JSLHR

Article

Toward Improved Spectral Measures of /s/: Results From Adolescents

Laura L. Koenig,[a,b] Christine H. Shadle,[a] Jonathan L. Preston,[a,c] and Christine R. Mooshammer[a]

Purpose: This article introduces theoretically driven acoustic measures of /s/ that reflect aerodynamic and articulatory conditions. The measures were evaluated by assessing whether they revealed expected changes over time and labiality effects, along with possible gender differences suggested by past work.
Method: Productions of /s/ were extracted from various speaking tasks from typically speaking adolescents (6 boys, 6 girls). Measures were made of relative spectral energies in low- (550–3000 Hz), mid- (3000–7000 Hz), and high-frequency regions (7000–11025 Hz); the mid-frequency amplitude peak; and temporal changes in these parameters. Spectral moments were also obtained to permit comparison with existing work.
Results: Spectral balance measures in low–mid and mid–high frequency bands varied over the time course of /s/, capturing the development of sibilance at mid-fricative along with showing some effects of gender and labiality. The mid-frequency spectral peak was significantly higher in nonlabial contexts, and in girls. Temporal variation in the mid-frequency peak differentiated ±labial contexts while normalizing over gender.
Conclusions: The measures showed expected patterns, supporting their validity. Comparison of these data with studies of adults suggests some developmental patterns that call for further study. The measures may also serve to differentiate some cases of typical and misarticulated /s/.

Key Words: acoustics, adolescents, development, speech production

Fricative spectra vary in complex ways that depend on articulatory and aerodynamic conditions. In this article, we present a set of automated measures for characterizing the spectrum of /s/ that reflect these underlying physical conditions. Our ultimate goal is to expand the battery of methods available for describing fricatives in adults, children, and clinical populations. The measures are based on past work with adult speakers and using mechanical models; for the current study, we adapted them for data from adolescent children taken in naturalistic recording environments. The proposed measures were primarily designed to account for patterns of coarticulation and the development of sibilance over time. Many past studies have quantified the spectral features of /s/ and other fricatives using spectral moments (Forrest, Weismer, Milenkovic, & Dougall, 1988), and we include moments among our measures to permit comparison with this literature. We will argue, however, that the measures introduced here offer some advantages over moments-based analyses.

Our focus on the sibilant /s/ is based on several considerations. First, /s/ has been widely investigated, so its acoustic features have been more fully described than those of many other fricatives. Indeed, a number of studies have, like the current work, exclusively evaluated /s/ (e.g., Boothroyd & Medwetsky, 1992; Daniloff, Wilcox, & Stephens, 1980; Flipsen, Shriberg, Weismer, Karlsson, & McSweeny, 1999a; Iskarous, Shadle, & Proctor, 2011; Karlsson, Shriberg, Flipsen, & McSweeny, 2002; Katz, Kripke, & Tallal, 1991; Munson, 2004; Shadle & Scully, 1995; Weismer & Elbert, 1982). Second, the age of acquiring perceptually accurate production of the sibilants /s z ʃ ʒ/ is quite variable across typically developing children, and on average it occurs later than many other sounds (Gruber, 1999; Sander, 1972; Smit, Hand, Feininger, Bertha, & Bird, 1990). The sibilants (along with the liquids) are also among the most common sounds to be misarticulated beyond the usual age of acquisition (Shriberg, 1994, 2009). These observations have motivated much of the past developmental and clinical work on /s/; they also suggest that /s/, as a late or challenging consonant, can provide a useful window into the extended time course of speech motor development. A thorough characterization of the acoustics of /s/ in development could improve our understanding of how children learn to produce this sound. For

[a]Haskins Laboratories, New Haven, Connecticut
[b]Long Island University, Brooklyn, New York
[c]Southern Connecticut State University, New Haven
Correspondence to Laura L. Koenig: koenig@haskins.yale.edu

Editor: Jody Kreiman
Associate Editor: Scott Thomson
Received January 27, 2012
Revision received July 7, 2012
Accepted November 25, 2012
DOI: 10.1044/1092-4388(2012/12-0038)

clinical application, acoustic descriptions are most useful if they allow inferences about the underlying aerodynamics and articulation, that is, the physical characteristics of the production. As discussed below, spectral moments are often ambiguous in this regard. Useful acoustic descriptions should also differentiate perceptually acceptable fricatives from misarticulations. This would provide an objective and reliable metric of misarticulation, and it could help in documenting improvements during a course of therapy. The current article evaluates typically speaking adolescents, but the methods proposed here, which include measures designed to capture sibilant "goodness," may ultimately have clinical utility.

Acoustic Characterization of Fricatives

Many studies of both adults and children have sought to differentiate fricatives across places of articulation. The results have indicated that these distinctions are reflected in the amplitude, frequencies, and duration of the fricative noise, and fricative-vowel formant transitions (e.g., Baum & McNutt, 1990; Behrens & Blumstein, 1988; Fox & Nissen, 2005; Jassem, 1965; Jongman, Wayland, & Wong, 2000; Maniwa, Jongman, & Wade, 2009; Nissen & Fox, 2005; Pentz, Gilbert, & Zawadzki, 1979; Soli, 1981; Strevens, 1960). Most research on sibilants has focused on the spectrum of the frication noise, which has been found to carry strong perceptual cues for discriminating among these sounds (e.g., Newman, Clouse, & Burnham, 2001; Whalen, 1991; Yeni-Komshian & Soli, 1981).

A common technique for measuring fricative spectra has been to identify, typically by eye, *spectral peaks,* that is, frequencies with high noise amplitudes (Behrens & Blumstein, 1988; Bladon & Seitz, 1986; Hughes & Halle, 1956; Iskarous et al., 2011; Jassem, 1965; Jongman et al., 2000; Maniwa et al., 2009; Pentz et al., 1979; Seitz, Bladon, & Watson, 1987; Shadle & Mair, 1996; Soli, 1981; Strevens, 1960; Yeni-Komshian & Soli, 1981). Some authors have also estimated the low-frequency bound of the region with high-amplitude noise (Bladon & Seitz, 1986; Jassem, 1965). Summarizing across these studies and measures, /s/ in adults has been described as having its lowest-frequency spectral peaks between 3.5 and 5 kHz, and most acoustic energy above 4 kHz (Behrens & Blumstein, 1988; Hughes & Halle, 1956; Jassem, 1965; Jesus & Shadle, 2002; Soli, 1981; Strevens, 1960; Yeni-Komshian & Soli, 1981).[1]

Forrest et al. (1988) described a quantitative approach to characterizing voiceless obstruents based on spectral characteristics, focusing on whether spectral moments could serve to classify obstruent place of articulation automatically.[2] Spectra of fricatives and stop bursts were treated as random probability distributions, and the first four moments (M1, M2, L3, L4) of the distributions were calculated.[3] These four moments represent, in turn, the mean (sometimes called the centroid or center of gravity), standard deviation, skewness, and kurtosis. Forrest et al. stated that M2 did not contribute to differentiating among the stops /p t k/ or fricatives /f θ s ʃ/, and only reported data on M1, L3, and L4; many subsequent studies likewise only reported data on a subset of moments, frequently just M1. For the fricatives, Forrest et al. observed that moments did not serve to separate nonsibilants well, but that /s/ and /ʃ/ could be discriminated fairly well based on the first 20 ms of the fricative noise (83% for moments calculated on a linear scale and 98% on a Bark scale). Discriminant analyses suggested that L3 was the primary factor differentiating between these two sounds.

The work of Forrest et al. (1988) was exploratory. Their database included 10 speakers, but the speech material was restricted to six repetitions per speaker of 14 words; five of these words contained fricatives, and results for sibilants were based on the syllables /si/ and /ʃi/. These authors also indicated that moments did not classify place of articulation as well for fricatives as for stops. Nevertheless, many researchers have gone on to apply these methods to fricatives across many places of articulation (and, as described below, have used them to evaluate coarticulatory patterns and differences across speaker populations). The studies using moments for classification purposes have generally supported the claim that /s/ and /ʃ/ can be differentiated by spectral moments, although results are more consistent for M1 than L3. Specifically, /s/ is reported to have a higher M1 than /ʃ/ (Jongman et al., 2000; Nissen & Fox, 2005; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Shadle & Mair, 1996; Tjaden & Turner, 1997). Most studies have found that /s/ has negative L3, that is, the energy is concentrated in high frequencies (Jongman et al., 2000; Nissen & Fox, 2005; Nittrouer, 1995; Shadle & Mair, 1996), but a few report the opposite (Avery

[1]Several studies have also assessed spectral slope over various regions (Bladon & Seitz, 1986; Evers, Reetz, & Lahiri, 1998; Fox & Nissen, 2005; Jesus & Shadle, 2002, Maniwa et al., 2009; Nissen & Fox, 2005; Seitz et al., 1987; Shadle & Mair, 1996), and some such measures appear to be useful for characterizing fricatives. However, the diversity of frequency ranges over which slope was calculated in these studies does not yield a simple summary, so we do not review them here.

[2]Although Forrest et al. (1988) used moments to classify obstruents across place of articulation, it is evident that these authors did not believe that the utility of moments was restricted to automatic classification; they specifically pointed to the possibility (p. 116) that moments might provide insight into the difference between correct and misarticulated fricatives.

