Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective[a]

Shawn L. Nissen[b]
Department of Audiology and Speech-Language Pathology (138 TLRB), Brigham Young University, Provo, Utah 84602

Robert Allen Fox
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

Scientists have made great strides toward understanding the mechanisms of speech production and perception. However, the complex relationships between the acoustic structures of speech and the resulting psychological percepts have yet to be fully and adequately explained, especially in speech produced by younger children. Thus, this study examined the acoustic structure of voiceless fricatives (/f, θ, s, ʃ/) produced by adults and typically developing children from 3 to 6 years of age in terms of multiple acoustic parameters (durations, normalized amplitude, spectral slope, and spectral moments). It was found that the acoustic parameters of spectral slope and variance (commonly excluded from previous studies of child speech) were important acoustic parameters in the differentiation and classification of the voiceless fricatives, with spectral variance being the only measure to separate all four places of articulation. It was further shown that the sibilant contrast between /s/ and /ʃ/ was less distinguished in children than adults, characterized by a dramatic change in several spectral parameters at approximately five years of age. Discriminant analysis revealed evidence that classification models based on adult data were sensitive to these spectral differences in the five-year-old age group. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2010407]

I. INTRODUCTION

A considerable amount of research has attempted to explain how the human perceptual system so accurately and efficiently perceives speech. Although many of these endeavors have made great strides toward understanding speech communication, the complex relationships between the acoustic structures of speech and the resulting psychological percepts have yet to be fully and adequately explained. Fortunately, advances in computer technology have allowed scientists to utilize more specific types of analysis on a variety of different acoustic parameters.

Previous investigations into the acoustic nature of speech have been aided by "spectral moments analysis" (e.g., Forrest *et al.*, 1988; Jongman *et al.*, 2000; Nittrouer, 1995). This type of analysis examines the spectral characteristics of discrete time segments of the speech signal in terms of multiple statistical moments (i.e., mean, variance, skewness, and kurtosis). By combining these various "statistical snapshots" of perceptually relevant segments of the physical waveform, researchers have attempted to identify unique and discrete patterns of acoustic energy within the speech signal.

Historically, speech research utilizing spectral moments has generally followed two lines of investigation. One line of

research has examined whether spectral moment data could provide a high rate of obstruent classification through discriminant function analysis. In particular, this research has aimed to identify acoustic parameters that might serve to differentiate or discriminate adult productions of stop and fricative consonants in terms of place of articulation. Utilizing spectral moments data as well as more traditional types of acoustic measures, several studies have found that the acoustic signal of adult obstruent productions can be characterized by distinct spectral patterns of acoustic energy (e.g., Forrest *et al.*, 1988; Jongman *et al.*, 2000; Tomiak, 1990). A second associated area of research has utilized spectral moments analysis to examine how the acoustic properties of stop bursts and/or fricatives change as a function of normal speech development and aging (e.g., Forrest *et al.*, 1990, 1994; Miccio, 1996; Nittrouer, 1995; Nittrouer *et al.*, 1989).

One of the earliest studies to utilize spectral moments data in discriminant classification of obstruents was conducted by Forrest *et al.* (1988). The authors reported that adult productions of voiceless stops and fricatives could be discriminated by an analysis of the first (mean), third (skewness), and fourth (kurtosis) spectral moments. Using discriminant functions based on linear data from a single window of analysis, stops were classified with an overall average of 79.9% and fricatives slightly lower at 74.5%. However, the authors concluded that the second spectral moment of variance did not make a significant contribution to the discrimination of different voiceless obstruents. Interestingly, subsequent spectral moments studies of child productions

(e.g., Forrest *et al.*, 1990; Nittrouer, 1995) have not included the measure of spectral variance in their analyses. Some researchers (Jongman *et al.*, 2000) have suggested that the disregard for the second spectral moment in recent literature may be due in part to this original finding by Forrest *et al.*

A more recent investigation of adult fricative productions (Jongman *et al.*, 2000) found that all four spectral moments (mean, variance, skewness, and kurtosis) could be used to distinguish among the fricatives in terms of place of articulation. Unlike previous spectral research on fricatives (e.g., Forrest *et al.*, 1988), Jongman *et al.* reported that spectral variance does differ significantly as a function of place of fricative articulation (labiodental, dental, alveolar, and palato-alveolar) and likewise contributes to the spectral classification of fricatives in adults. The authors reported that, in terms of place of articulation, the differences between four voiceless fricative productions /f, θ, s, ʃ/) remained robust even when comparing across differing window locations. Discriminant analysis based on spectral peak location and spectral moments at four locations for each fricative demonstrated a classification accuracy of 77%.

A second associated line of research has employed spectral moments analysis (i.e., mean, skewness, and kurtosis) to investigate the acquisition and development of speech in children. Several studies (Forrest *et al.*, 1994, 1990, 1988; Miccio, 1996; Nittrouer *et al.*, 1989) have documented age-related differences in the development of children's obstruent productions. Of particular interest has been the development of the voiceless /t/ vs /k/ stop contrast, as well as the fricative contrast of /s/ vs /ʃ/.

An influential spectral moments study on the developmental order of obstruent acquisition was conducted by Nittrouer in 1995. Following the methodology outlined in Forrest *et al.* (1988), Nittrouer examined the acoustic patterns of several obstruents (/t, k, s, ʃ/) as revealed by the first, third, and fourth spectral moments. Findings indicated that some of the obstruent articulatory gestures of children are not as precisely specified as those of adult speakers. From a developmental perspective, these results provide support for the theory that acoustic differences between the obstruent productions of young children and adults may be related to articulatory changes that occur with maturation. Interestingly, Nittrouer found that some developmental differences vary as a function of the place and type of obstruent contrast. For example, Nittrouer found that the difference in spectral mean values between /s/ and /ʃ/ was larger for adult speakers than for children ages 3 through 7, yet the differences between /t/ and /k/ were relatively similar. The author interpreted this finding as support for the conclusion that in terms of the first spectral moment the child speakers were continuing to "fine tune" sibilant fricative articulations toward an adult-like contrast.

Researchers have established a foundation for the effectiveness of spectral moments analysis as a means of evaluating the acoustic nature of obstruent productions. However, despite the above-mentioned findings, there is still a limited amount of knowledge on the acoustic structure and developmental nature of young children's fricative productions. Many findings reported with adult speakers (e.g., Jongman *et al.*, 2000) have yet to be fully addressed with younger speakers, especially children younger than 6 years of age. Thus, the specific aims of this study are threefold:

(1) Describe the acoustic structure of voiceless fricative productions of adults and typically developing children in terms of multiple acoustic parameters, including several measures (i.e., spectral slope and variance) and fricative types (i.e., /f/ and /θ/) not commonly included in spectral studies involving children.

(2) Examine to what extent the individual acoustic characteristics of voiceless fricative productions change as a function of age, gender, place of articulation, and vowel context.

(3) Utilize discriminant analysis to determine how well combinations of acoustic parameters based on specific groups of speakers successfully categorize fricative productions in terms of place of articulation.

II. METHODOLOGY

A. Subjects

Three groups of children between the ages of 3 and 6 years ($N=30$) and one comparison group of adults were recruited to participate in the present study ($N=10$). All subjects were monolingual speakers of American English and had minimal exposure to a second language. No children had a diagnosed history of speech, language, or hearing problems. At the time of their participation in the study all subjects had visible front incisors and were required to pass a hearing screening and an oral/motor exam. Prior to recording, the phonemic inventory of each child was informally evaluated by a certified SLP using the "Sounds-in-Words" subtest of the Goldman-Fristoe Test of Articulation (GFTA: Goldman and Fristoe, 1986). All children who participated in the study exhibited target appropriate fricatives, as measured by the GFTA. Since many of the targeted phonemes are typically acquired between the ages of 3 and 6, no subject was excluded from the study based on their inability to produce the target phonemes in a medial or final word position.

B. Stimuli

The corpus of elicited productions consisted of monosyllabic words with an initial syllable CV(C)(C) containing a combination of four different voiceless obstruents (/f, θ, s, ʃ/) in initial position followed by three different monophthongal vowels (/i, ɑ, u/). Specifically, the corpus included the following words: *field, fox, food, thief, thought, "Thoot," seal, sock, soup, sheep, shark*, and *shoe*. Each stimulus item was repeated five times; thus, the entire corpus of elicited productions yielded a total of 60 tokens per subject. To elicit relatively similar vocal emphasis across different productions, the targeted CV(C) syllable combinations were always in the initial and stressed position of each elicited word. Stimulus words in this set were produced in the carrier phrase "This is a ____." In one instance, an age-appropriate word did not exist that exemplifies the targeted syllable. For example, American English does not contain an age-appropriate word beginning with a /θu/ consonant-vowel combination. To ad-

dress this lexical "gap," an invented proper name that follows the phonotactic rules of Standard English was substituted for the missing word. In this case, subjects were instructed to produce the proper name "Thoot" in reference to a fictional character.

C. Acoustic analysis

Speech samples were recorded online to computer in a quiet room environment. More specifically, a low impedance dynamic microphone (Shure SM10A-CN) and a preamplifier (Samson Mixpad-4) were used to facilitate the recording of subject productions. The microphone was affixed to a headset and placed approximately 4 cm from the speaker's lips during recording. The speech tokens were sampled at a rate of 44.1 kHz with a quantization of 16 bits and low-pass filtered at 22.05 kHz.

The productions were elicited from all subjects through picture identification. Pictures representing the target words were presented in a randomized order (between presentations and across subjects) on a 15 in. computer screen. All pictures utilized in the study were selected to be age-appropriate for preschool aged children. The subjects were familiarized with the names of the pictures and the elicitation procedure prior to the recording session. If a subject incorrectly identified a picture as a different lexical item during the recording session, the correct target word was modeled by the experimenter and the child was instructed to repeat the identification of that particular item.

Segmentation of the onset and offset of the fricative target was conducted using waveform display assisted by spectrographic inspection (Adobe Systems Incorporated, 2003). The fricative onset was characterized by a rapid increase in zero-crossings and/or spectrographic identification of high frequency energy, whereas the offset was the intensity minimum prior to the onset of vowel periodicity and/or absence of high frequency energy (Jongman et al., 2000). Segmentation values were then recorded into a text file [in milliseconds (ms)] and later checked and corrected (and then rechecked) using a MATLAB program that displayed the segmentation marks superimposed over a display of the token's waveform. In addition, to test for segmentation accuracy and reliability, 240 randomly chosen tokens (one subject from each age group) were independently analyzed by a second individual. The fricative boundary measurements were subsequently correlated ($r=0.99, p<0.001$) to the original segmentation values extracted for these same tokens and were found to differ by an average of approximately 3 ms.

Acoustic and statistic analysis followed procedures outlined in previous research examining the acoustic properties of voiceless fricatives (i.e., Forrest et al., 1988; Jongman et al., 2000; Nittrouer, 1995). Fricative durations were reported in ms increments from raw time points extracted during segmentation. A measure of normalized amplitude was computed for each fricative segment. This measure was the rms amplitude in decibels of the entire segment of friction minus the rms amplitude of the strongest component within the initial 40 ms of the following vowel (Behrens and Blum-

stein, 1988a, b; Jongman et al., 2000). This type of amplitude measure served to normalize for differences in speaker intensity.

A series of three 40 ms Hamming windows, located at the beginning, middle, and end of each target, was used to analyze the acoustic structure of the fricative productions. Moments were only calculated for fricative tokens that were at least 80 ms in duration (following this rule one token was eliminated from the analysis). Each window interval was then pre-emphasized by first-differencing. Though the need for pre-emphasis is minimized when analyzing voiceless sounds, it was determined that such a procedure was necessary to more effectively compare subsequent results to previously published findings (e.g., Forrest et al., 1988; Jongman et al., 2000; Nittrouer, 1995). Using a 2048-point FFT with zero-padding, the spectral amplitudes of a series of frequency points were derived from the complex acoustic signal within each 40 ms window. The resulting power spectra were considered random distribution probabilities, from which the spectral moments were computed. The third and fourth spectral values were subsequently normalized by variance. These normalization procedures, as well as the other algorithms utilized in this study are specifically described in previous spectral moments studies (e.g., Forrest et al., 1988; Nittrouer, 1995). Measures of spectral slope were derived from the power spectra (1 to 15 kHz) generated during the spectral moments analysis. The slope was derived from a linear regression line fit to the relative amplitudes extracted from each analysis window, specifically describing the amplitude slope of the FFT power spectra.

All elicitation, recording, and analysis of stimuli were facilitated by custom designed computer programs (MATLAB) created by the authors. A corpus of test tokens comprised of known acoustic components was utilized to evaluate the accuracy and reliability of these computer programs. For example, a test token composed of several sinusoidal frequencies (1, 3, and 5 kHz) of equal amplitude was analyzed by the computer programs and found to have the appropriate values for the various acoustic measures.

D. Statistical analysis

Prior to statistical analysis all data were collapsed across repetitions of a given stimulus item. Spectral mean values were then transformed to a perceptually normalized scale, specifically, the Equivalent Rectangular Bandwidth (ERB) scale (Glasberg and Moore, 1990; Moore, 1997). The ERB auditory scale is a recently developed psychophysical metric, which employs a "notched-noise" method rather than traditional masking procedures to measure the auditory filter bandwidth of the human auditory system. The ERB metric is somewhat similar to the previously developed Bark scale (Zwicker, 1975; Zwicker and Terhardt, 1980), except that for lower frequencies, the slope of the ERB function decreases with decreasing center-frequency, while the Bark scale remains nearly linear.

Repeated measures analyses of variance (ANOVA) were then used to determine significant acoustic variation in the fricative productions as a function of place of articulation,

vowel context, gender, and age group. Results of significant *F*-tests include a measure of effect size, partial eta squared, or η^2 (the value of η^2 can range from 0.0 to 1.0, and can be considered a measure of the proportion of variance explained by a dependent variable when controlling for other factors). Greenhouse-Geisser adjustments were utilized to adjust *F*-tests with regard to degrees of freedom when significant deviations from sphericity were found. Furthermore, pairwise comparisons for significant within-subject factors were done using General Linear Model repeated-measures contrasts, comparison significance was determined using the appropriate *F*-tests (see SPSS, 1997). *T*-test statistics were reported for all significant between-subject post hoc results.

To allow comparison to the study by Jongman *et al.* (2000), a series of liner discriminant analyses were completed determining how well a set of acoustic measures obtained categorized fricatives from each group. Discriminant analysis is an application which derives a set of linear equations (discriminant functions) that classify individual cases (fricatives) into one of several mutually exclusive groups (place of articulation) based on their values for a set of predictor variables (acoustic measures). The analysis method utilized here selects predictor variables used in the discriminant functions on a stepwise basis (in terms of which of the available acoustic variables will significantly minimize Wilkes Λ). When all such variables have been selected, the stepwise procedure stops, leaving out of the discriminant functions all variables that do not significantly contribute (at the 0.01 level as determined by an *F*-test) to the categorization of the fricatives. A second use of Wilkes Λ is in an overall chi-square test of significance of the analysis; all the discriminant results to be discussed were significant at 0.01 or better.

Discriminant analysis allows us to examine how well a categorization model based on a set of acoustic parameters could classify individual fricative tokens by place of articulation. The accuracy of the classification was considered as the percentage of cases that were correctly classified into groups based on a particular discriminant function. These functions differed in terms of the acoustic parameters originally entered into the stepwise linear analysis and the particular individual or group of individuals upon which the function was developed. The discriminant functions were developed using both the cross-validation ("jack-knife") procedure and an approach that selects a subset of the data to train the classification model which is then used to classify the fricatives from the remaining data (the "specific group" approach).

In the cross-validation approach, data from one subject was excluded and the remaining data used to estimate a discriminant function for that particular subject, thus each case is classified on the basis of the discriminant functions derived from all cases other than itself. This process is repeated until each set of individual subject data has been successfully or unsuccessfully classified. When the number of individual cases is relatively small, this type of discriminant procedure provides a more accurate classification (Forrest *et al.*, 1990).

In the specific group approach, the overall data set is split into two subsets. The first subset is employed to develop or "train" a discriminant function and the other subset is then used to validate the resulting function. In our use of this procedure, discriminant functions were trained on only the adult data, the adult male data, and the adult female data. The resulting discriminant functions were then applied to specific speaker groups in the remaining data set. Predictor variables in both types of analysis included the duration measures, normalized amplitude, spectral slope, and spectral moments (mean, variance, skewness, and kurtosis).

III. RESULTS

Table I gives the acoustic measures (durations, normalized amplitude, spectral slope, spectral moments) from the four classes of voiceless English fricatives. Spectral moment values were taken from the analysis window centered over the temporal midpoint of the fricative. The data are tabularized according to speaker age, fricative place of articulation, and vowel context.

A. Fricative duration

For the measure of fricative duration, a main effect of place of articulation $[F(3,96)=14.16, p<0.001, \eta^2=0.31]$ was found to be significant. Subsequent pairwise comparisons indicated that the basis for this effect was derived from the significantly decreased ($p<0.001$) duration of labiodental fricatives /f/. Collapsed across speaker and vowel context, the mean fricative durations for /f, θ s, ∫/ were found to be 182, 205, 205, and 199 ms, respectively. There was also a significant place by age interaction effect $[F(9,96)=3.14, p<0.003, \eta^2=0.23]$. Adult speakers were found to have relatively shorter fricative durations than the groups of child speakers when the primary constriction of the consonant was more anterior in the oral cavity (/f/ and /θ/). However, this trend changed for alveolar and palato-alveolar fricatives (/s/ and /∫/), with the 5-yr-old speakers exhibiting the shortest durations.

