusob