Variability in /s/ Production in Children and Adults: Evidence From Dynamic Measures of Spectral Mean

Benjamin Munson
University of Minnesota,
Minneapolis

Previous research has found developmental decreases in temporal variability in speech. Relatively less work has examined spectral variability, and, in particular, variability in consonant spectra. This article examined variability in productions of the consonant /s/ by adults and by 3 groups of children, with mean ages of 3;11 (years;months), 5;04, and 8;04. Specifically, it measured the influence of age, phonetic context, and syllabic context on variability. Spectral variability was estimated by measuring dynamic spectral characteristics of multiple productions of /s/ in sV, spV, and swV sequences, where the vowel was either /a/ or /u/. Mean duration, variability in duration, and coarticulation were also measured. Children were found to produce /s/ with greater temporal and spectral variability than adults. Duration and coarticulation were comparable across the 4 age groups. Spectral variability was greater in swV contexts than in sV or spV sequences. The lack of consistent effects of phonetic context on spectral variability suggests that the developmental differences were related to subtle variability in place of articulation for /s/ in the children's productions.

KEY WORDS: coarticulation, variability, children, spectral mean

It is axiomatic that children, as a group, speak differently from adults. Early speech production is characterized by systematic substitutions, additions, deletions, and distortions of target sounds. Throughout the preschool and early elementary years, these differences gradually disappear, and children begin to produce sounds that adults identify as matching the target form. However, measurable differences between children and adults' speech persist even after children produce phonemes that are judged to be correctly produced. This has been documented by numerous acoustic studies of the duration, coordination, and intraspeaker variability of children's speech.

Smith (1978) found age-related declines in average duration of nonsense words produced by two groups of preschool children, and by adults. A subsequent study by Kent and Forner (1980) examined the durations of individual speech sounds in sentences produced by adults and by three groups of children: 4-year-olds, 8-year-olds, and 12-year-olds. On average, the youngest group of children produced sounds with longer durations than the adults, although considerable variability was noted among children, and among the different sounds measured. Smith, Kenney, and Hussain (1996) longitudinally examined age-related changes in duration. Again, group analyses showed that segment durations decreased with age, although these changes were not present for each individual participant.

Many of the same studies also examined trial-to-trial variability in duration and found that children produce speech with greater temporal variability than that of adults. The findings in these studies (Kent & Forner, 1980; Smith, 1978; Smith et al., 1996) closely parallel those for mean duration. Both Smith and Kent and Forner found that young children produce greater trial-to-trial variability than adults, although this was not true for every child and for every measure. Smith et al. found that, as a group, children's variability declines through development, although this was not evident for each individual.

A parallel line of research has examined variability in the spectral characteristics of children's speech. Early work by Eguchi and Hirsch (1969) found declines in variability in fundamental frequency and vowel formant frequencies in children ages 3 to 13 years. More recently, Lee, Potamianos, and Naryanan (1999) examined variability in temporal and spectral characteristics of repeated and read speech in a large cohort of children ages 5 to 18 years and in adults. Developmental changes were different for different acoustic measures. Children produced adult-like variability in vowel duration by age 7, but adult-like variability in other acoustic measures was not achieved until the middle teenage years, depending on the measure and the talker's sex.

Finally, one recent study (Munson, 2001) examined intraspeaker variability in the spectral characteristics of fricatives. Munson examined spectra of word-medial /sp/, /ʃp/, /st/, and /ʃt/ sequences produced by preschool children, early elementary-age children, and adults. Results indicated that both /s/ and /ʃ/ were spectrally more variable before /p/ than before /t/. The finding that fricatives were spectrally more variable when followed by a stop with a bilabial place of articulation suggested that this might be due to variable coordination between the tongue-tip raising articulatory gestures for the fricatives /s/ and /ʃ/ and the lip-closure gesture for the /p/. In addition, Munson found that variability of /s/ and /ʃ/ spectra declined with age in both phonetic contexts, although this did not achieve statistical significance.

A related line of research has also found that children's speech demonstrates patterns of coarticulation different from those of adults. In a series of articles, Nittrouer and colleagues (Nittrouer, 1993, 1995; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, Studdert-Kennedy, & Neely, 1996) reported on the development of coarticulation in children. In their studies, young children's fricative productions demonstrated greater coarticulation than older children's: Both /s/ and /ʃ/ were found to have lower F2 frequencies in the context of the vowel /u/ than the vowel /i/. This lower frequency before /u/ was attributed to anticipatory lip rounding, which lengthens the vocal tract, thereby lowering formant frequencies (Nittrouer et al., 1989, 1996). The greater coarticulation in children can occur across intervening segments: Children show more vowel-to-vowel coarticulation in schwa–consonant–vowel sequences than do adults (Nittrouer et al., 1996). These differences are not attributable to age-related differences in vocal-tract geometry, and may reflect children utilizing a production strategy in which speech movements are planned for syllable-sized units rather than segment-by-segment. Other studies have found no difference in coarticulation as a function of age (Katz, Kripke, & Tallal, 1991; Sereno, Baum, Marean, & Lieberman, 1987).

The picture that emerges from these studies is that phonetic development extends beyond the point where children produce phonemes that are identified as matching the adult forms. The timing and coordination of children's speech movements, as well as variability in those measures, continues to be identifiably different from adults throughout the elementary school and teenage years. As pointed out by Nittrouer (1993, 1995) and Stathopoulos (1995), these do not mature uniformly: Some aspects of children's speech production resemble those of adults earlier than others, and individual children progress at different paces. In any given comparison between a group of children and a group of adults, some measures will differ significantly and others will not. The production of speech in mature adults involves the coordination of many different subsystems to achieve the task-directed goal of producing intelligible speech. In development, these subsystems mature at different rates, leading to various degrees of achievement of adult-like speech.