In addition, the ANOVA also revealed a significant difference in fricative duration as a function of vowel context $[F(2,64)=35.30, p<0.001, \eta^2=0.52]$. Pairwise comparisons indicated that the significant differences ($p<0.001$) between contexts were characterized by a decrease in the fricative duration when followed by an /ɑ/ vowel. A significant difference ($p=0.05$) between the /i/ and /u/ vowel contexts was also found. The mean fricative durations for the /i/, /ɑ/, and /u/ vowel contexts were 200, 186, and 208 ms, respectively.

B. Normalized amplitude

A main effect of place $[F(3,96)=157.69, p<0.001, \eta^2=0.83]$ was obtained for normalized amplitude. As expected, Bonferroni-adjusted comparisons indicated that this effect was derived primarily from significant differences ($p<0.001$) between the nonsibilant and sibilant fricatives. The normalized amplitude of the nonsibilant fricatives (−13.7 dB for /f/, −11.9 dB for /θ/) was significantly less than that of the sibilant fricatives (−3.6 dB for /s/, −3.0 dB for /∫/).

Significant differences in the normalized amplitude of the fricative also depended on the articulation of the following vowel $[F(2,64)=16.08, p<0.001, \eta^2=0.33]$. Overall,

TABLE I. Acoustic measures from four classes of voiceless English fricatives, grouped as a function of speaker age, fricative place of articulation, and vowel context. Average durations in milliseconds (Dur), normalized amplitude in dB (NAmp), spectral slope (Slope), and the first four spectral moments (mean in ERB, M1; variance in MHz, M2; skewness, M3; and kurtosis, M4).

	Labiodental—/f/			(Inter)dental—/θ/			Alveolar—/s/			Palato-alveolar—/ʃ/		
	/i/	/ɑ/	/u/	/i/	/ɑ/	/u/	/i/	/ɑ/	/u/	/i/	/ɑ/	/u/
3 yr old												
Dur	203	198	201	233	216	260	215	211	240	204	210	224
NAmp	−15.88	−17.99	−13.72	−15.27	−16.69	−13.90	−5.66	−7.90	−4.72	−3.97	−3.61	−3.33
Slope	2.58	3.24	2.95	2.72	2.96	2.30	3.96	4.81	4.60	8.08	8.84	8.36
M1	27.94	28.20	28.80	28.94	29.29	28.94	29.06	30.67	30.74	31.29	30.78	31.21
M2	7.09	6.33	6.00	6.20	6.09	6.17	3.02	2.95	2.74	3.26	3.83	3.83
M3	−2.37	−2.32	−2.26	−2.19	−2.21	−2.02	−2.32	−2.08	−2.25	−0.92	−0.69	−0.92
M4	4.40	4.07	4.41	3.43	4.15	3.83	4.60	3.20	4.71	1.46	0.09	0.26
4 yr old												
Dur	187	180	207	214	202	217	215	206	208	197	168	218
NAmp	−11.84	−15.56	−11.29	−10.66	−11.40	−7.29	−2.57	−3.93	−2.79	−2.43	−2.81	−3.00
Slope	3.23	3.14	3.33	3.87	3.82	3.72	4.59	5.35	5.45	8.97	8.70	8.11
M1	28.85	28.63	27.26	29.11	30.60	29.88	30.88	30.93	30.70	31.77	31.13	31.19
M2	6.61	6.12	6.34	5.43	3.81	4.79	2.39	2.44	2.68	3.46	3.75	3.42
M3	−2.20	−2.32	−2.46	−2.36	−2.35	−2.30	−2.20	−2.02	−2.17	−0.45	−0.34	−0.49
M4	3.03	4.08	4.69	4.51	4.75	4.15	4.48	3.58	4.34	−0.09	−0.18	0.53
5 yr old												
Dur	195	158	174	206	188	210	201	180	189	187	173	214
NAmp	−11.63	−15.97	−11.55	−11.86	−12.26	−9.76	−2.64	−2.72	−2.49	−3.41	−3.74	−2.84
Slope	4.03	3.32	3.51	3.19	3.74	2.99	5.32	5.99	6.36	10.06	9.73	10.26
M1	28.48	28.17	27.92	29.78	30.22	29.51	31.27	31.93	31.67	30.82	30.14	30.10
M2	6.02	6.71	6.52	5.91	4.50	4.39	1.73	1.86	2.40	3.98	3.87	3.55
M3	−2.25	−2.27	−2.40	−2.22	−2.25	−2.35	−1.80	−1.79	−1.82	−0.02	0.29	0.27
M4	3.34	3.54	4.27	4.26	4.01	4.51	2.92	2.91	2.93	−0.58	−0.33	−0.29
Adult												
Dur	173	153	162	173	161	188	200	185	209	203	181	212
NAmp	−10.26	−15.76	−12.68	−10.67	−13.29	−10.90	−2.56	−2.81	−2.99	−1.91	−2.16	−2.78
Slope	4.24	3.57	3.89	4.30	3.69	3.55	6.02	6.26	6.88	9.71	8.85	9.39
M1	30.47	30.49	30.42	30.40	29.92	30.08	32.90	32.84	32.52	27.65	27.34	27.38
M2	5.84	6.13	5.44	5.22	6.29	5.81	1.85	1.91	2.81	2.31	2.28	2.17
M3	−1.91	−1.88	−2.21	−2.00	−1.83	−2.11	−1.59	−1.44	−1.20	1.80	2.14	1.92
M4	2.91	2.25	4.38	2.84	1.78	3.02	3.87	3.13	1.89	4.54	7.18	4.93

the normalized amplitude measures for the fricatives were significantly ($p < 0.001$) decreased when followed by an /ɑ/ vowel. The normalized values for the fricatives in the /i, ɑ, and u/ vowel context were −7.7, −9.3, and −7.2 dB, respectively. A place by vowel interaction was also found to be significant [$F(6,192) = 42.86, p < 0.001, \eta^2 = 0.18$]. The decreased normalized amplitude found in the /ɑ/ vowel context was significant ($p < 0.01$) for /f, θ, and s/, but was actually slightly increased for /ʃ/.

