Research Note

Perception of Sibilants by Preschool Children With Overt and Covert Sound Contrasts

Elizabeth Roepke^a and Françoise Brosseau-Lapré^a

Purpose: This study explores the role of overt and covert contrasts in speech perception by children with speech sound disorder (SSD).

Method: Three groups of preschool-aged children (typically developing speech and language [TD], SSD with /s/~/f/ contrast [SSD-contrast], and SSD with /s/~/f/ collapse [SSD-collapse]) completed an identification task targeting /s/~/f/ minimal pairs. The stimuli were produced by 3 sets of talkers: children with TD, children with SSD, and the participant himself/herself. We conducted a univariate general linear model to investigate differences in perception of tokens produced by different speakers and differences in perception between the groups of listeners.

Results: The TD and SSD-contrast groups performed similarly when perceiving tokens produced by themselves or other children. The SSD-collapse group perceived all speakers more poorly than the other 2 groups of children, performing at chance for perception of their own speech. Children who produced a covert contrast did not perceive their own speech more accurately than children who produced no identifiable acoustic contrast.

Conclusion: Preschool-aged children have not yet developed adultlike phonological representations. Collapsing phoneme production, even with a covert contrast, may indicate poor perception of the collapsed phonemes.

As a group, children with speech sound disorder (SSD) have more difficulties producing speech sounds accurately compared to other children of the same age (e.g., Preston, Irwin, & Turcios, 2015). Several studies have found that children with SSD also tend to have weaker speech perception abilities than their peers with typically developing speech and language skills (TD). For instance, Rvachew and Jamieson (1989) examined the speech discrimination abilities of adults, children with TD, and children with SSD for the voiceless fricative pairs /s/~/f/ and /s/~/θ/. They found that adults and children with TD could distinguish /s/ and /f/ along a continuum; however, of the children with SSD, five of the 12 children could distinguish /s/ and /f/, and seven could not. Similarly, children with TD were generally able to discriminate between /s/ and /θ/, whereas all the children with SSD had difficulties discriminating these phonemes in words. Rvachew and

Jamieson concluded that a subgroup of children with SSD have concomitant perceptual difficulties for the specific sounds they misarticulate.

Similarly, Broen, Strange, Doyle, and Heller (1983) found that some SSD children who misarticulated /r/ and /l/ had perceptual difficulties for these phonemes, but that children with TD were generally accurate in their perception of these phonemes. Other studies have found that children with SSD who misarticulate the /r/ sound are more likely to have difficulty discriminating this sound from /w/ than their peers with TD (Hoffman, Daniloff, Bengoa, & Schuckers, 1985; Ohde & Sharf, 1988). These studies suggest that the ability to perceive a contrast is relevant for producing that contrast accurately, though some children with SSD may have accurate perceptual knowledge and still lack articulatory accuracy. The reason that some children have perceptual deficits and others do not is still unknown.

Children with SSD appear to have especially poor speech perception when listening to imprecise speech targets such as their own speech or the speech of other children with SSD. For example, Rvachew, Ohberg, Grawburg, and Heyding (2003) tested the phoneme perception of children with TD and children with SSD using the Speech Assessment and Interactive Learning System (Rvachew, 1994). The Speech Assessment and Interactive Learning System

^aDepartment of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN

Correspondence to Elizabeth Roepke: john1973@purdue.edu

Editor-in-Chief: Bharath Chandrasekaran

Editor: Megan McAuliffe

Received March 21, 2019

Revision received May 13, 2019

Accepted July 4, 2019

https://doi.org/10.1044/2019_JSLHR-S-19-0127

Disclosure: The authors have declared that no competing interests existed at the time of publication.

assesses speech perception using tokens from speakers with typical and disordered speech. Children hear the target word and point either to a picture of the target word to indicate that the token was produced correctly or to an “X” to indicate that the token was produced incorrectly. Rvachew et al. (2003) found that children with SSD performed more poorly on this perception task than their peers with TD.

There have been few studies on how children with SSD perceive errors in their own speech despite the strong research evidence for perceptual deficits in some children with SSD. Overall, children are less accurate in judging the accuracy of their own speech than that of others (Aungst & Frick, 1964; Lapko & Bankson, 1975; Lof & Synan, 1997; Strömbergsson, Wengelin, & House, 2014; Woolf & Pilberg, 1971). Since children with speech errors must master self-monitoring of speech for carryover of therapeutic gains in articulation, understanding how these children perceive their own speech relative to that of others has salient clinical applications (Preston et al., 2015).

Locke and Kutz (1975) tested 15 kindergarteners with /r~/w/ collapse and 15 with correct /r~/w/ articulation on perception of adult speech and their own speech. In this task, children listened to an adult’s productions of “ring” and “wing” and pointed to a photograph of the intended target. Both groups of children demonstrated proficient perception of the /r~/w/ contrast when listening to adult speech; however, when listening to recordings of themselves saying “ring” and “wing,” the children with /r~/w/ collapse generally perceived both words as “wing,” whereas the children with correct articulation perceived a contrast between /r/ and /w/ in their own speech. In a similar study, Shuster recorded words produced by children with difficulty in correctly articulating /r/. These words were then digitally edited in order to correct the /r/ sound. Children then listened to the original and corrected recordings of themselves and another speaker with /r/ errors then judged each production as correct or incorrect. Shuster found that children with /r/ errors had poor perceptual judgment of misarticulated /r/, both those spoken by themselves and those spoken by others.

Target Contrast

We chose to investigate the /s~/f/ contrast because these are later acquired sounds (Smit, Hand, Freilinger, Bernthal, & Bird, 1990) with similar articulatory settings and can be distinguished by a number of acoustic cues (Li, Edwards, & Beckman, 2009). In adult English speakers, center of gravity and skewness together often differentiate /s/ from /f/ (Jongman, Wayland, & Wong, 2000; Li, Edwards, & Beckman, 2009), though mean, variance, skewness, kurtosis, and slope may differ significantly between these two phonemes (Nissen & Fox, 2005). Li et al. (2009) found that all four spectral moments can differentiate these phonemes, though the center of gravity was the most salient contrast in adults.

Children have not yet developed adultlike perception and production of /s/ and /f/. Perceptually, children rely

more on formant transitions for /s/ + vowel and /f/ + vowel combinations whereas adults use a segmental perception strategy in attending to the fricative noise (e.g., Nittrouer, 1992; Nittrouer & Studdert-Kennedy, 1987). In terms of production, Nissen and Fox (2005) report that spectral mean (center of gravity) is significantly different between /s/ and /f/ among 5-year-old children but not younger ones and that skewness for these phonemes begins to develop at 4 years of age. Nittrouer (1995) found that center of gravity, skewness, and kurtosis differentiate /s/ from /f/ among 3-, 5-, and 7-year-old children.

Because so many acoustic cues can signal the difference between these phonemes in adult and child speech, it is possible that children may form underlying representations of these sounds based on an inaccurate weighting of these acoustic cues. For example, a child may use the less salient cue of kurtosis to signal the difference between /s/ and /f/, producing a “covert contrast” that adult listeners do not perceive. We investigate the use of these acoustic cues in the speech of children with SSD to determine whether children with /s/ and /f/ errors can perceive their own speech accurately.

Aim of Current Study

The purpose of this preliminary study is to investigate the role of overt and covert contrasts in speech perception by children with SSD. To accomplish this aim, we investigated a single speech contrast (/s~/f/) among three groups of children: (a) children with SSD who collapse the target sounds, (b) children with SSD who contrast the target sounds, and (c) children with TD.

We addressed the following research questions:

1. Do TD and SSD children who produce the /s~/f/ contrast accurately discriminate these sounds when listening to recorded productions of children with TD, children with SSD, and themselves?
2. Do children with /s~/f/ collapse have poorer perception of the /s~/f/ contrast than children who produce a contrast between these sounds?
3. Do children with covert /s~/f/ contrast perceive the covert contrast in their own recorded utterances?

Method

Ethics Statement

This experiment was approved by the Purdue University Institutional Review Board. Parents provided written consent, and children provided verbal assent to participate in the experiment.

Participants

Twenty-one children between the ages of 4;0 and 5;11 (years;months) participated in this study. All children were monolingual English speakers, resident in the Midwest, and with normal hearing and structure of the oral speech

mechanism. Participants with autism spectrum disorder, global developmental delay or other neurodevelopmental disorders, or language impairment were excluded from participation. Children were first classified as presenting with TD or SSD according to these criteria: Children with TD had no previous history of speech or language therapy and performed within normal limits on standardized speech and language assessments. A certified speech-language pathologist completed an assessment including Oral Speech Mechanism Screening Evaluation–Third Edition (St. Louis & Ruscello, 2000), Kaufman Brief Intelligence Test–Second Edition (Kaufman & Kaufman, 2004), Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007), Expressive Vocabulary Test–Second Edition (Williams, 2007), and Goldman-Fristoe Test of Articulation–Third Edition (Goldman & Fristoe, 2015). Children also completed a language sample and the Structured Photographic Expressive Language Test–Preschool 2 (Dawson et al., 2005) in order to confirm that language skills were within normal limits. All children in both the TD and the SSD groups scored within normal limits on the measures of nonverbal intelligence (Kaufman Brief Intelligence Test–Second Edition), vocabulary (Peabody Picture Vocabulary Test–Fourth Edition, Expressive Vocabulary Test–Second Edition), and language (Structured Photographic Expressive Language Test–Preschool 2). Children were classified as presenting with SSD if they scored below a standard score of 85 on the Goldman-Fristoe Test of Articulation–Third Edition.

We further classified children with SSD according to phonetic transcription of /s/- and /ʃ/-initial words based on their productions of the stimulus list, listed in the section below. Two graduate students enrolled in a speech-language pathology program with additional training in phonetics completed phonetic transcriptions of each child's productions. The children whose transcriptions of /s/ and /ʃ/ overlapped at least 80% of the time were identified as having /s~/ʃ/ collapse, whereas those whose transcriptions differed at least 80% of the time were identified as having /s~/ʃ/ contrast. Interrater reliability for transcription of /s/ and /ʃ/ was 99.8%. The disagreements were resolved by consensus.

All children with TD presented with an /s~/ʃ/ contrast. The children with SSD were divided into two groups based on transcription: those who contrasted /s~/ʃ/ and those who collapsed /s~/ʃ/. All productions by children who collapsed the target sounds were perceived as /s/ by the adult listeners. Children with misarticulations of the target sounds were included. Some children produced interdental /s/ and either a perceptually accurate /ʃ/ or substituted [ʒ] for /ʃ/. One child substituted /t/ for /s/ but produced /ʃ/ correctly.

Overall, the children were placed into one of three groups: (a) TD with /s~/ʃ/ contrast ($N = 7$; five girls, two boys), (b) SSD with /s~/ʃ/ contrast (SSD-contrast; $N = 7$; four girls, three boys), and (c) SSD with /s~/ʃ/ collapse (SSD-collapse; $N = 7$; five girls, two boys). Acoustic analysis revealed that, of these seven children with /s~/ʃ/ collapse, two did not produce an acoustic contrast, and five produced

a covert contrast (see Table 3). Participant characteristics are presented in Table 1.

Stimuli

The word list included seven monosyllabic /s~/ʃ/ minimal pairs in a consonant–vowel (CV) or consonant–vowel–consonant (CVC) template: *sea-she*, *sew-show*, *sip-ship*, *seat-sheet*, *save-shave*, *sign-shine*, and *self-shelf*. The recordings of four speakers were kept constant across participants: a preschool-aged boy and a preschool-aged girl with SSD who collapsed /s/ and /ʃ/ (SSD speakers), and a young school-aged boy and a young school-aged girl with typical speech who produced the /s~/ʃ/ contrast (TD speakers). One production of each word by each speaker was included in the task. In addition, the children participating in the experiment recorded each word in the list twice, as described below. These recordings of the child's own speech were included in the listening stimuli. Thus, each word pair was heard six times: spoken twice by SSD children with /s~/ʃ/ collapse (SSD speakers), twice by TD children with /s~/ʃ/ contrast (TD speakers), and twice by the child himself/herself (Self).

Procedure

Participants recorded the stimulus words in a quiet room using a Marantz PMD661 MKII broadcast recorder at a sampling rate of 44.1 kHz. During recording, each child saw an image of the target word, saw the orthographic transcription of the target word, and heard an adult model of the target word before producing the word. We processed the sound files using Version 2.3.1 of Audacity(R) recording and editing software (Audacity Team, 2019) to remove background noise if present and equate root-mean-square amplitude to 70 dB SPLA across utterances.

The children returned to the lab for the perception task between 1 and 2 weeks after the recording session. Participants were seated in a quiet room used for research in the Purdue Speech, Language, and Hearing Sciences Department, and the stimuli were presented in the sound field on a tablet. The participants initially completed a training block for the identification task by listening to an adult's productions of the /m~/k/ contrast in word-initial position of five monosyllabic minimal pairs: *man-can*, *mall-call*, *mole-coal*, *mop-cop*, and *mat-cat*. The participants were instructed to touch the picture of the word they heard from two image options. The child completed the training until he or she achieved 90% accuracy. Most of the children reached 100% accuracy on the first attempt. No child required more than two attempts to master the task during training.

Following the training block, the children began the experimental blocks. The target word was played, and images of the corresponding minimal pair set appeared on screen. After listening to each word, the children touched the image of the word they heard using the touchscreen tablet. The children had previously seen the images associated

Table 1. Participant characteristics.