In this article I continued this line of inquiry by examining age-dependent changes in within-speaker variability for a single sound, /s/. Specifically, I examined whether children demonstrate more trial-to-trial variability in /s/ production than adults by measuring variability in /s/ spectra. This research differs from previous research in two ways. First, the majority of previous studies of acoustic variability have focused on measures of duration. The few studies that have examined spectral variability have examined variability in vowel-formant measures. This study examined whether similar age-related declines in variability are evident for consonant spectra.

Second, this research examined spectral variability as a function of phonetic context. The stimuli in this study contained /s/ in word-initial sV and sCV sequences, where the vowel was either /a/ or /u/ and the consonant was either /p/ or /w/. Previous research has shown that both adults and children produce fricatives with lower mean frequency prior to /u/ than /a/ (e.g., Nittrouer et al., 1996). The lowering of the frequency of /s/ is attributable to the lengthening of the vocal tract that occurs when the lips are rounded in anticipation of the following /u/. The articulatory movements for /a/ are predicted to have little influence on the acoustics of /s/, as they do not serve to lengthen the vocal tract. Although previous research has not examined the influence of the following consonant on fricative acoustics, predictions can be made based on models of the articulatory and acoustic characteristics of these sounds (e.g., Shadle, 1990; Stevens, 1998). These models predict that the lip rounding associated the labial–velar approximant consonant /w/ will cause the spectrum of /s/ to lower, much the way that lip rounding for /u/ has caused the spectrum to lower in previous studies. Like /w/, the sound /p/ is produced with a movement of the lips; however, this gesture is more rapid, and it is predicted to influence only the portion of the /s/ immediately adjacent to the stop closure. If the coordination between the lip-rounding movements in /u/ and /w/ and the tongue-tip raising movements of /s/ is variable within speakers, then /s/ should demonstrate more spectral variability in the /su/ and swV phonetic contexts than in the /sa/ and spV contexts, which are predicted to have little influence on /s/ spectra.

In this study, spectral variability in /s/ was estimated using a technique adapted from Munson (2001). Munson recorded speakers producing multiple tokens of fricative consonants in different phonetic contexts. The first spectral moment, spectral mean, was calculated for nonoverlapping 10 ms windows of frication noise in each token. Spectral means were calculated by treating the power spectrum of a fricative as a random distribution of numbers and by taking the mean of that distribution (Forrest, Weismer, Milenkovic, & Dougall, 1988). Previous research has shown that the spectral mean discriminates among the four fricative places of articulation in English (Jongman, Wayland, & Wong, 2000) and can be used to measure coarticulation in children's and adults' speech (Nittrouer, 1995). The spectral means were then plotted with respect to their location in the fricative, expressed as a proportion of the total duration of the fricative. Proportion of duration was used as the predictor variable in these regressions rather than raw duration, so that fricatives with different durations could be compared with one another. All of the repetitions by a speaker in a given context were plotted together. Nonlinear regression equations of the form $y = ae^{bx}$ were used to quantify the amount of dispersion in the resulting scatterplots. Specifically, nonlinear regression was used to predict values of the spectral mean of nonoverlapping 10 ms windows of frication noise from their time-normalized position in the fricative. Spectral moments from nonoverlapping windows were used to satisfy the requirement that pairs of observations analyzed in a regression be independent. The measure of model fit, R^2, was used as an index of intraspeaker variability. This measure of variability does not have the disadvantages associated with other measures of variability, such as standard deviation or the coefficient of variance (the standard deviation divided by the mean). Previous research (e.g., Weismer, 1991) has argued that both of these types of measures are problematic. Unlike the standard deviation, this R^2 is not correlated with the average value of the distribution whose variability is being calculated. Unlike the coefficient of variance, the R^2 measure is not a ratio, the statistical properties of which are potentially different from those of a nonratio variable.

The current study had two research objectives. First, a modified version of the method presented in Munson (2001) was used to examine whether children demonstrate more variable production of /s/ than adults. This hypothesis would be supported if children consistently demonstrate greater scatter in spectral mean values than adults in all contexts. Second, this article examined the influence of phonetic context on spectral variability. In the sV contexts, it was predicted that /s/ would be produced more variably before /u/ than before /a/. Variable coordination between the gestures required to produce /u/ and the tongue-tip raising gesture to produce /s/ would result in acoustic variability. Variability in timing between the tip-raising gesture for /s/ and the gestures associated with /a/ would not result in acoustic variability. In the sCV contexts, it was predicted that /s/ would show more variability before /w/ than before /p/, for similar reasons. In addition to spectral variability, measures of mean duration, variability in duration, and coarticulation are reported in this study. This was done to assess the extent to which any developmental decreases in spectral variability were related to changes in other temporal and static spectral measures.