C. Spectral measures

To better compare these results with previous findings concerning the spectral properties of young children's speech (i.e., Nittrouer, 1995), this study will present detailed findings from only the analysis window centered at the midpoint of the fricative segment. It is hoped that since the characteristics of the fricative are of primary interest in this study, utilizing the middle analysis window will limit the effects of stimulus onset variations and the formant transitions of the

subsequent vowel. However, it is important to report that the patterns of spectral differences across window locations were similar to those that have been previously found with adult speakers (Behrens and Blumstein, 1988a; Jongman et al., 2000). For example, window locations at the beginning and end of the fricative were generally found to have higher spectral peak and mean values than measures extracted from the middle portion of the fricative.

1. Spectral slope

A significant main effect of place of articulation [$F(3,96) = 266.16, p < 0.001, \eta^2 = 0.89$] was obtained for spectral slope. Subsequent pairwise comparisons ($p < 0.001$) demonstrated that three of the four places of fricative articulation were statistically different in terms of mean spectral slope, with no significant difference noted between nonsibilant fricatives (/f/ and /θ/). The mean spectral slope values for the different fricative types (/f, θ, s, ʃ/) were 3.42, 3.40, 5.46, and 9.08, respectively. In addition, the statistical

analysis indicated that when collapsed across age group and vowel context, the mean spectral slope of the four fricative types varied significantly $[F(3,96)=10.31, p<0.001, \eta^2=0.24]$ as a function of speaker gender. Female speakers exhibited higher slope values for all places of fricative articulation, except for the alveolar fricatives (/s/).

Interestingly, a main effect was obtained for age group $[F(3,32)=4.84, p<0.01, \eta^2=0.31]$, as well as a significant gender by age group interaction $[F(3,32)=3.08, p<0.05, \eta^2=0.22]$. Spectral slope values rise as the age of speaker increases (the mean slope values for 3-, 4-, 5-yr-old, and adult speakers were 4.61, 5.18, 5.70, and 5.86, respectively). Significant differences were found between the two youngest groups and the adult speakers [3 yr old, $t(9)=-3.42$, $p=0.008$; 4 yr old, $t(9)=-3.43$, $p=0.007$]. However, the extent of these differences between age groups varies widely as a function of gender. For males, the spectral slope increased with the age of the speaker. However, for female speakers, the 4-year-old age group exhibited the lowest slope values and the 5-yr-old group the highest.

2. Spectral mean

As expected, the spectral mean differed significantly across the place of fricative articulation $[F(3,96)=25.47, p<0.001, \eta^2=0.44]$. Pairwise comparisons indicated variation ($p<0.001$) between three of the four places of fricative articulation, with no statistical difference in spectral mean between nonsibilant /θ/ and /ʃ/ fricative tokens. There was also a significant place by gender interaction effect $[F(3,96)=4.46, p=0.008, \eta^2=0.12]$. This interaction is characterized by higher spectral means for male speakers on the three more anteriorly produced fricatives (/f, θ, s/) and a lower spectral mean for the more posterior /ʃ/ fricative. It is somewhat surprising to find that the male speakers exhibited elevated spectral means in the majority of fricative types. However, this result is somewhat explained by the finding that the elevated means for male speakers were only found in the child speakers. Since anatomical differences in vocal tract size are likely insignificant for child speakers of this age (Fitch and Giedd, 1999), gender differences may be a product of learned behavioral factors. However, data from future studies must be obtained before such a claim can be fully justified.

Similar to previous research (Nittrouer, 1995), a significant place by age group $[F(9,96)=10.90, p<0.001, \eta^2=0.51]$ interaction was noted for the dependent measure of spectral mean. As can be seen in Fig. 1, a contrast between /s/ and /ʃ/ was exhibited by the 5-year-old speakers and further differentiated by the adults, with the spectral mean of /ʃ/ being significantly lower than /s/. Interestingly, this within-sibilant contrast is not evident in the younger groups of speakers (3 and 4 yr old).

3. Spectral variance

A significant main effect of place $[F(3,96)=50.48, p<0.001]$, characterized by a relatively large effect size ($\eta^2=0.61$), was obtained for spectral variance. Subsequent pairwise comparisons indicated that all four fricative types (/f, θ, s, ʃ/) were significantly ($p<0.001$) different in terms of vari-
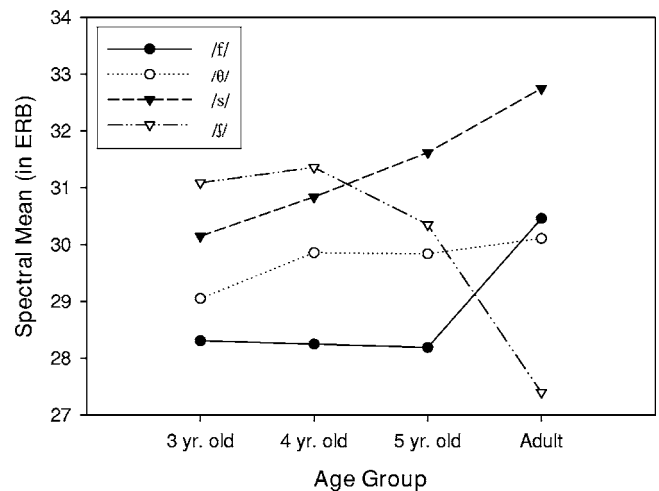

FIG. 1. Spectral mean as a function of place of fricative articulation and speaker age. Linear measures in hertz were converted to an ERB scale (Glasberg and Moore, 1990; Moore, 1997) prior to analysis.

ance, with means of 6.26, 5.38, 2.39, and 3.30 MHz, respectively. When collapsed into two groups according to sibilance, the measure of spectral variance was found to significantly ($p<0.001$) differentiate between sibilant and nonsibilant fricatives.