Measure	TD			SSD-contrast			SSD-collapse		
	(n = 7)			(n = 7)			(n = 7)		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Age (months)	61.9	6.0	52–68	62.8	5.7	55–68	58.9	4.0	53–65
KBIT-2	104.7	4.2	100–109	101.7	10.0	85–115	103.1	5.2	96–112
GFTA-3	100.1	11.1	87–118	71.1	13.4	61–83	56.4	12.2	40–77
PPVT-4	120.8	16.9	96–134	117.7	10.7	106–122	114.6	5.4	109–115
EVT-2	124.8	13.4	104–137	117.0	8.9	104–123	113.3	13.2	98–131
SPELT-P 2	108.2	15.5	92–125	111.7	5.0	108–120	107.7	6.3	98–115

Note. TD = typically developing speech and language; SSD = speech sound disorder; KBIT-2 = nonverbal matrices subtest of the Kaufmann Brief Intelligence Test–Second Edition (Kaufman & Kaufman, 2004); GFTA-3 = Goldman-Fristoe Test of Articulation–Third Edition (Goldman & Fristoe, 2015); PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); EVT-2 = Expressive Vocabulary Test–Second Edition (Williams, 2007); SPELT-P 2 = Structured Photographic Expressive Language Test–Preschool 2.

with the target words during the recording session, and the experimenter provided verbal labels of the images during the session if the child requested. The presentation of the audio stimuli was semirandomized with the following conditions: The same word by different speakers was not presented in two consecutive trials, and a single speaker's voice did not occur in two consecutive trials.

Each child completed a total of 84 trials, 28 words under each condition: children with TD with /s/~/f/ contrast (one boy, one girl), children with SSD with /s/~/f/ collapse (one boy, one girl), and their own voice. Children were free to play the audio for each trial multiple times before selecting a response. Most children responded after hearing the audio once, but the audio was replayed if the child was speaking during the automatic playback. No feedback on performance was given to the children during the identification task. After every seven trials, participants participated in a break with a reward activity such as playing a short game or placing a sticker on a sheet. Stimuli were presented, and responses were collected using E-Prime 3.0 (Schneider, Eschman, & Zuccolotto, 2012).

Acoustic Analysis

We tagged the onset and offset of the sibilants /s/ and /f/ for each child's 28 recorded single-word productions at zero crossings using Praat (Boersma & Weenink, 2012). Onset was defined as the first aperiodic noise on the waveform together with high-frequency noise on the spectrogram and, offset, as the first zero crossing of the periodic waveform of the following vowel, following Li et al. (2009). We used a Praat script based on the methods in Jongman et al. (2000) and used in Seidl, Brosseau-Lapr, and Goffman (2018) to extract the highest spectral peak, center of gravity (first spectral moment), standard deviation (second spectral moment), skewness (third spectral moment), kurtosis (fourth spectral moment), and duration of the fricative. The center of gravity, standard deviation, skewness, and kurtosis were calculated for the 40-ms full Hamming window at the middle of the fricative noise. We used the mid fricative value for these calculations to

minimize the effect of co-articulation with the following vowel.

Results

We conducted a univariate general linear model with fixed factors of the speaker group (TD speakers, SSD speakers, or Self speakers) and the listener group (TD listeners, SSD-contrast listeners, or SSD-collapse listeners) using SPSS Version 25 (IBM Corp.). The dependent variable was percentage of perceptual accuracy, calculated across all of the responses per target speaker group. Accurate responses matched adult perception, which was determined by two trained graduate students who independently transcribed each speaker's productions. The main effect of the speaker group on perceptual accuracy was significant, $F(2, 1755) = 13.386$, $MSE = .163$, $p < .001$, $\eta_p^2 = .015$, as was the main effect of the listener group, $F(2, 1755) = 72.337$, $MSE = .163$, $p < .001$, $\eta_p^2 = .076$.

There was also a significant interaction between the speaker group and the listener group, $F(4, 1737) = 3.315$, $MSE = .163$, $p = .01$, $\eta_p^2 = .007$, presented in Figure 1. Descriptive statistics for each interaction are reported in Table 2. We conducted post hoc comparison of the interactions between the speaker and listener groups using the Tukey's honestly significant difference (HSD) test.

We first compared the effect of the speaker group on the listener group. With regard to perception of the words produced by TD speakers, there was no significant difference between the TD listeners and SSD-contrast listeners ($p = 1.0$, $d = .05$). SSD-collapse listeners were significantly worse than TD listeners ($p < .001$, $d = .48$) and SSD-contrast listeners ($p = .001$, $d = .43$) at perceiving the difference between /s/ and /f/ produced by TD speakers.

There was not a significant difference between the TD listeners and SSD-contrast listeners for perception of SSD speakers ($p = .067$, $d = .31$). SSD-collapse listeners were significantly worse than TD listeners ($p < .001$, $d = .63$) and SSD-contrast listeners ($p = .009$, $d = .32$) at

Figure 1. Accuracy in identifying /s/ and /ʃ/ produced by the three speaker groups. Error bars represent 95% confidence interval.

perceiving the difference between /s/ and /ʃ/ by SSD speakers. Finally, there was not a significant difference between the TD listeners and SSD-contrast listeners for perception of Self ($p = .454$, $d = .24$). However, SSD-collapse listeners were significantly worse than TD listeners ($p < .001$, $d = .88$) and SSD-contrast listeners ($p < .001$, $d = .62$) at matching adult perception for /s/ and /ʃ/ for the speaker condition Self.