Method

Participants

Thirty children and 10 adults participated. All participants were native, monolingual speakers of English and had no reported history of speech or language disorder or sensorineural hearing loss. All child participants passed a bilateral 20 dB HL pure-tone air-conduction hearing screening at 0.5, 1, 2, and 4 kHz (American National Standards Institute, 1989); all adult participants completed a hearing screening or reported a recent normal audiometric evaluation. Adult participants were students at the University of Minnesota, had a mean age of 24;4 (years;months; *SD* = 2;2), and were naïve to the purposes of the experiment. Children were recruited from a local daycare center and through word of mouth. Children participated in a larger study examining lexical and phonological processing in children; this task was embedded randomly in the larger protocol. Children were compensated with stickers for their participation

Children completed the Goldman–Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000) and the Expressive Vocabulary Test (EVT; Williams, 1997) as screening measures of their speech and language skills. No child received a standard score less than 85, indicating performance greater than 1 *SD* below the mean, on either measure. Children's phonetic repertoires varied. Children had between zero and five phonemes consistently in error. Children's residual errors were either single-feature substitutions (i.e., f/θ, b/v) or /r/ distortions. Children were divided into three equal-sized age groups. The youngest age group had a mean age of 3;11 (*SD* = 0;03), the middle group of children had a mean age of 5;4 (*SD* = 0;06), and the oldest group of children had a mean age of 8;4 (*SD* = 0;08).

Sex distribution, mean GFTA-2 scores, and mean EVT scores for the three groups are presented in Table 1. The fact that the four groups were not matched for sex composition was a consequence of the populations from which the participants were drawn. The younger children were chosen to be typically developing age- and gender-matches for a group of children with phonological impairment. The sex ratios in these younger groups reflect the sex ratios in the population of children with phonological impairment (Shriberg & Austin, 1998). It is notable that the specific asymmetries (more boys in the child groups, more women in the adult group) would exacerbate, rather than attenuate, the age-related differences in speech production, under the presumption that adult-like speech is achieved earlier in girls than in boys.

Stimuli

The stimuli were two sV and four sCV syllables. The vowel was either /a/ or /u/, and the consonant was either /p/ or /w/. All of the two- and three-phoneme sequences were attested in the English lexicon. As in previous research (e.g., Nittrouer et al., 1996) the stimuli included both real words (/su/, /spa/) and nonwords (/sa/, /spu/, /swa/, /swu/). To minimize the potential effect of lexical status on production, all words were presented as the names of unfamiliar cartoon animals. To ensure that children received a consistent model, production prompts were prerecorded by a phonetically trained adult male speaker of American English and were digitized (44.1 kHz sampling rate, 16-bit quantization) for presentation from a computer.

Data Collection

Data collection took place in a quiet room at the child's daycare center or in the author's laboratory. The room at the daycare center was selected in advance because the experimenters determined it to have a low level of ambient noise. During the data collection, participants were told that the computer would show them pictures of made-up animals and would play the animals' names. Production prompts were presented concurrent with a color line drawing of an unfamiliar cartoon animal. Participants were instructed to repeat the word as soon as they heard it. After they repeated the word, participants were presented with a reinforcement picture of a different cartoon animal. The six target stimuli were embedded in a list of 15 stimuli. Five blocks of stimuli were presented to each child; stimuli were randomized within blocks. Two practice items were administered prior to the experiment. Production prompts were output from a pair of powered speakers (Audix PH5-4075; Audix Corp., Wilsonville, OR) at a level of approximately 65 dB SPL. Speech was recorded directly on to the hard drive of a Dell Pentium III laptop, using a table-mounted microphone (Sony ECM 220t) placed approximately 15 cm from the participant's mouth. Speech was digitized at a sampling rate of 22.05 kHz with 16-bit quantization and was low-pass filtered at 11.025 kHz to remove aliasing.

Table 1. Participant characteristics.

	Age (years;months)		GFTA-2		EVT		Sex	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	F	M
Adults	24;4	2;02	—	—	—	—	10	0
Children								
Oldest	8;04	0;08	104	8	114	8	6	4
Middle	5;04	0;06	106	11	112	10	2	8
Youngest	3;11	0;03	118	9	112	14	3	7

Note. GFTA-2 = Goldman–Fristoe Test of Articulation–Second Edition; EVT = Expressive Vocabulary Test.

Analysis

Tokens produced disfluently or with background noise were removed from the analysis. The eliminated tokens were distributed randomly among the targets. At least four nonconsecutive repetitions of each stimulus were collected from each participant, and most participants produced five tokens of each stimulus. The range in number of tokens per participant was from 24 to the full set of 30 (M = 28, SD = 2). Only correctly produced repetitions were analyzed. For the purposes of this study, correct production was defined as any /s/ that was produced without lateral airflow and without the tongue protruding between the upper and lower incisors, as judged by the experimenter's visual inspection of the child during the task. Tokens of /s/ that were produced with the tongue contacting the back of the upper or lower front teeth were considered acceptable, as these variants occur in the population of typical adult speakers of English (e.g., Dart, 1991). These variants occurred in 3 of the child participants.

The Praat (Version 4.0.7) signal-processing software (Boersma & Weenink, 2002) was used for acoustic analyses. The onsets and offsets of /s/ were labeled using the Praat software by the author, and acoustic analyses were completed using scripts that manipulated these files. The beginning of /s/ was defined as the onset of the aperiodic high-frequency noise characteristic of voiceless fricatives. Because the stimuli were uttered in isolation, this coincided with the onset of the response. The offset was defined as the end of this period of aperiodic noise, regardless of whether the fricative was partially voiced or the following vowel was partially voiceless. All of the acoustic measurements were made automatically using the Praat signal-processing program. Reliability was determined by having the author relabel three tokens from each participant (approximately 1% of the total data set) after a delay of approximately 4 months. The mean absolute difference in duration between the two intervals was 12 ms, and it ranged from –20 ms to 22 ms.

Duration

In addition to measures of spectral variability, measures of the mean duration and intraspeaker variability in duration of /s/ were also taken. Previous research (Hawkins, 1979) found that children's correctly produced /s/ clusters have temporal properties that are different from those of adults. Thus, temporal measures were made so that the relationship among spectral variability, mean duration, and variability in duration could be assessed.