4. Spectral skewness

As with many of the other spectral measures, statistical analyses of the data also indicated a significant main effect of place $[F(3,96)=242.97, p<0.001, \eta^2=0.88]$ for spectral skewness. Subsequent pairwise comparisons indicated that the main effect of place was due to the significantly ($p<0.001$) elevated skewness of palato-velar fricatives. Skewness values for the four fricatives types (/f, θ, s, ʃ/) were found to be −2.23, −2.18, −1.88, and 0.21, respectively.

In addition, a significant main effect of age group $[F(3,32)=24.96, p<0.001, \eta^2=0.70]$ and a significant place by age group $[F(9,96)=12.37, p<0.001, \eta^2=0.54]$ interaction effect was noted. As seen in Fig. 2, post hoc tests indicated significant differences between each of the groups of
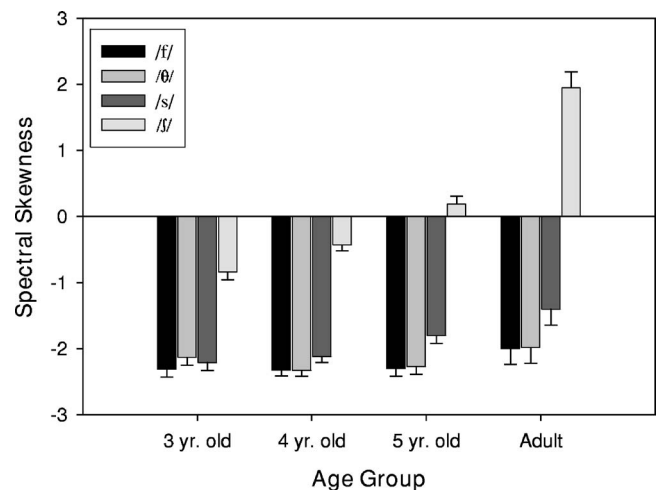

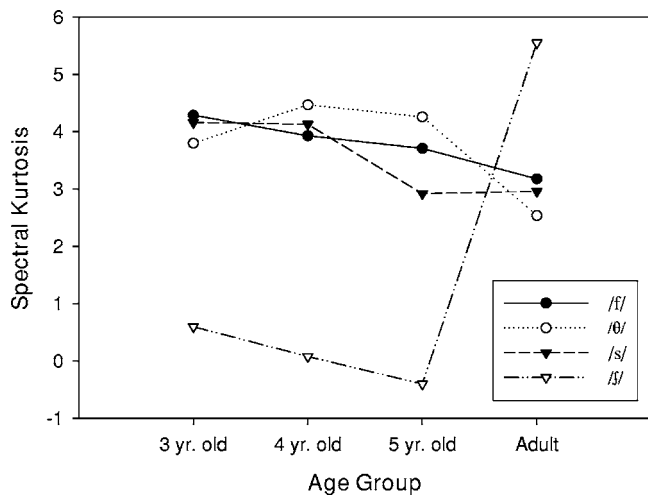FIG. 2. Spectral skewness as a function of speaker age group and place of fricative articulation.

Nissen and Fox: Acoustic/spectral characteristics of children's fricatives 2575

FIG. 3. Spectral kurtosis as a function of speaker age and place of fricative articulation.

TABLE II. Results of discriminant classification using both cross-validation and "specific group" procedures. All values represent % correct classification.

| | Classification model training set | | | |
Data set evaluated	Cross validation	All adults	Male adults	Female adults
Original training set	⋯	76.7	85	83.3
3-yr-old speakers	70	34.2	28.3	42.5
4-yr-old speakers	70	38.3	40	48.3
5-yr-old speakers	77.5	54.2	55.8	57.5
Adult speakers	65	⋯	⋯	⋯
Male only	75.0	⋯	⋯	60.0
Female only	78.3	⋯	61.7	⋯
Entire data set	69.2	42.2	44.3	51.0

children and the adults [3 yr old, $t(9)=5.51$, $p<0.001$; 4 yr old, $t(9)=7.18$, $p<0.001$; 5 yr old, $t(9)=3.94$, $p=0.003$], with no differences between the child groups found to be significant. Overall differences in age group are further explained by examining the place by age group interaction. Similar to findings reported by Nittrouer (1995), the contrast in spectral skewness between /ʃ/ and the other fricative types widened as the age of the speaker increased, with adult speakers exhibiting more positively skewed spectra for /ʃ/.

5. Spectral kurtosis

The main effect of place of articulation was statistically significant for spectral kurtosis [$F(3,96)=266.16$, $p<0.001$, $\eta^2=0.89$]. Pairwise comparisons demonstrated that only two of the four places of fricative articulation were statistically different ($p<0.001$) from each other in terms of mean spectral kurtosis. These differences were due mainly to the decreased kurtic value of palato-velar fricatives, with the mean kurtic values for the four fricative types (/f, θ, s, ʃ/) found to be 3.78, 3.77, 3.54, and 1.46, respectively. In addition, a significant place by gender interaction [$F(3,96)=10.31$, $p<0.001$, $\eta^2=0.24$] and subsequent pairwise comparisons indicated that for alveolar fricatives (/s/) the values of spectral kurtosis for male speakers were significantly lower ($t(19)=-3.92$, $p<0.001$) than female speakers (2.21 vs 4.90). Significant differences in gender were not found for the other fricative types.