Because the TD and SSD-contrast groups produced a distinction between the target phonemes, the word that they intended to produce for the Self speaker condition matched the adult perception of the word. However, when

the SSD-collapse children intended to produce /ʃ/, the adult listeners perceived /s/. We therefore ran our general linear model again where the dependent variable of perceptual accuracy was based on the child's intended production rather than on adult perception of the production to investigate whether the children perceived the word that they had intended to produce. Main effects of the speaker group, $F(2, 1755) = 128.219$, $p < .001$, $\eta_p^2 = .127$, and the listener group, $F(2, 1755) = 28.887$, $p < .001$, $\eta_p^2 = .032$, were significant, as were the interaction effects of the listener group by the speaker group, $F(4, 1755) = 12.180$, $p < .001$, $\eta_p^2 = .027$. Post hoc Tukey's HSD analysis of interactions for Self as speaker revealed no significant differences between TD listeners ($M = .88$, $SD = .33$) and SSD-contrast listeners ($M = .80$, $SD = .40$; $p = .626$, $d = .22$). SSD-collapse listeners ($M = .50$, $SD = .50$) performed significantly worse than TD listeners ($M = .88$, $SD = .33$, $p < .001$, $d = .90$) and SSD-contrast listeners ($M = .80$, $SD = .40$, $p < .001$, $d = .66$).

We then compared the effect of the speaker group on the listener group, using adult perception as the measure of perceptual accuracy. The Tukey's HSD test revealed that the TD listeners perceived TD speakers as well as they perceived SSD speakers ($p = .986$, $d = .13$) and Self ($p = 1.00$, $d = .06$). SSD-contrast listeners perceived TD speakers and Self with similar accuracy ($p = .372$, $d = .25$), but perceived SSD speakers less accurately than they perceived TD speakers ($p = .009$, $d = .38$). SSD-collapse listeners perceived TD speakers more accurately than they perceived SSD speakers ($p = .047$, $d = .27$) and more accurately than

Table 2. Accuracy of perception by speaker and listener groups.

Speaker	Listener	M	SD	95% CI
TD	TD	.903	.30	[.847, .960]
	SSD-contrast	.888	.32	[.831, .944]
	SSD-collapse	.719	.45	[.663, .776]
SSD	TD	.862	.35	[.806, .919]
	SSD-contrast	.740	.44	[.683, .796]
	SSD-collapse	.592	.49	[.535, .648]
Self: adult perception	TD	.883	.32	[.826, .939]
	SSD-contrast	.796	.40	[.739, .852]
	SSD-collapse	.515	.50	[.459, .572]
Self: intended production	TD	.878	.33	[.831, .929]
	SSD-contrast	.796	.40	[.739, .852]
	SSD-collapse	.500	.50	[.429, .570]

Note. CI = confidence interval; TD = typically developing speech and language; SSD = speech sound disorder.

they perceived Self ($p < .001$, $d = .43$). There was not a significant difference in how SSD-collapse listeners perceived SSD speakers and Self speakers ($p = .629$, $d = .16$).

Overall, TD listeners and SSD-contrast listeners performed similarly across all speaker groups. However, SSD-contrast listeners had less accurate perception of SSD speakers than of TD speakers. SSD-collapse listeners had poorer perception across all speaker groups than the TD and SSD-contrast listeners.

We next tested the hypothesis that children who produce a covert contrast can perceive that contrast in their own speech. To identify children with covert contrast, we compared the means of acoustic measures taken for /s/ and /ʃ/ for the SSD-collapse group using independent t tests. We found that five of the children contrasted the speech sounds covertly using one of the acoustic measures we analyzed, and two of the children did not produce significant acoustic differences between these sounds. Results are presented in Table 3. We then compared the performance on perception of self for the five SSD-collapse children who produced covert contrast. We conducted a one sample t test and tested the null hypothesis that children performed at chance or at 50% accuracy. The dependent variable for this comparison was accuracy of perceiving the word as the child had intended to produce it, rather than the adult perception of the word. We found that these children performed at chance for identifying whether they had intended to produce /s/ or /ʃ/ in their own speech, $M = 0.5$, $SD = .04$, $t(139) = -0.337$, $p = .737$. In other words, the children who produced covert contrast did not perceive this contrast in their own speech.