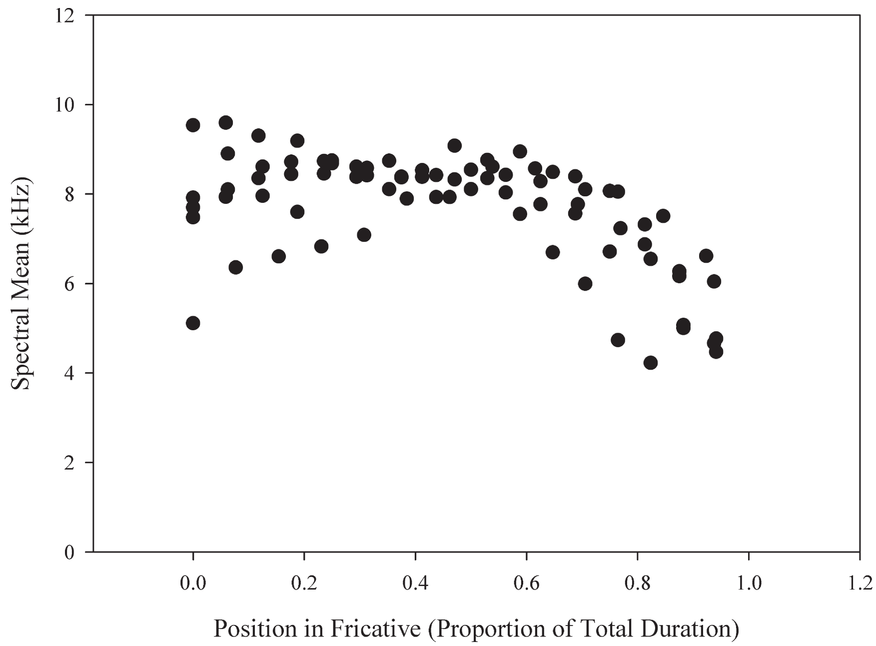
Duration was measured automatically in Praat. The coefficient of variance (the standard deviation divided by the mean) was used as the measure of variability in duration. Despite the limitations of this measure (Weismer, 1991), few alternatives have been presented in the research literature. For example, Tabain (2001) examined variability in acoustic and articulatory measures using the eta-squared test of variance. This test examines whether items in an analysis of variance (ANOVA) have significantly different degrees of variance. This measure could not be used to assess within-subject variability in the current experiment, as this would have required that individual ANOVAs be calculated separately for each participant. This would violate the assumptions of ANOVA. Moreover, only five observations were made per condition, meaning that these individual-participant ANOVAs would lack sufficient statistical power. Mean duration and mean coefficients of variance were calculated separately for each participant's production of the six targets.

Spectral Variability

Spectral variability was measured by examining changes in spectral mean through the course of the fricative. Spectral means were calculated using the method in Forrest et al. (1988), with one exception: A 500 Hz high-pass filter was applied to the spectrum before it was calculated so that values for partially voiced fricatives would be comparable to those of voiceless ones. Partially voiced fricatives represented 2% of the fricatives that were analyzed. For each token of /s/, the spectral mean was calculated for each nonoverlapping 10 ms hamming window of frication noise. Some previous research has used a 20 ms window to calculate fricative spectra (e.g., Forrest et al., 1988). A 10 ms window was used in this experiment so that rapid changes in the spectrum of /s/ could be detected. The spectral means were calculated for the 500–11025 Hz frequency range.

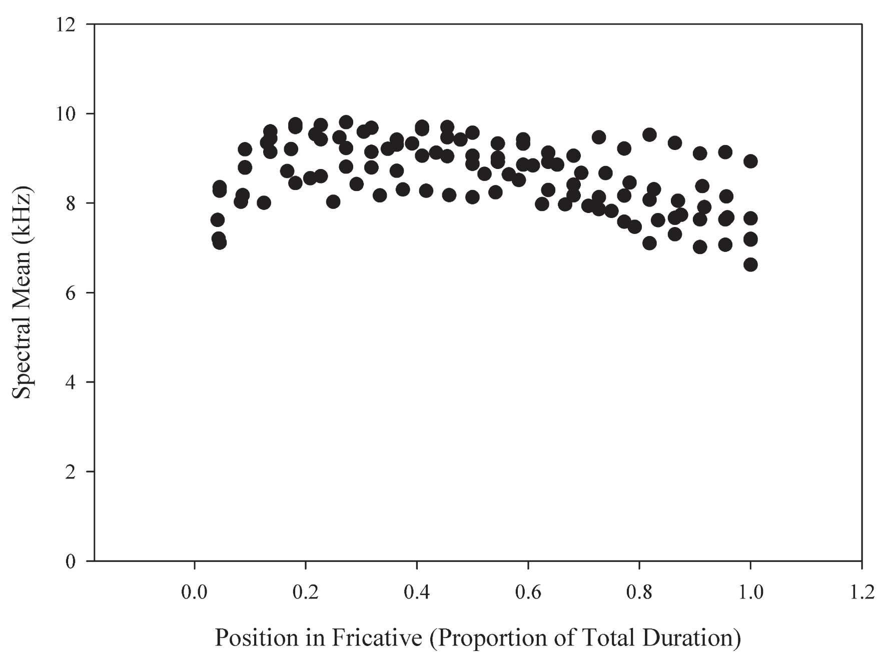
Spectral variability was measured using a modified version of the method presented in Munson (2001), in which regression analyses were used to characterize trial-to-trial variability in spectral mean. Six separate regressions were computed for each phonetic context, for each participant. In the current study, a second-order polynomial function was fitted to the data rather than an exponential function. Pilot analyses indicated that the $y = ae^{bx}$ function was not appropriate for the data from the sV stimuli, in which spectral mean was generally flat throughout the duration of the fricative, as in previous research (Shadle & Mair, 1996). This is shown in Figures 1 and 2. Figure 1 shows the spectral mean for 10 ms windows of frication noise in five productions of /swa/. The spectral means toward the end of the fricative are lower than those in the middle, likely reflecting the acoustic effects of the lip-rounding gesture associated with the upcoming /w/. Figure 2 shows five productions of /sa/. Values for the spectral mean in