As shown in Fig. 3, the ANOVA also yielded a significant place by age group interaction [$F(9,96)=4.64$, $p<0.001$, $\eta^2=0.30$]. Although an overall main effect of age group was not found to be significant, the interaction reveals significant age difference for each child age group when compared to the adult speakers [3 yr old, $t(9)=2.85$, $p=0.019$; 4 yr old, $t(9)=2.92$, $p=0.017$; 5 yr old, $t(9)=3.55$, $p=0.006$] for palato-velar fricatives (/ʃ/). This finding may indicate that with regard to kurtosis, the child speakers have not yet acquired an adult-like palato-velar fricative. Similar age differences concerning the measure of spectral kurtosis have not been reported in previous research (i.e., Nittrouer, 1995).

D. Discriminant analysis

As shown in Table II, when classification results are based on a cross-validation (jack-knife) procedure, 69.2% of the entire data set were classified correctly in terms of place of articulation. The adult speakers had a classification rating of 65.0%, while the 3-, 4-, and 5-yr-old speakers exhibited a rating of 70.0%, 70.0%, and 77.5%, respectively. Overall, differences between male (67.5%) and female (68.8%) speakers were minimal.

Utilizing the specific group approach, the discriminant analysis trained on the adult fricative productions resulted in two significant discriminant functions [($\Lambda=0.03$, $\chi^2(15,N=120)=416.81$, $p<0.001$) and ($\Lambda=0.32$, $\chi^2(8,N=120)=129.98$, $p<0.001$), respectively]. The two canonical functions were found to cumulatively accounting for 99.9% of the variance between productions. The first discriminant function had an eigenvalue of 11.24 and a canonical correlation of 0.96, thereby accounting for 84.4% of the variance. The second function accounted for 15.5% of variance, with an eigenvalue of 2.07 and an associated correlation of 0.82.

Of the seven input variables utilized in the analysis, five acoustic parameters (spectral mean, variance, skewness, kurtosis, and slope) were found to have a substantial contribution to the overall classification of the adult fricative tokens. Thomas (1992) found that the partial contribution of each of the selected variables to a particular discriminant function can be characterized by discriminant ratio coefficients (DRCs). The DRCs are calculated as the product of the absolute values of the structure coefficients and the corresponding standardized discriminant function coefficients. Thus, the larger the DRC, the greater the unique contribution of that respective variable to the discrimination between fricative types in terms of place of articulation. As shown in Table III, the acoustic variables of spectral skewness, variance, and slope were found to have the greatest relative contribution to the first discriminant function, whereas spectral mean, variance, and slope were found to be the top three contributors to the second discriminant function.

Classification models developed on adult-only productions resulted in an overall classification rate of 76.7% for data in the training set. High classification rates were found for the sibilant fricatives (95%), while the nonsibilant frica-

Nissen and Fox: Acoustic/spectral characteristics of children's fricatives

TABLE III. Discriminant variable coefficients for adult fricative data. Input variables are ordered according to their relative contribution to the discriminant function as measured by the DRCs (Thomas, 1992).

	First discriminant function				Second discriminant function		
Input variable	SDFCs[a]	SCs[b]	DRCs[c]	Input variable	SDFCs[a]	SCs[b]	DRCs[c]
Spectral skewness	−0.550	−0.666	0.366	Spectral mean	0.194	0.146	0.028
Spectral variance	0.915	0.258	0.236	Spectral variance	−0.375	−0.536	0.201
Spectral slope	−0.458	−0.506	0.232	Spectral slope	0.130	0.284	0.037
Spectral mean	0.598	0.312	0.187	Spectral skewness	0.845	0.920	0.777
Spectral kurtosis	0.274	0.073	0.020	Spectral kurtosis	−0.378	−0.034	0.013

[a]Standardized discriminant function coefficients.
[b]Structure coefficients.
[c]Discriminant ratio coefficients (DRC=SDFC×SC).

tives were correctly identified at a lower rate (70%). Errors in classification were usually the result of confusions within "neighboring" fricatives (e.g., /f/ vs /θ/) and rarely crossed the sibilant/nonsibilant distinction. When applied to data produced by child speakers, the classification model categorized fricatives with an overall accuracy of only 42.2%. Interestingly, the rate of successful classification improved as the age of the child speakers increased, with the mean classification rates for the 3-, 4-, and 5-yr-old age groups found to be 34.2%, 38.3%, and 54.2%, respectively. The higher classification rates for productions from the 5-year and adult speakers are primarily due to the increased categorization of the sibilant fricatives. For the two youngest groups of speakers, the discriminant model accurately identified /s/ and /ʃ/ productions only slightly above chance (26.6%), whereas, the categorization for 5-yr-old and adult speakers was dramatically higher (65% and 95%, respectively).

IV. DISCUSSION

Statistical analysis (ANOVA) indicated that the dependent measures of fricative duration, normalized amplitude, spectral slope, and all four spectral moments varied significantly as a function of place of articulation. However, only spectral variance was found to significantly distinguish all four places of fricative articulation. Interestingly, with the exception of Jongman et al. (2000) and Tomiak (1990), the measure of spectral variance has often not been analyzed or reported in the previous literature (e.g., Forrest et al., 1988; Miccio, 1996; Nittrouer, 1995). In general, these results are in agreement with the research of Jongman et al., where it was found that spectral variance could be used to acoustically distinguish four places of fricative articulation in adult productions.

Although spectral variance was the only measure to distinguish all four fricative types, many of the derived acoustic parameters were found to successfully distinguish three out of four places of fricative articulation. The measures of spectral slope, mean, and skewness were able to separate nonsibilant from sibilant fricatives, as well as between the two sibilant fricatives (/s/ and /ʃ/). However, contrary to research findings on the acoustic characteristics of fricatives in adults (Jongman et al., 2000), results of this study indicated that the majority of spectral parameters (with the exception of variance) did not differ significantly between nonsibilant frica-

tives (/f/ and /θ/). It is important to note that the spectral measure of slope was found to be an important cue in the distinction and classification of fricative productions.