One possible reason that the children with SSD-collapse performed at around 50% accuracy could be that their perception matched the adult listener's perception, who identified all intended productions of /ʃ/ as [s]. If this were true, then the SSD-collapse listener performance would be high on /s/-initial words and low on /ʃ/-initial words. We tested this hypothesis by comparing the performance of the two phonemes produced and perceived by the SSD-collapse group for the Self speaker condition, with the speaker's intended production as the measure of perceptual accuracy. Mean perception of /s/ was 51%, and mean perception of /ʃ/ was 49%. A one-way analysis of variance revealed no significant differences between performance on perception of /s/ and of /ʃ/, $F(1, 194) = 0.081$, $MSE = 0.02$, $p = .776$, $d = 0.04$. In other words, children with collapse were just as likely to

identify their own productions of intended /s/ as /ʃ/ as they were to identify intended /ʃ/ as /s/.

Discussion

In this preliminary study, we investigated perception of the /s/~/ʃ/ contrast by children with and without SSD. We expected that children who contrasted /s/~/ʃ/ in their own speech, whether TD or SSD, would accurately perceive that contrast in minimal pairs. We also expected that children with /s/~/ʃ/ collapse would have poorer perception of these sounds when produced by others but would discriminate these sounds in their own speech using covert acoustic contrast cues.

Our first research question was whether TD and SSD children who produce the /s/~/ʃ/ contrast will accurately discriminate these sounds when listening to recorded productions of TD children, children with SSD, and themselves. We found that children who produced the /s/~/ʃ/ contrast in their speech perceived the contrast well in TD speech, although their performance was not at ceiling. This finding supports previous research that preschool-aged children have not yet developed adultlike underlying representations of speech sounds (e.g., Sutherland & Gillon, 2005). In fact, Hazan and Barrett (2000) found that children did not develop adultlike perception of consonant contrasts until age 12 years. One aspect of adultlike perception that our study measured is the ability to identify speech sounds produced by speakers of various abilities. While children with SSD-contrast performed similarly to their peers with TD on perception of clear /s/~/ʃ/ tokens, their perception of unclear tokens, spoken by children with SSD, was significantly lower than their perception of TD speech. This finding suggests that children with SSD may have weaker phonological knowledge than children with TD, even for the sounds that they articulate correctly. These children may have correct discrimination of these sounds at the end points, when they are most distinct, but may not have adultlike knowledge of the within-category variability present in these sounds.

These findings are similar to those of Rvachew and Jamieson (1989) who presented /s/ and /ʃ/ on a synthesized continuum. Children with TD were closer in their perceptual responses to the adult response than were children with SSD. In addition, as the synthesized phoneme became less /s/-like and more /ʃ/-like, the gap in accuracy between

Table 3. Acoustic contrasts produced by speech sound disorder-collapse speakers.

Speaker	Acoustic measure	M (SD)		t	p	Effect size (Cohen's d)
		s	sh			
A	Kurtosis	2.0 (2.5)	5.3 (4.2)	-2.507	.020	0.95
B	Center of gravity	11047 (1797)	9701 (1652)	2.064	.049	0.78
	Skewness	-0.095 (.69)	0.44 (.48)	-2.396	.024	0.90
C	Center of gravity	9810 (1348)	8736 (637)	2.697	.012	1.02
D	Duration	211.4 (24.0)	236.3 (34.3)	-2.249	.033	0.84
E	Center of gravity	8058 (946)	6851 (1850)	2.325	.027	0.82

children with TD and children with SSD widened, as children with SSD were more likely to respond with /s/ than with /ʃ/. Similarly, the children in our study in the SSD-contrast group appear to have a weaker understanding than children with TD of the acoustic cues that signal the difference between the sounds despite producing these sounds correctly.

Overall, the children in the TD and SSD-contrast listener groups in the current study performed well on perception of their own speech. One interesting case in our study was a child with SSD whose contrast included substituting /t/ for /s/, but produced /ʃ/ accurately. This child correctly identified 100% of words produced by TD speakers. However, he did not perceive this contrast well in his own speech, performing a little above chance (60% accuracy). We expected that, because he had consistently identified /ʃ/ correctly in TD speech, he also would have identified /ʃ/ correctly in his own speech. However, examination of the data revealed that, for his own speech, he inconsistently identified /t/ as /s/ and /ʃ/ and also inconsistently identified /ʃ/ as /s/ and /ʃ/. Although he used the continuant feature to distinguish /s/ and /ʃ/ in his production, he did not use this cue to discriminate between these sounds perceptually. The other children with SSD-contrast ranged in accuracy for perception of their own speech from 71% to 93%. These children reliably produced a /s/~/ʃ/ contrast that adult listeners perceived correctly, but they themselves were not consistent in their discrimination of their own productions of these sounds. It is possible that children with SSD-contrast produced a less robust contrast between these phonemes than the TD speakers produced, and that this less robust contrast was more difficult to perceive.