Figure 1. Spectral means for 10 ms windows of the fricative /s/ in five repetitions of the sequence /swa/ produced by a male child aged 7;3 (years;months).



this context are generally flatter than those in the /swa/ tokens. Using the $y = ae^{bx}$ function, the fricatives in the sV contexts were estimated as more variable than those in the sCV contexts as an artifact of their unchanging spectra. In addition, it was observed that the R^2 measure used in Munson (2001) was sensitive to the number of data points contained in each regression: Fricatives with shorter durations were estimated as more variable than those with longer durations because they contained fewer data points.

Figure 2. Spectral means for 10 ms windows of the fricative /s/ in five repetitions of the sequence /sa/ produced by an adult female.



In order to estimate the relative variability of the spectral means for each context in the current study, the area occupied by data-point clouds like those illustrated in Figures 1 and 2 was estimated. This was done by summing the absolute values of the residuals of the second-order polynomial fit. This estimate of the area is positively correlated with the number of data points. To correct for this correlation, the estimate was weighted by the ratio of the maximum number of data points among all contexts for each participant to the number of data points for each context. This measure was called the weighted sum of average residuals (henceforth WSAR). Separate mean WSAR values were calculated for each participant's production of /s/ in each of the six contexts.

Results

The four dependent measures made in this experiment were examined separately. For each dependent variable, a three-factor mixed-model ANOVA was used. In each ANOVA, age group was the between-subjects factor; vowel (/a/, /u/) and context (sV, spV, swV) were the within-subjects factors. The alpha level for evaluating main effects and interactions was set at .01 for all analyses. In all analyses, a Huynh–Feldt correction for sphericity (Huynh & Feldt, 1976) was made for the within-subjects factors.

Mean Duration

The first analysis examined mean duration for the fricatives. For each participant, mean duration was calculated in the six contexts. A significant main effect of vowel context on /s/ duration was found. Longer durations were found in syllables containing /u/ (M = 248 ms, SD = 43 ms) than /a/ (M = 232 ms, SD = 43 ms). No effect of context was found, $F(2, 72) = 3.11$, $p > .01$. Vowels were produced with similar durations in the three syllable types (M = 249 ms, SD = 52 ms for sV; M = 233 ms, SD = 47 ms for spV, M = 238 ms, SD = 43 ms for swV). Finally, contrary to previous research, no effect of age was found, $F(3, 36) = 1.9$, $p > .01$. Fricatives were produced with similar duration by the youngest children (M = 230 ms, SD = 43 ms), by children in the middle group (M = 260 ms, SD = 53 ms), by the oldest children (M = 249 ms, SD = 36 ms), and by adults (M = 222 ms, SD = 20 ms). Notably, more variability between subjects was found in the three groups of children than in the adult group. None of the interactions was significant.

Variability in Duration

The second analysis examined intraspeaker variability in duration of /s/ in the six contexts. Table 2 shows mean coefficients of variance for /s/ in the six contexts spoken by the four groups. A significant main effect of age was found, $F(3, 36) = 8.0$, $p < .01$, partial $\eta^2 = 0.40$. Post hoc Tukey least significant difference (LSD) tests indicated that adults produced fricatives with less within-speaker variability in duration than all three groups of children, who did not differ from one another (M = 0.239, SD = 0.065 for the youngest children; M = 0.213, SD = 0.077 for the middle group of children; M = 0.212, SD = 0.010 for the oldest group of children; and M = 0.094, SD = 0.020 for adults). No effect of vowel was found, $F(1, 36) = 1.6$, $p > .01$. Fricatives were equally variable before the vowels /a/ (M = 0.18, SD = 0.09) and /u/ (M = 0.20, SD = 0.11). Finally, no effect of context was found, $F(2, 72) = 1.1$, $p > .01$. Fricatives demonstrated similar levels of variability in the three contexts (M = 0.19, SD = 0.12 for sV; M = 0.21, SD = 0.11 for spV; and M = 0.19, SD = 0.11 for swV). None of the interactions was significant.

Spectral Variability

For each participant, WSAR values were calculated separately for each of the six contexts. Table 3 shows mean WSAR values for the four groups' productions of /s/. A

Table 2. Mean intraspeaker temporal variability (coefficient of variance) for /s/ in sV and sCV contexts.

				Children					
		Adults		Oldest		Middle		Youngest	
Vowel	Consonant	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
a	(none)	.071	.042	.154	.064	.269	.213	.221	.122
	p	.118	.036	.146	.059	.235	.092	.229	.141
	w	.094	.035	.220	.189	.171	.071	.257	.135
u	(none)	.096	.034	.219	.236	.198	.077	.264	.142
	p	.105	.053	.297	.227	.237	.173	.276	.143
	w	.082	.033	.236	.149	.167	.061	.190	.095

Table 3. Mean spectral variability (WSAR) for /s/ in sV and sCV contexts.

