



Faculty Publications

2009-09-01

Acoustic and Spectral Patterns in Young Children's Stop Consonant Productions

Shawn L. Nissen
Shawn_Nissen@byu.edu

Robert Allen Fox

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Communication Sciences and Disorders Commons](#)

Original Publication Citation

Nissen, S. L., & Fox, R. A. (2009). Acoustic and spectral patterns in young children's stop consonant productions. *The Journal of the Acoustical Society of America*, 126(3), 1369-1378.

BYU ScholarsArchive Citation

Nissen, Shawn L. and Fox, Robert Allen, "Acoustic and Spectral Patterns in Young Children's Stop Consonant Productions" (2009). *Faculty Publications*. 123.
<https://scholarsarchive.byu.edu/facpub/123>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Acoustic and spectral patterns in young children's stop consonant productions^{a)}

Shawn L. Nissen

Department of Communication Disorders (138 TLRB), Brigham Young University, Provo, Utah 84602

Robert Allen Fox

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

(Received 15 April 2009; revised 6 July 2009; accepted 7 July 2009)

The aim of this study was to examine the acoustic and spectral patterns of stop articulation in the speech of pre-pubescent children. A set of voiceless stop consonants, /ptk/, produced by a group of adults and typically developing children 3–5 years of age were examined in terms of multiple acoustic and spectral parameters. Findings indicated that, with the exception of spectral kurtosis, the acoustic and spectral characteristics of the stop productions varied significantly as a function of place of articulation and vowel context. Sex-specific differences in spectral slope, mean, and skewness were found for the 5-year-old and adult speakers. Such differences in adult speakers can be explained in part by variation in vocal tract size across the sex of the speaker; however, vocal tract dimorphism is typically not present in pre-pubescent children. Thus, the findings of this study provide some support that sex-specific differences in the speech patterns of young children may be associated with learned or behavioral factors, such as patterns of obstruent articulation that depend in part on a culturally determined male-female archetype.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3192350]

PACS number(s): 43.70.Ep, 43.70.Bk, 43.70.Fq, 43.70.Aj [BHS]

Pages: 1369–1378

I. INTRODUCTION

Researchers continue to search for more accurate models and descriptions of the complex process by which humans develop the ability to produce and perceive speech in a remarkably efficient and accurate manner. Advances in instrumentation and computer technology have allowed scientists to utilize more specific types of acoustic analysis to investigate speech communication. Spectral moments analysis (e.g., Forrest *et al.*, 1988; Jongman *et al.*, 2000; Nittrouer, 1995) is an analysis method which allows the extraction of a number of spectral characteristics from discrete time segments of the speech signal (including spectral mean, variance, skewness, and kurtosis), providing researchers with the means to identify unique patterns of acoustic energy.

Utilizing spectral moments analysis, as well as more traditional types of acoustic analysis, researchers have investigated the differences in acoustic structure of obstruents produced by phonologically disordered children (e.g., Forrest *et al.*, 1994; Miccio, 1996). An associated line of research has focused on changes in obstruent production as a function of typical speech development and maturation (e.g., Forrest *et al.*, 1990; Fox and Nissen, 2005; Miccio, 1996; Nissen and Fox, 2005; Nittrouer, 1995; Nittrouer *et al.*, 1989).

Following the methodology employed with adult speakers by Forrest *et al.* (1988), a study by Nittrouer (1995) utilized spectral moments analysis to investigate the nature of obstruent acquisition in children between 3 and 7 years of

age and a comparison group of adult speakers. Nittrouer (1995) examined the acoustic nature of voiceless stop (/t/ and /k/) and fricative (/s/ and /ʃ/) productions in terms of spectral mean, skewness, and kurtosis. It was found that the children exhibited significantly fewer distinctions and less contrast in spectral mean between the /s/ and /ʃ/ fricatives than the adult speakers. The author suggested that this age-related difference may indicate that the children were continuing to “fine-tune” their fricative articulations with developmental maturation. Interestingly, it was found that the stop consonant productions differed very little as a function of age in terms of spectral mean, indicating that developmental differences in speech production may vary as a function of both the age of the speaker and the manner of the obstruent contrast.