Of particular interest, significant place by age interaction effects were noted for the measures of spectral mean, skewness, and kurtosis. In each of these measures, the interaction effect was due primarily to a widening in the acoustic distinction between /s/ and /ʃ/ as the age of speaker increased. In terms of spectral mean, a significant difference between /s/ and /ʃ/ was exhibited by the 5-yr-old speakers and further differentiated by the adults. For spectral skewness, an adult-like contrast appears to begin developing in the 4-yr-old speakers and widened as the speakers increased in age. These findings suggest that for several acoustic parameters the development of the sibilant contrast continues to be fine-tuned throughout young childhood toward an adult-like stage of articulation. This same general trend across age groups was reported by Nittrouer (1995), who also found age-related acoustic differences. Nittrouer concluded that these differences were the product of continuing articulatory development, even for children 7 years of age. The lack of acoustic difference between the sibilants in the youngest groups of speakers may also be partially due to the relatively smaller size of the vocal tract. In adults, the difference between an alveolar and palatal place of fricative articulation is likely to result in comparatively large acoustic variations, which may not be the case for young children.

In terms of place of articulation, a discriminant analysis based on the extracted acoustic measures was able to correctly classify the fricative data at a rate well above chance. Interestingly, the spectral measures of variance and slope were found to be important coefficients in the classification of the fricative productions. It is important to note that based on a cross-validation procedure the overall classification accuracy for the fricative productions (69%) was similar to the classification results obtained by Forrest et al. (1988). As expected, discriminant functions based on the adult tokens were less accurate in classifying the segments produced by the younger speakers. For adults, the majority of classification errors occurred with nonsibilant fricatives, with few categorization confusions crossing the sibilant/nonsibilant distinction. Conversely, the discriminant function classified the nonsibilant productions from the youngest child speakers at a rate higher than that of the sibilant fricatives. However, it is

interesting to note that the overall classification rate improved systematically as the age of the child speakers increased, with a marked improvement occurring at 5 years of age. Analysis of the classification of specific fricative types indicated that this improvement was primarily the result of the increased categorization of sibilant fricatives. Thus, these data once again support the theory that young children's sibilant fricative articulations are not differentiated in an adult-like manner, but are continuing to be "fine-tuned" as the child matures (Nittrouer, 1995). Of further interest is that unlike preliminary research conducted on the fricatives of older children (Fox and Nissen, 2003), discriminant functions trained on only male adult and only female adult data did not reveal systematic gender differences in the productions of the child speakers.

Future spectral research involving children's speech elicited in more naturalistic settings and across less predictable contexts would likely extend our understanding of the acoustic nature of speech in young children. Since acoustic studies provide only a "post-production" view of articulation, more studies are also needed that combine both kinematic and acoustic analysis techniques in a more dynamic investigation into the nature of fricative production. Finally, our use of discriminant analysis was designed to allow a comparison between the different age and gender groups in terms of the acoustic characteristics of the fricatives and, in particular, in terms of how well these acoustic measures could be used to categorize these fricatives. However, there are two limitations to our approach. First, as in previous studies such as Jongman *et al.* (2000), all the acoustic measures utilized here represented only static acoustic cues. Clearly, any future study should also examine the utility of dynamic cues (e.g., the change in the spectral mean over the fricative's duration) in the discriminant functions. Second, one should determine whether listeners make use of the same acoustic cues (and whether they rely on them to the same degree) as those used in the discriminant analyses discussed. Despite these limitations, this study contributes valuable results concerning the spectral nature of a broad set of voiceless fricative productions from a substantial number of speakers; findings which serve to establish a more comprehensive knowledge of how fricative consonants develop in younger speakers.

ACKNOWLEDGMENTS

Adobe Systems Incorporated. (**2003**). Adobe Audition (Version 1.3) [Computer software], San Jose, CA, Adobe Systems Incorporated.

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

Behrens, S., and Blumstein, S. E. (**1988b**). "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants," J. Acoust. Soc. Am. **84**, 861–867.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Forrest, K., Weismer, G., Elbert, M., and Dinnsen, D. A. (**1994**). "Spectral analysis of target-appropriate /t/ and /k/ produced by phonologically disordered and normally articulating children," Clin. Linguist. Phon. **8**, 267–281.

Forrest, K., Weismer, G., Hodge, M., and Dinnsen, D. A. (**1990**). "Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children," Clin. Linguist. Phon. **4**, 327–340.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Fox, R. A., and Nissen, S. L. (**2003**). "Age-related acoustic changes in voiceless English fricatives," Proceedings from the Meikai Linguistics and Phonetics Conference, Meikai University, Japan.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldman, R., and Fristoe, M. (**1986**). *Goldman-Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Miccio, A. W. (**1996**). "A spectral moments analysis of the acquisition of word-initial voiceless fricatives in children with normal and disordered phonologies," Doctoral dissertation, Indiana University.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, New York).

Nittrouer, S. (**1995**). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," J. Acoust. Soc. Am. **97**, 520–530.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Hear. Res. **32**, 120–132.

SPSS, Inc. (**1997**). *SPSS Advanced Statistics* 7.5 [Computer Software Manual] (SPSS Incorporated, Chicago).

Thomas, D. (**1992**). "Interpreting discriminant functions: A data analytic approach," Multivar. Behav. Res. **27**, 335–362.

Tomiak, G. R. (**1990**). "An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents," Doctoral dissertation, SUNY Buffalo.

Zwicker, E. (**1975**). "Scaling," in *Handbook of Sensory Physiology*, Auditory System Vol. 3, edited by D. Keidel and W. Neff (Springer, Berlin), pp. 401–408.

Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.