		Adults		Children					
				Oldest		Middle		Youngest	
Vowel	Consonant	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
a	(none)	42.5	13.7	72.9	20.9	102.6	44.0	84.5	39.9
	p	44.7	19.5	86.6	20.7	104.5	40.0	80.9	27.4
	w	62.1	16.6	95.5	26.9	108.8	39.2	100.6	39.7
u	(none)	54.1	16.5	73.1	33.4	82.2	23.7	87.8	43.2
	p	52.2	18.5	103.9	25.6	97.8	32.2	84.4	31.8
	w	62.0	19.0	95.0	18.8	118.5	44.7	107.8	45.2

Note. WSAR = weighted sum of average residuals.

significant main effect of age was found, $F(3, 36) = 7.5$, $p < .01$, partial $\eta^2 = 0.39$. Post hoc Tukey LSD tests indicated that the adults ($M = 53$, $SD = 14$) produced fricatives with less spectral variability than the three groups of children, who did not differ from one another ($M = 90$, $SD = 33$ for the youngest children; $M = 102$, $SD = 31$ for the middle group of children; and $M = 88$, $SD = 15$ for the oldest group of children). A significant main effect of context was found, $F(2, 72) = 11.7$, $p < .01$, partial $\eta^2 = 0.25$. Post hoc Tukey LSD tests indicated that fricatives were more variable in the swV context ($M = 94$, $SD = 36$) than either the sV ($M = 75$, $SD = 35$) or the spV ($M = 81$, $SD = 32$) contexts, which did not differ from one another. No effect of vowel was found, $F(1, 36) = 1.7$, $p > .01$. Fricatives were produced with similar levels of spectral variability before the vowels /a/ ($M = 85$, $SD = 30$) and /u/ ($M = 82$, $SD = 32$). None of the interactions was significant.

Coarticulation

The absence of an effect of vowel on spectral variability in the sV contexts was contrary to predictions. It is hypothesized that the WSAR measure of spectral variability relates to variability in coordination. The absence of an effect of vowel on spectral variability may have been due to a lack of overlap between the lip-rounding gestures associated with /w/ and /u/ and the tongue-tip raising gesture associated with /s/. Although coarticulation was not the focus of this analysis, a post hoc analysis of coarticulation in the six target contexts was completed to determine whether patterns of spectral variability were related to patterns of coarticulation. In these analyses, spectral means were computed for the 40 ms window of frication noise centered three-quarters of the way through the fricative. A larger window size was used for this analysis than was used in the analysis of spectral variability. This analysis did not need to detect rapid changes in the spectral characteristics of the fricatives; hence, a larger sized window could be used. The three-quarter point was used so that the influence of sounds immediately adjacent to the fricative could be detected. For each participant, the average spectral mean was calculated for /s/ separately in each of the six phonetic contexts.

Two significant effects were found. First, a significant main effect of context was found, $F(2, 72) = 67.5$, $p < .01$, partial $\eta^2 = 0.65$. The fricative /s/ was produced with a lower spectral mean in the swV context ($M = 5745$ Hz, $SD = 1072$ Hz) than in the sV context ($M = 7018$ Hz, $SD = 1147$ Hz). Values for the spV context were intermediate ($M = 6618$ Hz, $SD = 1077$ Hz). All post hoc Tukey LSD comparisons were significant. A significant effect of vowel was also found, $F(1, 36) = 18.4$, $p < .01$, partial $\eta^2 = 0.34$. As in previous research (e.g., Nittrouer et al., 1996), the spectral mean of /s/ was higher when followed by /a/ than when followed by /u/ when pooled across syllable types ($M = 6329$ Hz, $SD = 971$ Hz for /s/ before /u/; $M = 6591$ Hz, $SD = 1099$ Hz for /s/ before /a/). Contrary to previous research, no effect of age was found on spectral means, $F(3, 36) < 1$, $p > .01$. This is very likely attributable to the different sex ratios for the groups of participants: All of the male participants were in the three groups of children, and all of the adult participants were female. No interactions were found. The lack of an Age × Vowel interaction stands in contrast to previous research (e.g., Nittrouer et al., 1996), which found a larger effect of vowel context on fricative second-formant frequency in children than adults. This finding is, however, consistent with Katz et al.'s (1991) and Sereno et al.'s (1987) finding that adults and children show similar degrees of coarticulation.

Sex Differences

The three groups of child participants were not matched for gender. Participants in the three groups of children were chosen to be typically developing same-age matches for children with phonological impairment who were tested in a separate study. There is a higher incidence of phonological impairment in boys than in girls (Shriberg & Austin, 1998); hence, there were more boys than girls in the child groups. In addition, the ratios were different for the different groups, although this difference did not achieve significance in a Pearson's chi-square test, $\chi^2(2, N = 30) = 3.7$, $p > .10$. Previous research has found that girls master speech-sound production accuracy earlier than boys (e.g., Smit, Hand, Freilinger, Bernthal, & Bird, 1990), suggesting that these asymmetries might confound the interpretation of group differences.

To test this possibility, a second series of analyses was completed, examining whether sex affected each of the four dependent measures. Each of these analyses was a four-factor mixed-model ANOVA. As in the other ANOVAs, vowel and context were the within-subjects factors. Age and sex were the between-subjects measures. Only the three groups of children were included in this analysis. An alpha level of .01 was used to evaluate the significance of individual factors and interactions. Using this criterion, there was no significant main effect of sex on any of the four dependent measures. In addition, sex did not interact significantly with any of the other factors. Thus, the asymmetries in sex composition of the groups do not appear to limit the interpretation of the age-related effects, both in terms of the age-related differences in variability and the lack of a developmental effect on mean duration and coarticulation.