The second spectral moment of variance has typically not been utilized in earlier spectral studies of obstruent production in children (e.g., Forrest *et al.*, 1990, 1994; Nittrouer, 1995). The exclusion of the spectral variance measure from these studies may be due to the fact that Forrest *et al.* (1988) did not find a significant contribution of spectral variance to the discrimination of a set of voiceless obstruents produced by a group of adult speakers. However, several recent studies have found that the spectral variance, as well as the spectral slope, can be useful in discriminating between fricative contrasts (Jongman *et al.*, 2000; Fox and Nissen, 2005; Nissen and Fox, 2005).

Previous research has not only found that speech development varies as a function of age but also according to the sex of the speaker. For example, research investigating voice onset time (VOT) has found sex differences in the production of stop consonants. In general (see Sweeting and Baken,

^{a)}Portions of this work are contained in the unpublished doctoral dissertation, “An acoustic study of voiceless obstruents produced by adults and typically developing children,” The Ohio State University, 2003.

1982), adult female speakers of American English have been reported to exhibit significantly longer VOT values than adult males (e.g., Ryalls *et al.*, 1997; Swartz, 1992; Whiteside and Irving, 1997). Sex-related acoustic differences have also been noted in speaking rate (Byrd, 1992, 1994; Klatt and Klatt, 1990) and the occurrence of pauses (Whiteside, 1996). It has been shown that for some consonants, adult female speakers exhibit a greater degree of devoicing when compared to male speakers (Fant *et al.*, 1991; Whiteside, 1996). Moreover, research has revealed significant differences between male and female speakers in prosodic patterns (Fitzsimmons *et al.*, 2001; Graddol, 1986) and voice quality (Klatt and Klatt, 1990; Mendoza *et al.*, 1996).

Sex-related acoustic differences have also been noted in the speech of younger speakers. Several studies involving preadolescent children (Bennett, 1981; Busby and Plant, 1995; Perry *et al.*, 2001; Whiteside and Hodgson, 2000) have found that formant frequency patterns differed significantly as a function of speaker sex. A significant contribution to this area of research was a large scale study ($N=436$) by Lee *et al.* (1999) and a subsequent reanalysis of the data by Whiteside (2001). Although some researchers have attributed dissimilarities in formant patterns across speaker sex to non-uniform anatomical changes in the vocal tract size (e.g., Bennett, 1981; Fant, 1966), Lee *et al.* (1999) concluded that the acoustic variation in formant patterns across speaker sex could not be solely attributed to differences in vocal tract morphology.

Less emphasis has been placed on examining sex-related developmental differences in the productions of voiceless obstruents. A study examining voiceless fricative productions from children as young as 6 years of age found the presence of sex-related differences in multiple individual acoustic and spectral parameters, with findings from a discriminant analysis also indicating differences in combinations of speech parameters (Fox and Nissen, 2005). In addition, an examination of fricative productions from younger children, 3–5 years of age, also found sex-specific differences in spectral mean and slope (Nissen and Fox, 2005). However, it is unclear if such variation will be found in other types of obstruent productions (i.e., stop consonants) and at what age possible differences might occur. If sex-specific differences are learned phenomena, such differences may occur across phonetic classes of sounds.

Previous research has greatly enhanced the understanding of speech communication; however, the complex relationships between the acoustic structures of speech, as well as the manner and time in which they are acquired by children, have yet to be fully and adequately explained. Research examining stop consonant production in pre-pubescent children as a function of age and sex of the speaker, including several measures (i.e., spectral slope and variance) frequently not included in spectral studies involving children may lead to additional insights into the developmental nature of speech production. In particular, it is of interest to investigate if sex-specific differences noted with fricative production will also occur with stop consonant productions.

Considering recent large-scale magnetic resonance imaging (MRI) studies investigating sexual dimorphism in the

oral and pharyngeal portions of the vocal tract (Fitch and Giedd, 1999; Vorperian *et al.*, 2005, 2009), it is of interest to examine sex-specific differences in the speech acoustics of young children to developmental growth patterns of anatomic structures in the vocal tract. Such comparisons are important to establish anatomic-acoustic correlates (Vorperian *et al.*, 2009), as well as provide further insight into the source (anatomic or learned) of male-female speech differences. Thus, this study aims to describe the acoustic patterns (normalized amplitude, slope, mean, variance, skewness, and kurtosis) of the voiceless stop consonants, /p t k/, produced by a group of adults and typically developing children 3–5 years of age, as well as investigate to what extent the individual amplitude and spectral characteristics of the target productions change as a function of age, sex of the speaker, place of articulation, and vowel context.

II. METHODOLOGY

A. Participants

Participants included three groups of children between the ages of 3 and 5 years of age ($N=30$) and one comparison group of adults ($N=10$). Speakers in the 3-year-old group were between 3:0 and 3:11 years of age ($M=3:6$), the 4-year-old group were between 4:0 and 4:11 years of age ($M=4:8$), and the 5-year-old group contained children between 5:0 and 5:11 years of age ($M=5:7$). The adult subjects within the comparison group were between 18 and 40 years of age. Each group was composed of an equal number of male and female subjects. All participants were monolingual speakers of American English, with no diagnosed history of speech, language, or hearing problems. At the time of their participation all of the speakers exhibited pure-tone air-conduction thresholds ≤ 15 dB hearing loss at octave frequencies from 125 to 8000 Hz and had visible front incisors. Prior to recording, the phonemic inventory of each child was evaluated by a certified speech language pathologist 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 stop productions, as measured by the GFTA.

B. Stimuli

Target phonemes were elicited from a series of words with an initial syllable containing a combination of one of three voiceless obstruents (/p/, /t/, or /k/) in initial position followed by a monophthongal vowel (/i/, /a/, or /u/). Specifically, the corpus included the following words: *peanut*, *pocket*, *Poohbear*, *teapot*, *Thomas*, *toothbrush*, *key*, *car*, and *cougar*. Participants produced each word three times while embedded in the carrier phrase “This is a__ again.” The targeted syllable combinations were in the initial and stressed position of each word to elicit relatively similar vocal emphasis across productions. Occasionally the child participants produced the target word or carrier phrase in a dysfluent manner or incorrectly identified the picture as a different lexical item; in which case the utterance was rerecorded. As expected, the older children displayed fewer instances of dysfluent speech, as well as fewer misidentifications.

C. Elicitation procedures

The speech productions were recorded online to computer in a quiet room environment with a low impedance dynamic microphone (Shure SM10A-CN) and preamplifier (Samson Mixpad-4). The microphone was affixed to a headset and placed approximately 4 cm from the speaker's lips during recording. The productions were digitized at a sampling rate of 44.1 kHz and a quantization of 16 bits, and subsequently low-pass filtered at 22.05 kHz. Target productions were elicited from the participants through the verbal identification of age-appropriate pictures representing the target words. Custom software programmed in MATLAB was utilized to randomly present the elicitation pictures on a 15 in. computer screen and subsequently capture the participants' responses. The participants were familiarized with the names of the pictures and the elicitation procedure prior to the recording session by the test administrator modeling the procedure for the subjects prior to data collection. If a participant 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.