[3]Some authors prior to Forrest et al. also computed fricative moments (e.g., Miller, Mathews, & Hughes, 1962; Strevens, 1960). Abbreviations used for the four spectral moments have varied across studies. Forrest et al. used L1–L4 to indicate the first four moments calculated on a linear scale, and then further defined l3 and l4 to represent the coefficients of skewness and kurtosis, respectively. Subsequent work has generally used M1 and M2 for the mean and standard deviation, and some authors also use M3 and M4 for l3 and l4. For this work, to maintain consistency, we will refer to the four moments as M1, M2, L3, and L4, basically corresponding to L1, L2, l3, and l4 of Forrest et al. (with the variations noted in the text regarding cutoff frequencies, preemphasis, and the use of multitaper spectra).

& Liss, 1996; Tjaden & Turner, 1997). M2 appears to differ mainly between sibilants and nonsibilants, with high M2 values in nonsibilants reflecting a broad noise spectrum (Jongman et al., 2000; Nissen & Fox, 2005; Shadle & Mair, 1996). Few authors report data on L4. Jongman et al. (2000) found that kurtosis was high (i.e., the spectrum was rather "peaked") for both /s z/ and /f v/, but flatter for /ʃ ʒ/ and /ð θ/, whereas Nissen and Fox (2005) obtained higher kurtosis for /ʃ/ compared to /f/, /θ/, and /s/. As discussed in the next section (see also Flipsen et al., 1999a; Forrest et al., 1988), some discrepancies across studies may result from differences in analysis procedures (e.g., sampling rate and whether or not spectral averaging was used).

Methodological Considerations

In an early article on fricative acoustics, Strevens (1960) observed that some speakers could have considerable acoustic energy above 8 kHz for /s/. Sampling frequencies of 16 kHz or less provide no information on this high-frequency content. As such, much past work may not have used an analysis range capable of adequately capturing the high-frequency energy of /s/, particularly for child speakers. Also, many authors have made measures from single spectral slices rather than averaging over tokens or time windows. Because of the random variations inherent in frication noise, single spectral slices—and measures taken from them—are associated with a high degree of error (Bendat & Piersol, 2000). Finally, the use of preemphasis and the presence of background noise can affect measures of peak frequencies, spectral tilt, and spectral moments. Some inconsistencies observed in the results of past research may reflect differences in these methodological factors.

Spectral moments can be obtained automatically, and they yield a small set of dependent measures to characterize the fricative noise. They have been found to differentiate /s-ʃ/ on average and to differ with clear speech and gender (Fox & Nissen, 2005; Maniwa et al., 2009). However, several considerations indicate that results of moments-based analyses can be difficult to interpret unambiguously. In particular, spectral moments do not permit straightforward interpretation in terms of either source or filter effects. For example, shifting the place of /s/ articulation posteriorly could lower M1 in two ways. The increased size of the downstream cavity would lower the frequencies of the front-cavity resonances (a filter effect), but moving the constriction further from the teeth would also weaken the fricative noise that excites those resonances (a source effect; Shadle, 1985, 1990, 1991). M1 can also be increased in multiple ways: Changes in speaking effort (a source effect) may increase M1 by yielding more energy at high frequencies (Shadle & Mair, 1996), but reducing lip rounding can also increase M1 by raising the frequency of the main resonance peak (a filter effect). Because M1 and L3 can be strongly correlated (Blacklock, 2004; Newman, 2003; Newman et al., 2001), the same ambiguity holds for L3 as for M1. Finally, M1 cannot by definition be any higher than half the sampling rate, and it varies considerably as sampling rate is changed (Shadle & Mair, 1996); this can complicate comparisons of moments across studies.

Given these issues with moments-based analyses, the primary goal of the present work was to develop theoretically justified measures of /s/ that would reflect articulatory and aerodynamic conditions, including context effects and changes over time consistent with how turbulent noise is generated. We also evaluated the possibility that the measures would show gender differences.

Temporal, Context, and Speaker Effects in /s/

Temporal variation. Results of several past studies indicate that spectral characteristics can vary considerably over the time course of a fricative (Iskarous et al., 2011; Jesus & Shadle, 2002; Jongman et al., 2000; Munson, 2001; Nissen & Fox, 2005; Shadle & Mair, 1996). Along with coarticulatory patterns, such changes may relate to the development of frication noise. During the time-course of a well-produced voiceless fricative, the spectral balance should change. As the minimum constriction area is achieved and the vocal folds open, low-frequency resonances should be cancelled and the noise source should increase in amplitude for frequencies above about 3 kHz (Jesus & Shadle, 2002; Shadle, 1985, 2012; Shadle & Mair, 1996). The increase in air particle velocity resulting from constriction formation and higher airflow in voiceless fricatives should also yield a stronger balance of high- compared with mid-frequency energy at midpoint (Jesus & Shadle, 2002). These time-varying features of noise generation have been relatively neglected in studies of fricatives, and they have not been considered at all in developmental or clinical studies. The current work includes two measures designed to capture this kind of temporal variation.

Coarticulation. There is an extensive literature on anticipatory coarticulation in fricatives, particularly regarding labialization, usually due to rounded vowels, but sometimes also to labial consonants. The lip approximation involved in both rounded vowels and labial consonants lowers the frequencies of the fricative noise (e.g., Heinz & Stevens, 1961; Munson, 2004). Labial contexts additionally have a reliable lowering effect on the second formant frequency (F2) at voicing onset for vowels following the fricative (Jongman et al., 2000; Katz et al., 1991; Maniwa et al., 2009; Nittrouer et al., 1989). F2 differences can also be observed in the fricative noise itself, particularly (but not only) in transitions to the more open postures of the flanking vowels (Jassem, 1965; McGowan & Nittrouer, 1988; Nittrouer et al., 1989; Soli, 1981; Yeni-Komshian & Soli, 1981). Finally, the peak-amplitude frequency tends to be lower for /s/ in the context of /u/ compared to unrounded vowels like /i/ and /a/ (Jongman et al., 2000; Yeni-Komshian & Soli, 1981), although when /s/ is whistly, as often occurs in rounded contexts, the opposite may occur (Shadle & Scully, 1995).

Studies using moments to evaluate anticipatory coarticulation in children and adults generally report that M1 is lowered by labiality (Katz et al., 1991; Nittrouer, 1995; Nittrouer et al., 1989). Nittrouer (1995) found no significant effects of rounded vowels for L3 or L4 in children or adults. In contrast, Shadle and Mair (1996) observed minimal change in M1 in an u_u context for one of their two adult speakers, but M2 was lower and L4 was greatly increased at /s/ midpoint. These M2 and L4 values are consistent with the speaker producing whistly /s/ in rounded contexts (cf. Shadle & Scully, 1995). This example shows how changes in a particular moment can reflect multiple articulatory or aerodynamic causes.

Many developmental studies have asked whether the degree or extent of anticipatory coarticulation is comparable in children and adults. McGowan and Nittrouer (1988) and Nittrouer et al. (1989) reported greater effects of following vowel on fricative and vowel-onset F2s in children than adults, but other authors have found no Age × Vowel context interactions in fricative F2s or centroids (Katz et al., 1991; Munson, 2004; Sereno, Baum, Marean, & Lieberman, 1987). The cited studies considered children as old as age 8. It is not clear at what age any developmental differences might disappear. The work of Flipsen et al. (1999a) on speakers 9–15 years of age did not explicitly compare labial–nonlabial contexts, nor child–adult coarticulatory patterns. For the present study, our expectation was simply that adolescents would show coarticulatory patterns in the same direction as adults. However, given considerable evidence that token-to-token variability remains high in children compared to adults well into the school-age years (e.g., Koenig, Lucero, & Perlman, 2008; Munson, 2004; Sereno et al., 1987; Walsh & Smith, 2002), adolescent coarticulatory patterns may be more variable than those reported for adults.

Gender. Studies of adults have shown that /s/ has higher frequency M1s and spectral peaks in women compared to men (Jongman et al., 2000; Maniwa et al., 2009; Nittrouer et al., 1989; Schwartz, 1968; Yeni-Komshian & Soli, 1981). Higher frequencies in women could arise from smaller vocal tract sizes and/or from articulatory differences such as the degree of lip retraction (Avery & Liss, 1996) or a more anteriorly placed constriction (Fuchs & Toda, 2010). Jongman et al. (2000) also reported gender differences in M2 (greater in women than men), L3 (lower in women than men; see also Fuchs & Toda, 2010), L4 (greater in women than men), and duration relative to the syllable (lower in women than men). Some authors have found similar patterns in children. Fox and Nissen (2005) reported gender differences in adults and children 6–14 years of age in measures of spectral peak, spectral slope, M1, L3, and L4; Nissen and Fox (2005) observed gender differences for 3–5-year-olds in spectral slope; and Flipsen et al. (1999a) found reliable gender differences in speakers 9–15 years of age for M1 and L3 at fricative midpoint and for M2 at onset and offset. However, Pentz et al. (1979) did not find gender differences in spectral peak frequencies for 8.5–11.3-year-olds. The anatomical data of Vorperian et al. (2011) suggest that gender differences in young adolescents (ages comparable to those used here) are restricted to nasopharynx size, which should not affect the resonant frequencies of fricatives in the oral cavity; however, as with adults, gender differences in adolescents could arise from any variations in articulatory postures that affect front cavity length. Thus, we evaluated whether our proposed measures showed gender differences.

Age. One might expect smaller dimensions to yield higher frequencies overall for fricatives produced by children compared to adults, and child–adult differences have been reported for peak frequencies in studies including children up to 11 years of age (Daniloff et al., 1980; Pentz et al., 1979; Sereno et al., 1987). Studies using moments have not generally arrived at the same conclusion, however. Nittrouer et al. (1989) found that, compared to adults, children 3–7 years of age had M1s that were higher for /ʃ/ but lower for /s/, with the net effect that /s-ʃ/ were less differentiated in children than adults. This pattern was also observed by Nittrouer (1995) and by Nissen and Fox (2005) for 3- and 4-year-olds (but not for 5-year-olds). These studies evaluated adult–child differences and not age effects within child groups. However, of importance for the current work, no age trends were found among the 9–15-year-olds studied by Flipsen et al. (1999a); those authors concluded, based on comparison with adult data from Weismer and Bunton (1999), that adolescents had reached adultlike mean values for /s/ moments.

Current Study

To summarize, the goal of this work was to develop measures of fricatives that could be automated (as are moments), and which would also permit interpretation in terms of articulatory and aerodynamic conditions (which can be difficult using moments). Previous studies of coarticulation and temporal changes within fricatives provide clear expectations for patterns that ought to be reflected in measures of /s/ and provide a means of evaluating the adequacy of the measures. Some but not all past studies have also reported gender effects in adolescent /s/; the statistical methods used here accordingly took gender into account.

Method

Speakers and Recording Methods

Data were drawn from Preston and Edwards (2007) for 12 adolescents (six boys and six girls), 10–15 years of age, with typical production of /s/. Ages for the girls were 10;0–15;3 (years; months; mean = 12;6); those for the boys were 10;2; 14;5 (mean = 12;4). This age range is slightly narrower than the 9;7–15;2 age range of Flipsen et al. (1999a) and the 9;1–15;7 age range used by Karlsson et al. (2002), and comparable to the oldest (12–17-year-old) group in Cheng, Murdoch, Goozée, and Scott (2007). All participants were native speakers of American English from central New York State with no history of neurological, cognitive, orostructural, or fluency problems. Participants passed a hearing screening at 20 dB HL at 500, 1000, 2000, and 4000 Hz; had vocabulary skills within normal limits on the Peabody Picture Vocabulary Test–III (Dunn & Dunn, 1997); and showed speech production skills within normal limits as judged by a certified speech-language pathologist (the third author) and confirmed by a second listener experienced in child speech sound development and disorders. Children were recorded in a quiet room in their schools or in a laboratory setting using a Shure WH20 head-mounted microphone fed to a Rolls MS 54s Pro Mixer Plus, sampled into .wav files at 22 kHz onto a laptop computer.

Speech Materials

Tasks were selected from a protocol administered by Preston and Edwards (2007) that provided a range of speech materials: picture naming, repeating nonwords after a recorded model, sentence repetition following an adult model, and paragraph reading. Thus, the corpus includes connected speech along with single-word productions. This allowed us to verify that the proposed measures were robust across speech tasks. The full set of words with /s/, along with more information on the tasks, is provided in Table 1 of the online supplemental materials. From the recordings, syllables containing /s/ in onset position of a stressed syllable were located. Words containing the cluster /stɹ/ were excluded, because this sequence is undergoing sound change to /ʃtɹ/ in many speakers (Janda & Joseph, 2003; Lawrence, 2000; Mielke, Baker, & Archangeli, 2010; Rutter, 2011; Shapiro, 1995). The corpus provided a maximum of 58 possible words per speaker, including a few cases where the same word was repeated within or across tasks. Not all words were produced by all speakers, chiefly because of errors in repetition or reading. The statistical methods take this variation into account.

As reviewed above, much past work indicates that anticipatory labialization has reliable effects on fricative noise frequencies in children and adults (e.g., Iskarous et al., 2011; Katz et al., 1991; McGowan & Nittrouer, 1988; Munson, 2004; Nittrouer et al., 1989; Soli, 1981; Yeni-Komshian & Soli, 1981). To evaluate whether the measures were sensitive to such coarticulatory effects, each word was coded according to whether or not there was a following labial sound within the syllable. This included not only rounded vowels but also clusters containing /p/, /ɹ/, and/or /w/ (all of which should have similar acoustic effects on /s/ spectra). Preceding context was not coded because many words were produced in isolation or at the beginning of a sentence. In the process of locating the /s/ productions (described in the next section), all words were evaluated by the first author to be acceptable at the level of broad transcription. Thus, words coded as labialized were perceived to contain rounded vowels or labial consonants following the /s/.

Signal Processing

The /s/ segments were located in a spectrographic display using Praat (Boersma & Weenink, 2008), with a frequency display range of 0–10 kHz and preemphasis at 6 dB. Exclusion criteria for extraction were cases where the /s/ could not be segmented with confidence because of background noise, overtalk, or continuous acoustic changes related to surrounding sounds. Clusters in which the following stop was spirantized (which occurred frequently for /st/) were retained. The final data set for the 12 children consisted of 620 productions (mean = 51.7 per speaker; range = 32–58).

The excerpted /s/ tokens were saved as .wav files and read into MATLAB (Version R2008b) for subsequent processing. The duration of each /s/ token was computed, and measures were made for 25-ms segments at the beginning, middle, and end (b, m, e) of each /s/. The b segment started immediately at /s/ onset, and the e segment ended at the final sample of the /s/. The m segment was centered on the midpoint of the /s/, that is, the sample equal to the total sample length of the /s/ divided by 2 and rounded.

For each b, m, and e segment of individual /s/ productions, multitaper spectra were computed (Blacklock, 2004) using eight tapers. Each taper is a weighting function applied to the samples that make up the 25-ms segment. The tapers shape the samples in the segment much as a Hanning window would, but each of the eight tapers shapes them in a different way, and all eight tapers are applied to each segment. Tapers are defined to be orthogonal, so that the differently weighted samples are statistically independent. A transform is then found for each taper, and the results are averaged at each frequency to produce a single multitaper spectrum for that segment. Obtaining multitaper spectra provides for both low error and temporal precision and represents an improvement over the processing used in much past work. In particular, single discrete Fourier transforms (DFTs) are temporally precise but may have large error because of the random nature of frication noise; averaged DFTs have reduced spectral error but involve averaging over time, reducing temporal precision (Shadle, 2012).

For all calculations, no preemphasis was applied, and frequencies below 550 Hz were excluded to remove low-frequency ambient noise along with any lower harmonics related to voicing intruding into the fricative. In the current data, about 10% of the productions had voicing persisting 30 ms or more into the fricative. These occurred chiefly where the fricative followed voiced sounds in the reading and sentence-repetition tasks.

Measures

Moments. The first four spectral moments were computed following Forrest et al. (1988). In this method, spectra are normalized so that the amplitudes of a given spectrum sum to 1; Moments 2–4 are centralized (i.e., M1 is subtracted out); and Moments 3 and 4 are normalized by M2 to yield L3 (coefficient of skewness) and L4 (coefficient of kurtosis).

Forrest et al. (1988) and most other authors who have applied moments analyses did not employ a low-frequency cutoff as was done here. Further, our processing did not apply preemphasis to the data, in contrast to many past studies (including Forrest et al., 1988). Finally, moments (and all other measures) were calculated from multitaper spectra instead of single DFTs. Nevertheless, as described in the Results section, the moment values obtained here are generally consistent with the adolescent data provided by Flipsen, Shriberg, Weismer, Karlsson, and McSweeny (1999b), so

Table 1. Definitions of parameters.

Measure	Definition
Amp_{LMin}	Minimum amplitude over low-frequency range (550–3000 Hz)
Amp_M	Peak amplitude within mid-frequency range (3000–7000 Hz)
$Freq_M$*	Frequency at Amp_M
$Level_M$	Sound level over entire mid-frequency range (3000–7000 Hz)
$Level_H$	Sound level over entire high-frequency range (7000–11025 Hz)
$AmpD_{M-LMin}$*	Amp_M–Amp_{LMin}
$LevelD_{M-H}$*	$Level_M$–$Level_H$
ΔFreq*	Drop in $Freq_M$ at end of fricative, compared to the beginning–middle average: $\Delta F = \frac{Freq_M(b) + Freq_M(m)}{2} - Freq_M(e)$

Note. The four dependent measures used to characterize /s/ are indicated with an asterisk.

these differences in preprocessing do not preclude comparisons to past work. As stated above, moments were included here primarily to provide some commonality with the existing literature and, in particular, with Flipsen et al. (1999a, 1999b).

New measures. A set of measures was designed to capture hypothesized variation over time in /s/ and effects of labiality and gender. These are summarized in Table 1 and are illustrated for two words (one labial, one nonlabial) in Figure 1. As described below, past work suggests that a thorough characterization of /s/ should measure spectral energy in different frequency regions. For example, /s/ has been found to have a mid-frequency peak which, in adults, is in the range of about 3.5–5.0 kHz; it also has relatively little energy in low frequencies (cf., e.g., Bladon & Seitz, 1986; Jassem, 1965). To permit automatic determination of such features, three frequency bands—low, middle, and high—were defined (see next paragraph). The frequency bands constrain the measures so that the highest-amplitude peak in the middle frequency band is highly likely to be the main front-cavity resonance.[4]

The literature is somewhat conflicting on whether fricative frequencies show systematic differences between adolescents and adults. Although Flipsen et al. (1999a) determined that adolescents 9–15 years of age had reached adultlike values for fricative moments, Pentz et al. (1979) reported differences in spectral peaks between adults and children 8.5–11.3 years of age. For the current work, the boundaries of the three frequency bands did not rely on adult values but, rather, were defined based on an examination of sample spectra for words produced by all 12 speakers. Iterative piloting, comparing various frequency ranges and checking the output against sample spectra, was used to establish the values indicated in Table 1 and in Figure 1. The final dependent measures consisted of four quantities, indicated with asterisks in Table 1 and explained in the following paragraphs.