Our data supported our second research question, whether children with /s/~/ʃ/ collapse will have poorer perception of the /s/~/ʃ/ contrast than children who produce a contrast between these sounds. This finding adds to previous research that children with SSD may have specific perceptual deficits affecting the speech sounds they misarticulate, especially if that misarticulation is substitution for a phoneme of the same manner class (e.g., Broen et al., 1983; Rvachew & Jamieson, 1989). However, an unexpected finding was that children with SSD-collapse were able to discriminate /s/ and /ʃ/ better in TD speech than in their own or SSD speech. We had expected similarly poor perception of all speakers. These data suggest that the children in the SSD-collapse group do have underlying representations of both the /s/ and /ʃ/ phonemes, though these representations may be less well defined than those of the SSD-contrast and TD groups.

Our third research question, whether children with covert /s/~/ʃ/ contrast will perceive the covert contrast in their own recorded utterances, was not supported by our data. The children with covert contrast performed at chance on discriminating the target sounds in their own recorded speech. Our preliminary finding that children with covert contrast do not perceive the contrast that they produce adds to previous research that suggests that covert contrast is a sign of imprecise but developing phonological knowledge.

For example, Tyler et al. (1993) concluded that children who produced covert contrast required fewer speech therapy treatment sessions than children who produced no contrast. Covert contrast has been suggested as a step that the child takes while acquiring his or her native phonological system, by marking a phonological contrast before mastering the articulatory differences between the phonemes (e.g., Glaspey & MacLeod, 2010). As the child acquires some perceptual knowledge of the target contrast, he or she may produce covert acoustic contrasts to signal this contrast before arriving at the correct motor targets to produce a perceptible distinction (Tyler, Edwards, & Saxman, 1990; Tyler & Saxman, 1991). Overall, our preliminary findings suggest that covert contrast may exist in the absence of accurate perceptual contrast. However, we did not test the children's perception of adult speech, which may include more distinct exemplars of the contrast that children are able to perceive more reliably.

Future directions for this research include further systematic investigation of children's ability to discriminate phonological contrasts in their own and others' speech for children with and without covert contrast. Furthermore, future research may include more speech sound contrasts than the one investigated in this research note.

Acknowledgments

The work reported in this article was supported in part by a grant from the National Institutes of Health (awarded to Françoise Brosseau-Lapr, NIDCD Grant R21DC016142). We would like to thank the children and their parents who participated in the study. We acknowledge the contributions of the Purdue Child Phonology Laboratory personnel with data collection and data entry, and we would like to thank Krista Riegsecker for her assistance with acoustic analyses and Kathryn Bower for her assistance with transcription.

References

- Audacity Team.** (2019). Audacity(R): Free Audio Editor and Recorder (Version 2.3.1) [Computer application]. Retrieved from <https://audacityteam.org/>
- Aungst, L. F., & Frick, J. V.** (1964). Auditory discrimination ability and consistency of articulation of /r/. *Journal of Speech and Hearing Disorders*, 29, 76–85. <https://doi.org/10.1044/jshd.2901.76>
- Boersma, P., & Weenink, D.** (2012). Praat: Doing phonetics by computer (Version 5.3.82) [Computer software]. Amsterdam: Institute of Phonetic Sciences.
- Broen, P. A., Strange, W., Doyle, S. S., & Heller, J. H.** (1983). Perception and production of approximant consonants by normal and articulation-delayed preschool children. *Journal of Speech and Hearing Research*, 26(4), 601–608. <https://doi.org/10.1044/jshr.2604.601>
- Dawson, J. I., Stout, C., Eyer, J., Tattersall, P. J., Fonkalsrud, J., & Croley, K.** (2005). *Structured Photographic Expressive Language Test—Preschool 2 (SPELT-P 2)*. DeKalb, IL: Janelle Publications.
- Dunn, L. M., & Dunn, D. M.** (2007). *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4)*. Pearson Assessments.