Discussion

Hypotheses Revisited

The first objective of this article was to examine whether children demonstrate more spectral variability in /s/ than adults. Children's production of /s/ in both sV and sCV sequences in this study demonstrated significantly more spectral variability than adults' productions, when the WSAR measure is used to characterize variability. In addition, children produced speech with more temporal variability than adults. These age-related differences do not appear to be related to the duration of the sounds, as no significant effect of age was found on /s/ duration in any of the contexts studied. Moreover, these differences do not appear to be related to differences in the extent of coarticulation. Although all three groups of children produced greater spectral variability than adults, the effect of the following vowel and consonant on /s/ spectra was similar for the four age groups studied. This finding stands in contrast to some previous research, which has found a larger effect of vowel context on spectral characteristics of children's /s/ productions than those made by adults (e.g., Nittrouer et al., 1996). One potential interpretation of this null finding is that the current study used relatively small participant groups ($n = 10$). However, other studies have found age-related differences in coarticulation using comparably sized participant groups. Thus, although statistical power in the current study was certainly not optimal, the differences between the current study and previous research are not likely to be due to differences in statistical power.

In contrast, patterns of spectral and temporal variability were similar to one another. When pooled across the four age groups, significant correlations were found between coefficients of variance for /s/ duration and WSAR measures ($r^2 = .26$, $p < .01$ for sV; $r^2 = .16$, $p = .01$ for spV; $r^2 = .16$, $p = .01$ for swV). This finding is consistent with previous literature on children's spectral and temporal variability (e.g., Kent & Forner, 1980; Lee et al., 1999). The findings regarding duration and coarticulation are not comparable to some previous findings: Many previous studies have found age-related declines in duration (e.g., Kent & Forner, 1980) and coarticulation (e.g., Nittrouer et al., 1996). This may be attributable to differences in the complexity of the task used to elicit /s/ in the current experiment and in previous experiments. Many previous studies have used picture-naming tasks to elicit /s/ (e.g., Nittrouer et al., 1996). The current study used a repetition task. Picture-naming tasks, particularly ones in which a novel picture is paired with a novel word-shape, require talkers to complete a number of cognitive processes, including learning a picture–word-shape association, accessing a newly learned word from memory when confronted with the picture, and forming and executing a response. The lack of developmental effects on duration, and the lack of a consistent developmental decrease in coarticulation may be due to the relative ease of the repetition task. Future research on this topic should examine the influence of task complexity on developmental differences in speech production.

The second objective of this study was to examine whether spectral variability in /s/ was dependent on phonetic context. These predictions were not supported consistently. Contrary to predictions, spectral variability of /s/ was not consistently greater in the /u/ context than in the /a/ context. Consistent with predictions, however, the highest degree of spectral variability was noted in the swV context. It was predicted that /s/ would show greater variability in this context than in the sV or spV contexts, as variability in the onset of the lip-rounding movement for /w/ would lead to greater spectral variability in /s/. The lack of a consistent effect of phonetic context suggests that the differences in spectral variability between adults and children may be due to variability in the production of the /s/ phoneme. Although early research claimed that subtle differences in the place of articulation of /s/ would not affect its acoustics (Stevens, 1989), more recent research has found significant correlations between static articulatory measures and static acoustic measures (Tabain, 2001). This suggests that the acoustics of /s/ may be influenced by subtle variations in its place of articulation. Thus, at least some of the spectral variability captured by the WSAR measure may be due to trial-to-trial variability in the place of articulation of /s/. Following Munhall (1989), this is referred to this as *point variability*, which can be distinguished from variability in coordination. An example of point variability can be seen in Figure 2, which shows an adult participant's production of /s/ in five tokens of the sequence /sa/. This shows that the fricative /s/ was produced slightly differently each time it was uttered, even though this is a context in which only a minimal influence of phonetic context on articulatory or acoustic characteristics of /s/ would be expected. Further research examining concurrent kinematic and acoustic measures of variability is needed to quantify the relative contribution of point variability and variability in coordination on the variability of /s/ spectra in adults and children.

General Discussion

In summary, /s/ is produced with greater spectral variability by children than by adults. In addition, it is produced with greater variability in swV sequences than in spV or sV sequences. The influences of phonetic context and syllable type are similar for the four age groups that participated in this experiment. This study expands on previous studies of variability in children's speech by showing that children produce the fricative /s/ with greater spectral variability than adults. The finding that even children in the oldest age group have not achieved adult-like measures of spectral variability in /s/ is consistent with previous studies on spectral variability in vowels (Lee et al., 1999), in which adult-like values for some spectral measures were not found until the teenage years.

As with previous studies on variability in children's speech, there are a number of potential interpretations of this finding. The presumption in many previous studies is that the observable differences between children and adults are the result of a lack of skill in children's speech production. Although these studies may differ in whether they posit that differences between children and adults are due to neuromotor immaturity, inexperience in producing sounds, or differences in the level of detail in phonological representation, the general consensus is that developmental differences are an indication of the immaturity of children's speech production.

In contrast, some recent studies have emphasized that the parameters observed to be "deficient" in children's speech, such as duration and coarticulation, are modified by normal adult speakers to maximize their speech intelligibility. For example, research has shown that coarticulation is both planned (Whalen, 1990) and language-specific (Manuel, 1990). Some researchers have claimed that coarticulation may have a functional benefit (Bradlow, 2002) in that it serves to extend a feature in time and may make perceptually vulnerable features more robustly perceptible by virtue of their being longer. Thus, the extended coarticulation seen in some children's speech may be the result of children consciously attempting to maximize the intelligibility of their speech.