$Freq_M$. Many studies have presented data on fricative spectral peaks (Behrens & Blumstein, 1988; Bladon & Seitz, 1986; Hughes & Halle, 1956; Iskarous et al., 2011; Jassem, 1965; Jongman et al., 2000; Maniwa et al., 2009; Pentz et al., 1979; Seitz et al., 1987; Soli, 1981; Strevens, 1960; Yeni-Komshian & Soli, 1981). The $Freq_M$ measure is an automatically obtained mid-frequency peak. Following past work, we predicted that $Freq_M$ would decrease with labiality (Iskarous et al., 2011; Jongman et al., 2000; Yeni-Komshian & Soli, 1981). Notice in Figure 1 that the nonlabialized token (top) has a $Freq_M$ value of about 6.7 kHz, whereas the labialized token (bottom) has a value of about 5.2 kHz. Longer front cavities and/or smaller lip openings in labial contexts should both have the effect of lowering $Freq_M$. This prediction follows from early theoretical work showing that poles (and zeros) fitted to fricative spectra can be interpreted as resonances (and antiresonances) of the upper vocal tract (Heinz & Stevens, 1961). Shadle's (1985) work with mechanical models also showed that the frequency of the main resonance was lower when the front cavity was longer. Studies, mostly on adults, suggest that the main spectral peak might also show gender effects (Fox & Nissen, 2005; Jongman et al., 2000; Maniwa et al., 2009).

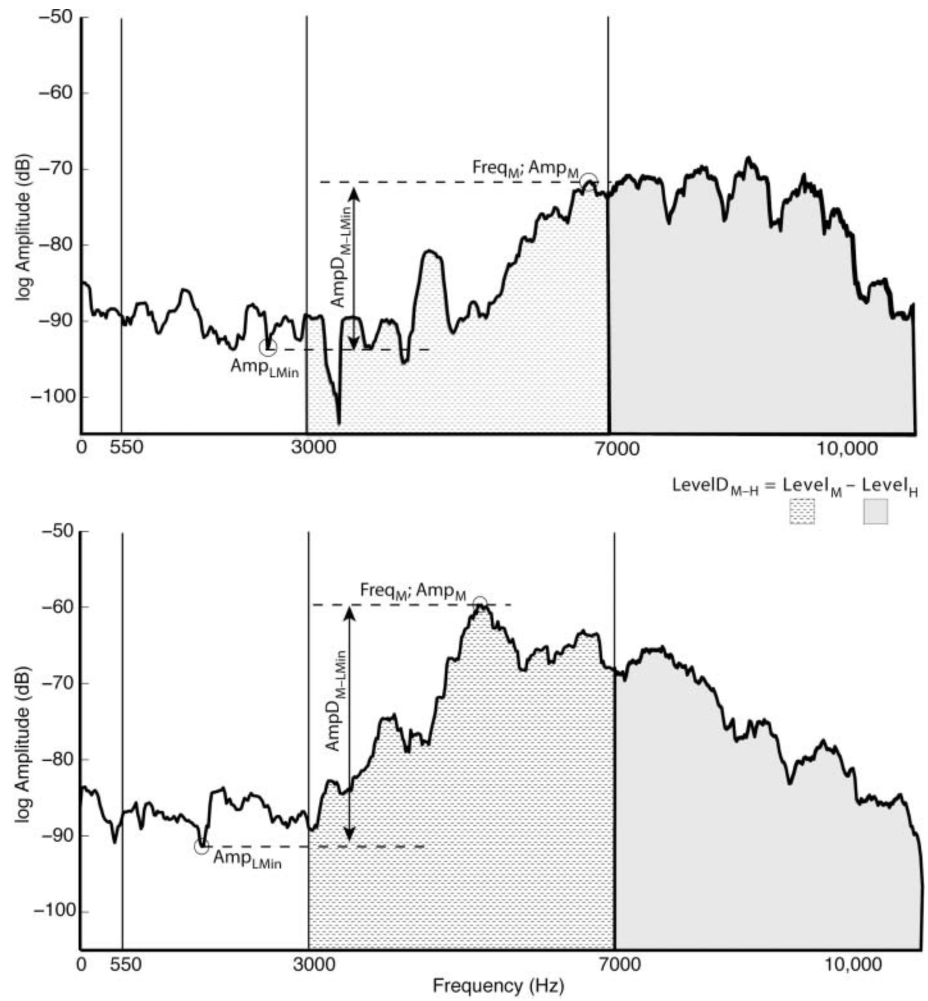
ΔFreq. This parameter was designed to reveal increasing effects of anticipatory labialization toward the end of the fricative. It quantifies the change in $Freq_M$ at the last measurement window, $Freq_M(e)$, as compared to the average of the initial and middle ones, $Freq_M(b)$ and $Freq_M(m)$. It was predicted that labialized tokens would have a larger frequency drop, that is, a higher ΔFreq, than nonlabialized tokens. Because this value represents a frequency change, it should also serve to normalize somewhat over any absolute frequency differences across speakers (e.g., between boys and girls).

$AmpD_{M-LMin}$. This measure, the amplitude difference between the low-frequency minimum and the mid-frequency peak, was intended to capture the degree of sibilance (or /s/ "goodness") and was expected to vary over time. Mechanical modeling experiments (Shadle, 1985, 1990) showed that the sibilant–nonsibilant contrast was effectively modeled as the presence versus absence of an obstacle (representing the teeth) downstream of the constriction, and that $AmpD_{M-LMin}$[5] was

[4]We note that some authors have measured a parameter called the "spectral peak" obtained as the frequency of the maximum-amplitude point over the entire range of the spectrum (e.g., Jongman et al., 2000; Maniwa et al., 2009). This may sometimes but will not necessarily correspond to the lowest front cavity resonance. For our study, we constrained the frequency range because we sought the particular spectral peak corresponding to the lowest front cavity resonance, following the method of Jesus and Shadle (2002).

[5]$AmpD_{M-LMin}$ in Shadle (1985, 1990) was measured manually, not automatically.

Figure 1. Sample spectra from a 10-year-old girl showing measures used to characterize /s/. Top panel: nonlabialized token (the word *seventy-three*); bottom panel: labialized token (the word *squirrel*).



systematically larger in the sibilant (and obstacle) case. The measure quantifies the difference between the low-frequency antiresonance and the mid-frequency peak representing the front-cavity resonance (Heinz & Stevens, 1961). In fricatives produced by adults, enough noise is generated early in the fricative to excite the main peak and generate the low-frequency antiresonance, yielding a positive $AmpD_{M–LMin}$. Moving into the midpoint of the fricative, acoustic and aeroacoustic conditions contribute to a further increase in $AmpD_{M–LMin}$. Continued reduction in the constriction area during fricative formation increases acoustic decoupling, leading to cancellation of back-cavity resonances and lower amplitudes in the low-frequency region. The decreased constriction area also results in greater air particle velocity, which generates more turbulence and raises the amplitude of the main peak. Together, reduction of low-frequency energy and an increase of mid-frequency energy produce higher values of $AmpD_{M–LMin}$ at fricative midpoint. At fricative release, conditions should become similar to those at onset. In short, $AmpD_{M–LMin}$ was predicted to show a mid-/s/ peak. In adult speakers, values typically range from 20–45 dB for sibilants (Shadle, 1985), with the mid-/s/ value approximately 5–15 dB higher than beginning or end values (Jesus & Shadle, 2002).

$LevelD_{M–H}$. $LevelD_{M–H}$, like $AmpD_{M–LMin}$, was designed to show the degree of sibilance and was expected to vary over time. This measure quantifies the balance of acoustic power in mid- and high-frequency ranges (see Figure 1); it is conceptually related to the high-frequency slope measures of fricatives used by Shadle and Mair (1996) and Jesus and Shadle (2002). High-frequency energy content in a well-formed /s/ should be greatest mid-fricative: Reduced constriction area yields an increase in turbulent noise via increased air particle velocity. This leads to a change in the source spectrum, with energy increasing more at high frequencies than at lower frequencies (Shadle, 2012). Jaw raising during the /s/ may also move the lower teeth into the path

of the air jet, enhancing noise source production (Iskarous et al., 2011). Thus, we predicted that $LevelD_{M–H}$ would decrease mid-fricative and increase again at the end. The actual values of $LevelD_{M–H}$ may vary across speakers, especially if their main resonances are at quite different frequencies, but the pattern of change over time should be consistent.

Statistics

Linear mixed-effects (LME) models (Baayen, 2008; Pinheiro & Bates, 2000) were calculated using R 2.10.0 (see R Development Core Team, 2009) in order to test effects of labiality and gender for variables M1, M2, L3, L4, $Freq_M$ (at beginning, middle, and end [b, m, e] time slices), ΔFreq, $AmpD_{M–LMin}$ (b, m, e), and $LevelD_{M–H}$ (b, m, e). Fixed effects were labiality (within speakers) and gender (between speakers). These factors were coded and centered in order to reduce collinearity between factor levels. Speaker and item (i.e., word) were included as random variables. This option represents an advantage of LME models because it permits an assessment of the degree to which observed effects varied across speakers and words. Specifically, for all models we tested whether the fit improved by allowing the slopes of the fixed effects to vary with subject and items by using a log-likelihood test for goodness of fit. Inclusion in the random term can be interpreted as speakers or items differing in the direction and/or magnitude of the effect. In the following, if item and speaker are not mentioned, it implies that the patterns were stable across speakers and words. Because the words were almost entirely different across the tasks (the word *himself* occurred in the reading and sentence-repetition tasks, and the word *stage* occurred in the sentence-repetition and naming tasks), task was not entered as a separate fixed effect because this would have resulted in a highly unbalanced design with many empty cells. Instead, the words were also coded for tasks and entered together as random effects. For all variables, we checked whether there was a systematic clustering of tasks in the modeled random intercepts. Such clustering was not observed, suggesting that the current measures did not vary consistently as a function of task.