- Glaspey, A. M., & MacLeod, A. A. N. (2010). A multi-dimensional approach to gradient change in phonological acquisition: A case study of disordered speech development. *Clinical Linguistics & Phonetics*, 24(4–5), 283–299. <https://doi.org/10.3109/02699200903581091>
- Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation—Third Edition (GFTA-3)*. San Antonio, TX: Pearson.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396. <https://doi.org/10.1006/jpho.2000.0121>
- Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. H. (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders*, 50(1), 46–53. <https://doi.org/10.1044/jshd.5001.46>
- IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: Author.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252–1263. <https://doi.org/10.1121/1.1288413>
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test—Second Edition (KBIT-2)*. Circle Pines, MN: AGS.
- Lapko, L. L., & Bankson, N. W. (1975). Relationship between auditory discrimination, articulation stimulability, and consistency of misarticulation. *Perceptual and Motor Skills*, 40, 171–177. <https://doi.org/10.2466/pms.1975.40.1.171>
- Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37(1), 111–124. <https://doi.org/10.1016/j.wocn.2008.10.001>
- Locke, J. L., & Kutz, K. J. (1975). Memory for speech and speech for memory. *Journal of Speech and Hearing Research*, 18(1), 176–191. <https://doi.org/10.1044/jshr.1801.176>
- Lof, G., & Synan, S. (1997). Is there a speech discrimination/perception link to disordered articulation and phonology? A review of 80 years of literature. *Contemporary Issues in Communication Sciences and Disorders*, 24, 62–77. Retrieved from <https://www.asha.org/uploadedFiles/asha/publications/cicsd/1997IsThereaSpeechDiscrimination.pdf>
- Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118(4), 2570–2578. <https://doi.org/10.1121/1.2010407>
- Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20(3), 351–382.
- Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, 97(1), 520–530. <https://doi.org/10.1121/1.412278>
- Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, 30(3), 319–329. <https://doi.org/10.1044/jshr.3003.319>
- Ohde, R. N., & Sharf, D. J. (1988). Perceptual categorization and consistency of synthesized /r-w/ continua by adults, normal children and /r/-misarticulating children. *Journal of Speech and Hearing Research*, 31(4), 556–568. <https://doi.org/10.1044/jshr.3104.556>
- Preston, J. L., Irwin, J. R., & Turcios, J. (2015). Perception of speech sounds in school-aged children with speech sound disorders. *Seminars in Speech and Language*, 36(4), 224–233. <https://doi.org/10.1055/s-0035-1562906>
- Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research*, 37(2), 347–357. <https://doi.org/10.1044/jshr.3702.347>
- Rvachew, S., & Jamieson, D. G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders*, 54(2), 193–208. <https://doi.org/10.1044/jshd.5402.193>
- Rvachew, S., Ohberg, A., Grawburg, M., & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. *American Journal of Speech-Language Pathology*, 12(4), 463–471. [https://doi.org/10.1044/1058-0360\(2003\)092](https://doi.org/10.1044/1058-0360(2003)092)
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools Inc.
- Seidl, A., Brosseau-Lapr, F., & Goffman, L. (2018). The impact of brief restriction to articulation on children's subsequent speech production. *The Journal of the Acoustical Society of America*, 143(2), 858–863. <https://doi.org/10.1121/1.5021710>
- Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research*, 41(4), 941–950. <https://doi.org/10.1044/jslhr.4104.941>
- Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55, 779–798. <https://doi.org/10.1044/jshd.5504.779>
- St. Louis, K. O., & Ruscello, D. M. (2000). *Oral Speech Mechanism Screening Examination—Third Edition (OSMSE-3)*. San Antonio, TX: The Psychological Corporation.
- Strmbergsson, S., Wengelin, ., & House, D. (2014). Children's perception of their synthetically corrected speech production. *Clinical Linguistics & Phonetics*, 28(6), 373–395. <https://doi.org/10.3109/02699206.2013.868928>
- Sutherland, D., & Gillon, G. T. (2005). Assessment of phonological representations in children with speech impairment. *Language, Speech, and Hearing Services in Schools*, 36(4), 294–307. [https://doi.org/10.1044/0161-1461\(2005\)030](https://doi.org/10.1044/0161-1461(2005)030)
- Tyler, A. A., Edwards, M. L., & Saxman, J. H. (1990). Acoustic validation of phonological knowledge and its relationship to treatment. *Journal of Speech and Hearing Disorders*, 55(2), 251–261. <https://doi.org/10.1044/jshd.5502.251>
- Tyler, A. A., Figurski, G. R., & Langsdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech and Hearing Research*, 36(4), 746–759. <https://doi.org/10.1044/jshr.3604.746>
- Tyler, A. A., & Saxman, J. H. (1991). Initial voicing contrast acquisition in normal and phonologically disordered children. *Applied Psycholinguistics*, 12(4), 453–479. <https://doi.org/10.1017/S0142716400005877>
- Williams, K. T. (2007). *Expressive Vocabulary Test—Second Edition (EVT-2)*. Circle Pines, MN: AGS.
- Woolf, G., & Pilberg, R. (1971). A comparison of three tests of auditory discrimination and their relationship to performance on a deep test of articulation. *Journal of Communication Disorders*, 3, 239–249. [https://doi.org/10.1016/0021-9924\(71\)90030-X](https://doi.org/10.1016/0021-9924(71)90030-X)