Similarly, adults actively manipulate duration to improve speech intelligibility. Again, the slower rate of children's speech in previous studies may be due to an active attempt to maximize intelligibility. Future research on children's speech development should examine whether children's non-adult-like production of coarticulation and duration, as well as the variability in these measures, are due truly to immaturity in the control of these parameters or to children's conscious efforts to maximize the intelligibility of their developing speech. Goffman, Schwartz, and Marton (1996) found that children produced words representing contextually new information more accurately than words representing contextually old information. Future research might extend this line of inquiry and examine the spectral characteristics of speech produced in predictable and unpredictable contexts. If children were utilizing duration and coarticulation to maximize the intelligibility of their developing speech, then it would be predicted that these properties would be exaggerated in the unpredictable contexts. Results from such studies could provide important information on the locus of developmental differences in children's speech production.

Acknowledgments

This research was supported by a Grant-in-Aid for Research, Scholarship, and Artistry to the author from the University of Minnesota Graduate School, and by laboratory set-up funds from the University of Minnesota College of Liberal Arts. I acknowledge the University of Minnesota Child Care Center for allowing collection of data at their facility. I gratefully acknowledge Cyndie Swenson and Shayla Manthei for assistance in data collection, and Edward Carney for helping to develop to the WSAR measure of spectral variability. Portions of this work were presented at the spring 2002 meeting of the Acoustical Society of America. I acknowledge participants at that conference for their helpful feedback. As always, generous acknowledgment is given to Pauline Welby for assistance in writing the Praat scripts to automate data analyses.

References

American National Standards Institute. (1989). *Specifications for audiometers*. Washington, DC: Author.

Boersma, P., & Weenink, D. (2002). Praat 4.0.7 [Computer software]. Amsterdam: Institute of Phonetic Sciences.

Bradlow, A. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven, T. Rietveld, & N. Warner (Eds.), *Papers in laboratory phonology VII* (pp. 241–273). Cambridge, U.K.: Cambridge University Press.

Dart, S. (1991). *Articulatory and acoustic properties of apical and laminal articulations*. Unpublished doctoral dissertation, Department of Linguistics, University of California, Los Angeles.

Eguchi, S., & Hirsch, I. (1969). Development of speech sounds in children. *Acta Otolaryngologica Supplementum, 257*, 1–51.

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

Goffman, L., Schwartz, R., & Marton, K. (1996). Information level and young children's phonological accuracy. *Journal of Child Language, 23*, 337–347.

Goldman, M., & Fristoe, R. (2000). *The Goldman–Fristoe Test of Articulation–Second Edition*. Circle Pines, MN: American Guidance Service.

Hawkins, S. (1979). Temporal co-ordination of consonants in the speech of children: Further data. *Journal of Phonetics, 7*, 235–267.

Huynh, H., & Feldt, L. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69–82.

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

Katz, W., Kripke, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of Speech and Hearing Research, 34*, 1222–1232.

Kent, R., & Forner, L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics, 8*, 157–168.

Lee, S., Potamianos, A., & Naryanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America, 105*, 1455–1468.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America, 88,* 1286–1298.

Munhall, K. (1989). Articulatory variability. In P. Square-Storer (Ed.), *Acquired apraxia of speech in aphasic adults* (pp. 64–83). Hillsdale, NJ: Erlbaum.

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

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., Studdert-Kennedy, M., & McGowan, R. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32*, 120–132.

Nittrouer, S., Studdert-Kennedy, M., & Neely, S. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech, Language, and Hearing Research, 39*, 379–389.

Sereno, J., Baum, S., Marean, G., & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *Journal of the Acoustical Society of America, 81*, 512–519.

Shadle, C. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 187–209). Dordrecht, The Netherlands: Kluwer.

Shadle, C., & Mair, S. (1996). Quantifying spectral characteristics of fricatives. *Proceedings of the International Conference on Speech and Language Processing (ICSLP), 3,* 1521–1524.

Shriberg, L., & Austin, D. (1998). Co-morbidity of speech-language disorder: Implications for a phenotype marker for speech delay. In R. Paul (Ed.), *Exploring the speech-language connection* (pp. 73–117). Baltimore: Paul H. Brookes.

Smit, A., Hand, L., Freilinger, J., Bernthal, J., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779–798.

Smith, B. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics, 6*, 37–67.

Smith, B., Kenney, M. K., & Hussain, S. (1996). A longitudinal study of duration and temporal variability in children's speech production. *Journal of the Acoustical Society of America, 99*, 2344–2349.

Stathopoulos, E. (1995). Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech. *Journal of Phonetics, 23*, 67–80.

Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3–45.

Stevens, K. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production. *Language and Speech, 44*, 57–94.

Weismer, G. (1991). Assessment of articulatory timing. In J. Cooper (Ed.), *Assessment of speech and voice production:*

Research and clinical applications (NIDCD Monograph, Vol. 1, pp. 84–95). Bethesda, MD: National Institutes of Health.

Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics, 18*, 3–35.

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Received February 15, 2003

Accepted June 4, 2003

DOI: 10.1044/1092-4388(2004/006)

Contact author: Benjamin Munson, PhD, Department of Communication Disorders, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive, SE, Minneapolis, MN 55455. E-mail: munso005@umn.edu