The p values presented here are not derived from F values but, rather, are based on Markov Chain Monte Carlo samples with 1,000 simulations. This method is generally preferred to calculating the F values because it is more stable for small sample sizes, and the correct calculation of degrees of freedom is still controversial for mixed linear models (Baayen, 2008). Hypothesis testing based on LME, using MCMC sampling, is known to be anticonservative for small samples ($N < 20$); nevertheless, LME models offer a number of advantages over traditional repeated-measures analyses of variance (ANOVAs). Of particular importance to the present work is that these methods are much more flexible in dealing with missing data (Baayen, Davidson, & Bates, 2008), and our data set was unbalanced (not all speakers produced each word or the same number of repetitions). Because all tokens can be included in the statistical analysis, LME models offer more statistical power than ANOVAs. Reduced statistical power and other issues associated with ANOVAs have been reviewed in some detail by several authors (e.g., Guo, Owen, & Tomblin, 2010; Max & Onghena, 1999; Owen, 2010).

Results

Spectral Moments

Table 2 presents descriptive statistics for the four moments at b, m, e windows. All values are within the band defined by 1 *SD* below the lowest mean and 1 *SD* above the highest mean for individual speakers reported by Flipsen

Table 2. Descriptive statistics for spectral moments (M1, M2, L3, L4), measured at the beginning, middle, and end (b, m, e) of the fricative.

			Measurement window *M* (*SD*)		
Measure	**Labiality**	**Gender**	**b**	**m**	**e**
M1	–	F	6260 (1467)	7643 (890)	6507 (1399)
	–	M	4759 (1068)	5985 (781)	5034 (1112)
	+	F	5759 (1516)	7306 (867)	6098 (1280)
	+	M	4433 (1100)	5605 (857)	4513 (1320)
M2	–	F	2086 (536)	1492 (419)	2092 (557)
	–	M	1850 (407)	1675 (255)	1932 (394)
	+	F	2054 (467)	1629 (319)	2020 (463)
	+	M	1743 (448)	1738 (313)	1909 (417)
L3	–	F	–0.520 (1.003)	–0.836 (0.766)	–0.626 (0.794)
	–	M	0.573 (0.957)	0.342 (0.825)	0.318 (0.698)
	+	F	–0.248 (1.021)	–0.543 (0.785)	–0.102 (0.741)
	+	M	0.928 (1.190)	0.602 (1.085)	0.841 (0.843)
L4	–	F	1.410 (3.805)	2.181 (2.858)	0.807 (1.947)
	–	M	1.652 (3.842)	0.445 (1.980)	0.593 (1.769)
	+	F	1.195 (3.836)	0.872 (2.196)	0.184 (1.683)
	+	M	3.389 (7.679)	1.000 (5.493)	1.368 (2.740)

Note. M1 and M2 values are in Hz; L3 and L4 are dimensionless.

et al. (1999b). It is evident that all four moments vary over the time course of the fricative (cf. also Jongman et al., 2000); however, following Nissen and Fox (2005) and Nittrouer (1995), we calculated statistics on midpoint values only. Flipsen et al. (1999a) also concluded that midpoint values were the most optimal for characterizing gender differences, although they might miss some coarticulatory effects.

Results of the LME models for spectral moments are given in Table 3. These statistics showed significant effects of labiality and gender, but no interactions, for both M1 and L3: M1 was higher and L3 was lower/more negative (a greater balance of high-frequency energy) in nonlabial contexts and in girls. These two measures showed significant model improvements if the factor gender was varied within the random factor of word, M1: $\chi^2 = 14.04$, $pMCMC < .01$; L3: $\chi^2 = 10.27$, $pMCMC < .01$, because three words (*screwdriver, student,* and the nonsense word [sə'sɪdəbi]) showed smaller gender differences. M2 was significantly higher in labial contexts, but the effect was small (<100 Hz). The pattern was consistent across all speakers but one, who had a small reversal, and some extreme outliers in both directions. L4 showed no main effect of labiality or gender, but a significant Labiality × Gender interaction: Whereas the girls consistently showed higher L4 (more peaky distributions) for nonlabial contexts, the boys did not show such an effect, and one boy had some extremely high L4 values in clusters involving labial sounds (possibly reflecting whistled fricatives).

$Freq_M$ and $\Delta Freq$

Descriptive data for $Freq_M$ and $\Delta Freq$, along with the two spectral difference/balance measures (discussed in the next section), are provided in Table 4, and results of the LME models for these measures are given in Table 5. $Freq_M$ values for the three time windows (b, m, e) and all speakers are plotted in Figure 2.

Based on past work, $Freq_M$ was predicted to be higher in nonlabial contexts and, possibly, in girls. The statistics indicate that labiality and gender had significant effects on $Freq_M$ at all time points. Figure 2 shows that $Freq_M$ was generally lower for labial contexts than nonlabial, but speakers differed considerably in the magnitude of the difference, and one girl showed a small reversal for the beginning and middle windows. For the middle window, this led to a significant improvement of the model, $\chi^2 = 6.66$, $pMCMC < .05$, if the slope of labiality was included in the random factor of subjects. For the end window the model improved when the fixed factor of gender was included in the random factor of term for word, $\chi^2 = 11.14$, $pMCMC < .01$. The significant gender effect indicated that values were higher on average in girls than in boys (see Figure 2).

Comparison of the labial–nonlabial differences for $Freq_M$(b) and $Freq_M$(m) (first and second panels of Figure 2) with $Freq_M$(e) (third panel) reveals that the labial influence became stronger at the end of the fricative: Labial and nonlabial values within individual speakers are generally more widely separated in the bottom (e) panel than in the top two (b, m). This change over time is quantified by $\Delta Freq$, shown in Figure 3. Recall that $\Delta Freq$ was intended to capture effects of labiality while normalizing for any overall frequency differences between boys and girls. The statistics (Table 5) indicate that the measure was successful in this respect: $\Delta Freq$ showed a significant effect of labiality, but not of gender. The fixed effect of gender within the random-effect word improved the model significantly, $\chi^2 = 7.58$, $pMCMC < .05$, meaning that a gender effect was seen for some words but not all.

Table 3. Results of linear mixed-effects (LME) models at midpoint values of the four moments with fixed effects of labiality and gender.

Measure	Fixed effect	*t*	*pMCMC*
M1	Labiality	–4.39	.001*
	Gender	–6.33	.001*
	Labiality × Gender	–0.48	.633
M2	Labiality	2.92	.004*
	Gender	1.87	.062
	Labiality × Gender	–1.46	.145
L3	Labiality	4.176	.001*
	Gender	4.115	.001*
	Labiality × Gender	–0.185	.856
L4	Labiality	–0.827	.409
	Gender	–1.406	.160
	Labiality × Gender	3.496	.001*

Note. Values significant at *pMCMC* < .05 are indicated with asterisks.

$AmpD_{M–LMin}$ and $LevelD_{M–H}$

These parameters were designed to capture the development of sibilance, and they were mainly expected to differ over time. However, context and gender effects might also be observed, owing to differences in the relative spectral balance. The time-varying patterns, plotted in Figure 4, show that both measures varied as predicted over the time course of the fricative. Specifically, at mid-fricative, $AmpD_{M–LMin}$ was higher and $LevelD_{M–H}$ was lower as compared to the beginning and end of the fricative. For both $AmpD_{M–LMin}$ and $LevelD_{M–H}$, the statistics showed a significant improvement if the fixed factor time was included in the random factors subject and item. In the case of $AmpD_{M–LMin}$, midpoint values were significantly higher than beginning and endpoint values ($pMCMC < .0001$), but beginning and end values did not differ. For $LevelD_{M–H}$, midpoint values were significantly lower than beginning and endpoint values ($pMCMC < .0001$), and end values were also significantly lower than beginning values ($pMCMC = .00756$). For both of these analyses, significance testing used a Holm adjustment for multiple factor levels ($df = 2$).

For $AmpD_{M–LMin}$, the LME results also showed that labiality was significant at fricative midpoint (see Table 4), but the effect was quite small (around 1 dB on average; cf. Table 3), and speakers varied in the magnitude and direction of the difference. At fricative end, a significant Gender × Labiality interaction was observed: Boys had equivalent values in labial and nonlabial contexts, whereas girls had

Table 4. Descriptive statistics for ΔFreq, $Freq_M$, $AmpD_{M-LMin}$, and $LevelD_{M-H}$, measured at beginning, middle, and end of the fricative.

Measure	Labiality	Gender	Measurement window *M (SD)* n/a	b	m	e
ΔFreq	–	F	19 (274)			
	–	M	64 (160)			
	+	F	424 (247)			
	+	M	456 (183)			
$Freq_M$	–	F		5727 (1031)	6320 (633)	5953 (1012)
	–	M		4593 (887)	4843 (771)	4672 (804)
	+	F		5176 (1070)	5742 (888)	5030 (1035)
	+	M		4118 (676)	4401 (718)	3857 (712)
$AmpD_{M-LMin}$	–	F		21.04 (6.64)	27.09 (6.91)	21.38 (7.35)
	–	M		22.27 (7.00)	29.37 (5.41)	23.25 (6.47)
	+	F		21.22 (6.61)	27.91 (5.70)	24.13 (7.24)
	+	M		22.92 (6.36)	30.36 (5.86)	22.95 (7.68)
$LevelD_{M-H}$	–	F		–0.24 (5.37)	–4.55 (4.33)	–1.44 (3.90)
	–	M		6.68 (4.33)	3.53 (3.78)	5.44 (3.98)
	+	F		1.90 (4.92)	–2.20 (4.35)	1.25 (4.35)
	+	M		8.30 (4.97)	4.54 (4.37)	6.12 (5.13)

Note. ΔFreq and $Freq_M$ are in Hz; $AmpD_{M-LMin}$ and $LevelD_{M-H}$ are in dB.

Table 5. Results of LME models with fixed effects of labiality and gender for $Freq_M$, ΔFreq, $AmpD_{M-LMin}$, and $LevelD_{M-H}$ measures.

Measure	Time	Fixed effect	*t*	*pMCMC*
$Freq_M$	b	Labiality	–4.07	0.001*
		Gender	–4.71	0.001*
		Labiality × Gender	0.61	0.592
	m	Labiality	–4.71	0.001*
		Gender	–6.36	0.001*
		Labiality × Gender	0.52	0.240
	e	Labiality	–8.58	0.001*
		Gender	–5.31	0.001*
		Labiality × Gender	0.10	0.474
ΔFreq	n/a	Labiality	3.34	0.001*
		Gender	–0.38	0.674
		Labiality × Gender	0.10	0.956
$AmpD_{M-LMin}$	b	Labiality	0.89	0.358
		Gender	0.52	0.508
		Labiality × Gender	0.57	0.574
	m	Labiality	2.07	0.026*
		Gender	0.86	0.300
		Labiality × Gender	0.21	0.870
	e	Labiality	0.57	0.518
		Gender	–0.02	0.962
		Labiality × Gender	–3.39	0.002*
$LevelD_{M-H}$	b	Labiality	2.65	0.002*
		Gender	5.76	0.001*
		Labiality × Gender	–0.80	0.432
	m	Labiality	4.05	0.001*
		Gender	6.40	0.001*
		Labiality × Gender	–1.17	0.022*
	e	Labiality	4.14	0.001*
		Gender	4.82	0.001*
		Labiality × Gender	–3.27	0.002*

Note. Analyses for $Freq_M$, $AmpD_{M-LMin}$, and $LevelD_{M-H}$ were performed at the beginning, middle, and end of the fricative. Values significant at *pMCMC* < .05 are indicated with asterisks.

higher $AmpD_{M-LMin}$ values in labial contexts (i.e., showed a coarticulatory effect). Since the degree of coarticulation may vary with speech rate (e.g., Gay, 1978), a follow-up LME analysis was done on /s/ durations. This showed that the girls had slightly shorter /s/ durations than the boys (21 ms on average; $t = 2.474$, $pMCMC = .024$, with a significant improvement in slope for gender within word because of exceptionally long durations in two boys for the word *sixty-eighth*). Thus, the gender effect for coarticulation may be mediated by speech rate in this sample.

$LevelD_{M-H}$ showed significant labiality and gender effects at all three time points. Values were higher in labial contexts and in boys (see Figure 4). The gender effect can be interpreted as showing a stronger balance of high-frequency energy in girls. There was also a significant Context × Gender interaction at fricative midpoint and end; similar to the case with $AmpD_{M-LMin}$, girls showed larger effects of labiality than boys.

Discussion

Spectral Moments

On the whole, the results for spectral moments are similar to those of past studies. The significant labiality effects for M1 are consistent with those previously observed for adults and for younger children (Katz et al., 1991; Nittrouer, 1995; Nittrouer et al., 1989). In the current data, labiality also affected M2 and L3, but the effect for M2 was of very small magnitude. L4 varied considerably across speakers and largely seemed to reflect speaker-specific differences in coarticulatory effects (including whistle mechanisms in rounded contexts). Significant effects of gender were also observed for M1 and L3. This is comparable to the findings of Flipsen et al. (1999a), who recorded a larger

Figure 2. Individual speaker means and standard deviations (girls on left, boys on right) for FreqM measured at beginning, middle, and end time slices in the top, middle, and bottom panels, respectively.

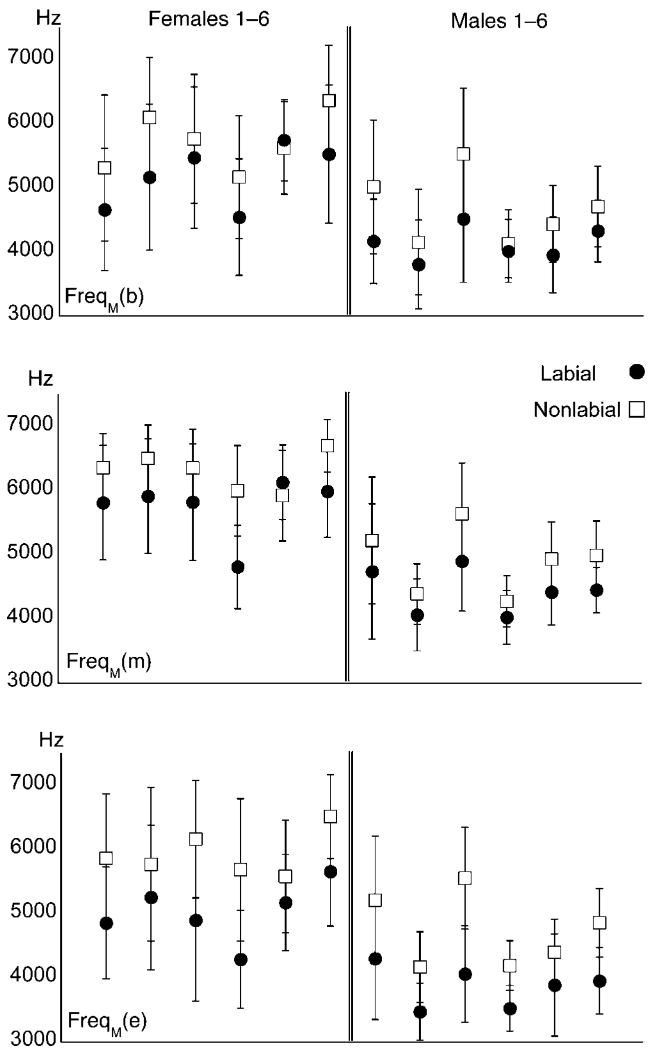
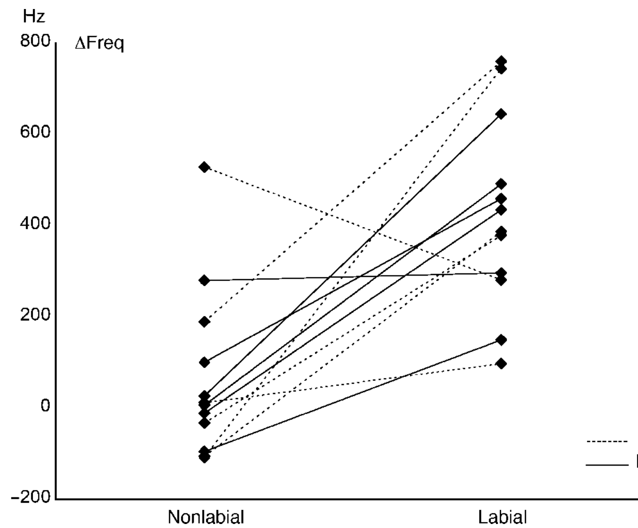


Figure 3. Individual speaker data for ΔFreq in labial and nonlabial contexts.



number of adolescent speakers ($N = 26$) but used much more limited speech contexts (10 words produced in citation form). Thus, these patterns appear to be generally representative of this speaker population. This parallelism also suggests that our speaker group, although of more modest size, is not atypical. Because the moments were mainly provided to permit comparison with past work, the rest of the discussion is focused on the new measures.

New Measures: Patterns and Validity

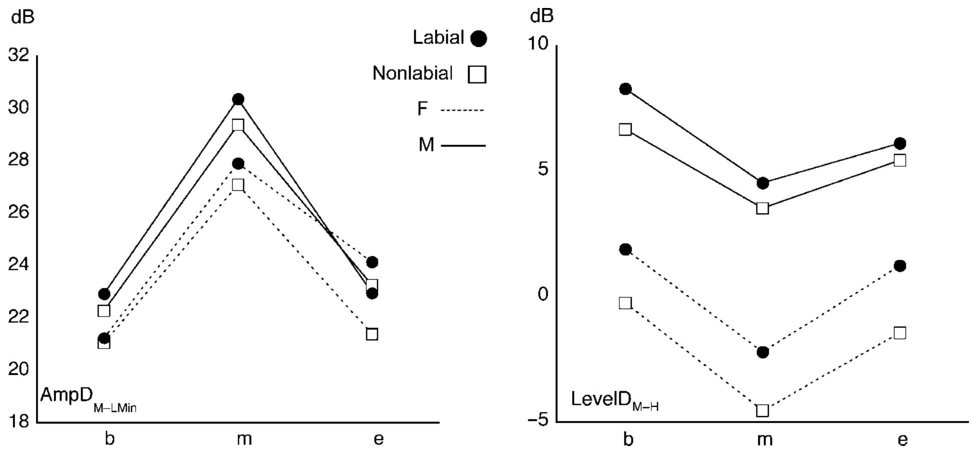
The primary goal of this work was to develop theoretically driven, automatable measures of fricative noise spectra and evaluate whether they revealed expected changes over time, coarticulation, and possible gender differences. The data support the validity of the measures in that they behaved as predicted: Freq_M varied with labiality as well as gender; ΔFreq differentiated labial and nonlabial contexts and did not vary by gender; and the amplitude and level difference parameters, $\text{AmpD}_{\text{M–LMin}}$ and $\text{LevelD}_{\text{M–H}}$, showed changes over time consistent with the buildup of sibilance during /s/ production, along with some small effects of gender and phonetic context. By design, the proposed measures allow interpretation in terms of articulatory and aerodynamic conditions.

$Freq_M$ and ΔFreq. Freq_M was designed to be an automated measure of the main mid-frequency peak of the fricative. The finding of lower Freq_M in labial contexts can be interpreted as a result of longer front cavities and/or smaller lip openings (Heinz & Stevens, 1961). M1 was also affected by labialization in our data, similar to previous work (Katz et al., 1991; Nittrouer, 1995; Nittrouer et al., 1989). However, as outlined in the Introduction, M1 varies not only with front cavity size but also changes in spectral skew and speaking effort. Thus, Freq_M, as a simple spectral peak measure, is more transparent than M1 as an indicator of filter effects.

The observed gender effects on Freq_M are consistent with past reports of higher frequency spectral peaks in female speakers compared to male speakers (Fox & Nissen, 2005; Jongman et al., 2000; Schwartz, 1968; Yeni-Komshian & Soli, 1981), which could arise either from differences in vocal-tract dimensions or articulatory postures (cf. Avery & Liss, 1996; Fuchs & Toda, 2010) that affect front cavity size. Given the minimal anatomical differences observed by Vorperian et al. (2011) for boys versus girls in the age range 10–14; 11 years, a sociophonetic explanation appears to account best for the current data.

Temporal variation in Freq_M was quantified by ΔFreq. As a difference measure, ΔFreq was designed to normalize over any gender effects, and indeed gender effects were significant for Freq_M but not for ΔFreq. A few studies have measured coarticulatory effects throughout fricatives (Iskarous et al.,

Figure 4. AmpDM–LMin (left panel) and LevelDM–H (right panel) measures plotted over time.



2011; Munson, 2004), but more typically authors have assessed coarticulation at a few distinct time windows within fricatives (e.g., Katz et al., 1991; Nittrouer et al., 1989; Sereno et al., 1987). Because coarticulation is inherently a temporal phenomenon, a temporally based parameter like ΔFreq may be a useful alternative to single-point quantities. An analogous ΔFreq measure could also be established to compare anticipatory and carryover coarticulation, comparing the value of $Freq_M$(b) to the $Freq_M$(m) and $Freq_M$(e) average.

$AmpD_{M\text{–}LMin}$ and $LevelD_{M\text{–}H}$. These amplitude and level difference measures were designed to quantify the degree of sibilance (or /s/ "goodness") and the changes in noise source strength over time in different frequency regions of the spectrum. Both measures varied significantly over time in the expected directions.

All children exhibited positive $AmpD_{M\text{–}LMin}$ values, with peak values occurring mid-/s/, as predicted. The increase from beginning to mid values ranged from around 6–10 dB (cf. Table 3), compared to 5–15 dB in adult spectra (Jesus & Shadle, 2002; Shadle, 2012). The adolescents had a mid-sibilant peak of 27–30 dB, compared to 28–45 dB observed for adults (Shadle, 2012). The slightly lower average values obtained here could arise from a number of sources. One is speech material. The current adolescent data included connected speech, whereas most work on fricatives (in both adults and children) has evaluated single-word productions or sustained fricatives. One feature of running speech is that voicing assimilation may occur. As noted in the Method section, about 10% of our /s/ tokens included some intrusive voicing. This was almost always observed as carryover from a preceding voiced sound, however, and thus mostly affected the first time window. As such, it does not account for differences at fricative midpoint or end. Further, the low-frequency cutoff of 550 Hz covered the first two harmonics for most speakers, so it should exclude most effects of voicing. It does not appear, then, that carryover voicing can account for much of this difference. It may be that running speech contexts simply lend themselves to less extreme /s/ constrictions, leading to less formant cancellation in the low-frequency band. Alternatively, the degree of /s/ constriction may differ between children and adults even when the speaking task is held constant. McGowan and Nittrouer (1988) observed stronger F2s in fricatives produced by 3–7-year-old children compared to adults, and suggested that children did not produce a sufficiently tight constriction to cancel back-cavity resonances. Lastly, it may be that adolescent speakers do not direct the airflow toward the teeth as consistently as adults, leading to an inefficient frication source and, hence, a lower $AmpD_{M\text{–}LMin}$ value. Further study is needed to determine whether the relatively low $AmpD_{M\text{–}LMin}$ values observed here are an effect of speaking task or represent a real developmental phenomenon.

Along with showing considerable variation over time, $AmpD_{M\text{–}LMin}$ varied, to a much smaller degree, with labiality and gender. It is not immediately clear how the labiality effect might arise, but the gender effect might indicate a need to modify the definitions of low-, mid- and high-frequency bands for future research in adolescents. The current work used equivalent cutoff frequencies for boys and girls given that the existing literature did not solidly establish gender differences in this age range, but post hoc inspection of the data suggests that for some adolescent girls, the main resonance (i.e., the value targeted for $Freq_M$) may occasionally occur above 7 kHz. Cases where parameter settings failed to choose the main peak could have led to a reduced magnitude of the mid-frequency value chosen as Amp_M in these tokens, reducing the value for $AmpD_{M\text{–}LMin}$ in some girls and accounting for this small gender difference.

$LevelD_{M\text{–}H}$, like $AmpD_{M\text{–}LMin}$, varied as predicted over time. As explained in the Method section, it was expected that measures of $LevelD_{M\text{–}H}$ might differ over speakers but would change consistently over time, and the inferential statistics showed this to be the case. The statistics also revealed a small effect of labiality, and a somewhat larger effect of gender. The gender difference can be explained as follows. $LevelD_{M\text{–}H}$ was designed as an estimate of a source difference, namely an increased source strength at high frequencies relative to mid frequencies, but it is also affected by filter properties: If all else remains the same, as the frequency of the main mid-range peak increases, the energy in the mid-frequency band decreases, and the value of $\text{LevelD}_{\text{M–H}}$ drops. Context can also affect $\text{LevelD}_{\text{M–H}}$ values, based on similar logic. Because a labial context tends to lower the frequency of the main peak, the energy in the mid-frequency band will tend to increase, leading to a higher $\text{LevelD}_{\text{M–H}}$ value.

New Measures: Issues to Explore

The results generally support the validity of the proposed measures and also suggest some possible improvements for future work. First, one of the main heuristic settings, the upper frequency limit of the mid-frequency band, was held constant for male and female speakers here, given that gender differences in the upper vocal tract in this age range seem to be minimal (Vorperian et al., 2011). Yet it appears that some adolescent girls may sometimes have main front-cavity resonances higher than the current high-frequency bound of 7 kHz, especially at fricative midpoint and in nonlabial contexts. Future work might implement a more complex algorithm for finding the mid-frequency peak, taking into account expected variation as a function of speaker, time point, and/or phonetic context. Second, for $\text{AmpD}_{\text{M–LMin}}$, the smaller magnitude in our adolescent speakers as compared to adults is worth exploring. The range of amplitudes in the low-frequency region could be measured to test whether the differences between child and adult fricative spectra noted by McGowan and Nittrouer (1988) can account for smaller $\text{AmpD}_{\text{M–LMin}}$ values in younger speakers. Refining the method of finding Freq_{M} may also affect the $\text{AmpD}_{\text{M–LMin}}$ values computed for girls and may account for the observed gender effect. Third, $\text{LevelD}_{\text{M–H}}$ showed the expected change over time, but the values are confounded with filter effects, a likely explanation for the gender and labiality effects. It would be useful to find a way to lessen the filter effects. Adult values for $\text{LevelD}_{\text{M–H}}$ are also needed for comparison.

In sum, the data reported here suggest that the proposed measures, which have strong theoretical justification based on extensive studies of mechanical models and adult speakers, may also have utility in characterizing aspects of fricative production across populations. The current study assessed temporal, coarticulatory, and gender effects in a modest number of speakers. The consistency between our moments results and those of Flipsen et al. (1999a) provides some assurance regarding the representativeness of our sample, but these methods should be evaluated further using data from more speakers and different populations. It is likely that the division into low-, mid-, and high-frequency bands will need to be adjusted based on the age and gender of the speakers, as is the case, for example, with formant- or f0-tracking algorithms. We plan to apply these measures to fricative misarticulations to evaluate their efficacy in differentiating typical and atypical productions. The measures of $\text{AmpD}_{\text{M–LMin}}$ and $\text{LevelD}_{\text{M–H}}$, which were designed to capture aspects of sibilance, may be particularly useful in this regard. We will also compare the acoustic measures with simultaneously collected physiological data.

Conclusions

The current data support the validity of the proposed measures and provide preliminary descriptive data on typically speaking adolescents. Freq_{M} is an automatically identified estimate of the main spectral peak, used in much past work. ΔFreq appears to be a promising measure for evaluating coarticulation, reducing three values of Freq_{M} into a single parameter and in the process normalizing for gender. It can also be modified to compare anticipatory and carryover coarticulation. The amplitude and level difference measures, $\text{AmpD}_{\text{M–LMin}}$ and $\text{LevelD}_{\text{M–H}}$, as indices of sibilance, might serve to differentiate some cases of typical and misarticulated /s/. In future work, we plan to extend these methods to studies of other fricatives and to children who misarticulate fricatives. We will also explore differences between adolescents and adults in more detail, using comparable corpora.

Acknowledgments

This work was supported, in part, by National Institute on Deafness and Other Communication Disorders (NIDCD) Grants DC006705 (awarded to Christine Shadle) and DC008780 (awarded to Louis Goldstein) as well as by National Institutes of Health Grant T32HD7548 (awarded to Carol Fowler). Preliminary portions of this work were presented at the 157th (Portland, 2009) and 161st (Seattle, 2011) meetings of the Acoustical Society of America, and the 9th International Seminar on Speech Production (Montreal, 2011). We express our appreciation to Khalil Iskarous for assistance in programming early versions of the MATLAB analyses; Mary Louise Edwards for assistance with stimulus development and verifying typical speech patterns in the speakers; Dyala Sophia Eid, and Natoy Gayle (both supported by graduate assistantships at Long Island University, Brooklyn Campus) for assistance in preliminary data processing; and Anders Löfqvist, Doug Whalen, Jody Kreiman, and Scott Thomson for comments on earlier versions of this work.

References

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *The Journal of the Acoustical Society of America, 99,* 3738–3748.

Baayen, H. (2008). *Analyzing linguistic data*. Cambridge, United Kingdom: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

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

Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics, 16,* 295–298.

Bendat, J. S., & Piersol, A. G. (2000). *Random data: Analysis and measurement procedures* (3rd ed.). New York, NY: Wiley.

Blacklock, O. S. (2004). *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods*. Unpublished doctoral dissertation, University of Southampton, England.

Bladon, R. A. W., & Seitz, P. F. D. (1986). Spectral edge orientation as a discriminator of fricatives. *The Journal of the Acoustical Society of America, 80,* S18.

Boersma, P., & Weenink, D. (2008). Praat [Computer program] (Version 5.1.30). Retrieved from www.praat.org

Boothroyd, A., & Medwetsky, L. (1992). Spectral distribution of /s/ and frequency response of hearing aids. *Ear and Hearing, 13,* 150–157.

Cheng, H. Y., Murdoch, B. E., Goozée, J. V., & Scott, D. (2007). Physiologic development of tongue-jaw coordination from childhood to adulthood. *Journal of Speech, Language, and Hearing Research, 50,* 352–360.

Daniloff, R. G., Wilcox, K., & Stephens, M. I. (1980). An acoustic-articulatory description of children's defective /s/ productions. *Journal of Communication Disorders, 13,* 347–363.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—III*. Circle Pines, MN: AGS.

Evers, V., Reetz, H., & Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics, 26,* 345–370.

Flipsen, P., Jr., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999a). Acoustic characteristics of /s/ in adolescents. *Journal of Speech, Language, and Hearing Research, 42,* 663–677.

Flipsen, P., Jr., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999b). *Acoustic reference data for /s/ in adolescents* (Technical Report No. 8). University of Wisconsin-Madison, Waisman Center on Mental Retardation and Human Development. Retrieved from www.waisman.wisc.edu/phonology/techreports/TREP08.PDF

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

Fox, R. A., & Nissen, S. L. (2005). Sex-related acoustic changes in voiceless English fricatives. *Journal of Speech, Language, and Hearing Research, 48,* 753–765.

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. Żygis (Eds.), *An interdisciplinary guide to turbulent sounds* (pp. 281–302). Berlin, Germany: Mouton deGruyter.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America, 63,* 223–230.

Gruber, F. A. (1999). Variability and sequential order of consonant normalization in children with speech delay. *Journal of Speech, Language, and Hearing Research, 42,* 460–472.

Guo, L.-Y., Owen, A. J., & Tomblin, J. B. (2010). Effect of subject types on the production of auxiliary *is* in young English-speaking children. *Journal of Speech, Language, and Hearing Research, 53,* 1720–1741.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America, 33,* 589–596.

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America, 28,* 303–310.

Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory-acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America, 129,* 944–954.

Janda, R. D., & Joseph, B. D. (2003). Reconsidering the canons of sound-change: Towards a "big bang" theory. In B. Blake & K. Burridge (Eds.), *Historical linguistics 2001: Selected papers from the 15th International Conference on Historical Linguistics* (pp. 205–219). Melbourne, Australia: John Benjamins.

Jassem, W. (1965). The formants of fricative consonants. *Language and Speech, 8,* 1–16.

Jesus, L. M. T., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics, 30,* 437–464.

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

Karlsson, H., Shriberg, L. D., Flipsen, P., Jr., & McSweeny, J. (2002). Acoustic phenotypes for speech-genetics studies: Toward an acoustic marker for residual /s/ distortions. *Clinical Linguistics & Phonetics, 16,* 403–424.

Katz, W. F., 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.

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of Functional Data Analysis. *The Journal of the Acoustical Society of America, 124,* 3158–3170.

Lawrence, W. P. (2000). /str/→/ʃtr/: Assimilation at a distance? *American Speech, 75,* 82–87.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America, 125,* 3962–3973.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research, 42,* 261–270.

McGowan, R. S., & Nittrouer, S. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *The Journal of the Acoustical Society of America, 83,* 229–236.

Mielke, J., Baker, A., & Archangeli, D. (2010). Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In C. Fougeron & M. P. D'Imperio (Eds.), *Laboratory phonology* (Vol. 10, pp. 699–730). Berlin, Germany: Mouton de Gruyter.

Miller, J. E., Mathews, M. V., & Hughes, G. W. (1962). Spectral analysis of fricative consonants. *The Journal of the Acoustical Society of America, 34,* 743.

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

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

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America, 113,* 2850–2860.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America, 109,* 1181–1196.

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,* 2570–2578.
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,* 520–530.
Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (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.
Owen, A. J. (2010). Factors affecting accuracy of past tense production in children with Specific Language Impairment and their typically developing peers: The influence of verb transitivity, clause location, and sentence type. *Journal of Speech, Language, and Hearing Research, 53,* 993–1014.
Pentz, A., Gilbert, H. R., & Zawadzki, P. (1979). Spectral properties of fricative consonants in children. *The Journal of the Acoustical Society of America, 66,* 1891–1893.
Pinheiro, J., & Bates, M. D. (2000). *Mixed effects models in S and S-PLUS*. New York, NY: Springer-Verlag.
Preston, J. L., & Edwards, J. (2007). Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech and Hearing Services in Schools, 38,* 297–308.
R Development Core Team. (2009). R: A language and environment for statistical computing [Computer program]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.
Rutter, B. (2011). Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English. *Journal of the International Phonetic Association, 41,* 27–40.
Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37,* 55–63.
Schwartz, M. F. (1968). Identification of speaker sex from isolated, voiceless fricatives. *The Journal of the Acoustical Society of America, 43,* 1178–1179.
Seitz, P. F. D., Bladon, R. A. W., & Watson, I. M. C. (1987). Across-speaker and within-speaker variability of British English sibilant spectral characteristics. *The Journal of the Acoustical Society of America, 82,* S37.
Sereno, J. A., Baum, S. R., Marean, G. C., & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *The Journal of the Acoustical Society of America, 81,* 512–519.
Shadle, C. H. (1985). *The acoustics of fricative consonants* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.
Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 187–209). Dordrecht, the Netherlands: Kluwer Academic Publishers.
Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics, 19,* 409–424.
Shadle, C. H. (2012). Acoustics and aerodynamics of fricatives. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *Handbook of laboratory phonology* (pp. 511–526). New York, NY: Oxford University Press.
Shadle, C. H., & Mair, S. J. (1996). Quantifying spectral characteristics of fricatives. *Proceedings of the International Conference of Spoken Language Processing [ICSLP]* (pp. 1521–1524). Philadelphia, PA: ICSLP.
Shadle, C. H., & Scully, C. (1995). An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics, 23,* 53–66.
Shapiro, M. (1995). A case of distant assimilation: /str/→/ʃtr/. *American Speech, 70,* 101–107.
Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders, 4,* 38–53.
Shriberg, L. D. (2009). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In R. Paul & P. Flipsen (Eds.), *Speech sound disorders in children* (pp. 1–34). San Diego, CA: Plural.
Smit, A. B., Hand, L., Feininger, J. J., Bertha, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55,* 779–797.
Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America, 70,* 976–984.
Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech, 3,* 32–49.
Tjaden, K., & Turner, G. S. (1997). Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 40,* 1358–1372.
Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research, 54,* 995–1010.
Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research, 45,* 1119–1133.
Weismer, G., & Bunton, K. (1999). Influences of pellet markers on speech production behavior: Acoustical and perceptual measures. *The Journal of the Acoustical Society of America, 105,* 2882–2894.
Weismer, G., & Elbert, M. (1982). Temporal characteristics of "functionally" misarticulated /s/ in 4- to 6-year-old children. *Journal of Speech and Hearing Research, 25,* 275–287.
Whalen, D. H. (1991). Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *The Journal of the Acoustical Society of America, 90,* 1776–1785.
Yeni-Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America, 70,* 966–975.