D. Acoustic and spectral analysis

Segmentation of the onset and offset of the obstruent target segments was conducted through waveform display assisted by spectrographic inspection using ADOBE AUDITION Version 1.3 (Adobe Systems Incorporated, 2003). The onset of the stop burst was characterized by a sharp increase in diffuse noise energy and the rapid increase in zero crossings, with the burst offset defined by a sharp decrease in diffuse noise energy. Segmentation values were then recorded into a text file (in milliseconds) and later checked, corrected, and re-checked 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, 540 tokens (three subjects randomly chosen from each age group) were independently analyzed by a second person and subsequently correlated ($r=0.99$, $p<0.0001$) to the original segmentation of these same tokens, differing by an average of approximately 1 ms.

A measure of normalized amplitude was computed for each stop burst, calculated by subtracting the root-mean-square (rms) amplitude in decibels of the segment of the stop burst from the rms amplitude of the strongest component within the initial 40 ms of the following vowel (Behrens and Blumstein, 1988a, 1988b; Jongman *et al.*, 2000). This amplitude measure served to normalize differences in speaker intensity.

Spectral moments measures (mean, variance, skewness, and kurtosis) were computed for the stop consonants following the approach of Forrest *et al.* (1988) and Nittrouer (1995). Normalized power spectra were derived from a 20 ms Hamming window centered +10 ms from the release of the stop burst, which was then pre-emphasized by first-differencing. Though the need for pre-emphasis is minimal 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, 1992, 1995). Using a 1024-point fast Fourier transform with zero-padding, the spectral amplitudes of a series of frequency points were derived from the complex acoustic signal within the 20 ms window. The resulting power spectra were considered random distribution probabilities, from which the first (mean), second (variance), third (skewness), and fourth (kurtosis) spectral moments were computed for each of the target stimuli. The third and fourth spectral values were subsequently normalized by the spectral variance for that same token. Measures of spectral slope were derived from the power spectra generated during the spectral moments analysis and calculated from a linear regression line fit to the extracted relative amplitudes of acoustic energy from 1 to 10 kHz. The slope values were reported as a ratio of amplitude to frequency (i.e., dB/kHz). 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; Fox and Nissen, 2005; Nissen and Fox, 2005; Nittrouer, 1995).

The stimuli in this study were elicited, recorded, and analyzed using 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 the spectral analysis. For example, a test token composed of several sinusoidal frequencies (1, 3, and 5 kHz) of equal strength was analyzed by the computer programs and found to have the appropriate values for the various acoustic measures.

E. Statistical analysis

Data were collapsed across repetitions of a given stimulus item and the spectral mean values were transformed to a perceptually normalized scale prior to statistical analysis. Specifically, the equivalent rectangular bandwidth (ERB-2) scale (Glasberg and Moore, 1990; Moore, 1997) was used to normalize the spectral mean measurements, thereby increasing the validity of comparisons across individual speakers. Repeated measures analyses of variance (ANOVAs) were used to examine possible acoustic differences in the stop productions as a function of place of articulation, vowel context, speaker sex, and age group. Results of significant *F*-tests include a measure of effect size (partial eta squared or η^2), which 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.

III. RESULTS

Detailed listings of the acoustic and spectral measures for the male and female speakers are found in Tables I and II, respectively. The data are tabularized according to speaker age, stop place of articulation, and vowel context.

TABLE I. Acoustic measures from male speakers for three classes of voiceless American English stop consonants, grouped as a function of speaker age, place of articulation, and vowel context. Normalized amplitude in dB (NAmp), spectral slope in dB/kHz (slope), and the first four spectral moments (mean in ERB, M1, variance in kHz², M2; skewness, M3; and kurtosis, M4).

	Bilabial-/p/			Alveolar-/t/			Velar-/k/		
	/i/	/a/	/u/	/i/	/a/	/u/	/i/	/a/	/u/
3 yr. old									
NAmp	-3.76	-11.96	-7.06	-2.70	-5.75	-3.93	-2.08	-7.38	-3.64
Slope	5.11	-5.11	-5.57	30.18	22.99	15.56	14.40	-20.90	-20.90
M1	28.36	28.06	27.29	30.95	30.97	30.07	29.56	25.85	26.46
M2	6.54	6.65	6.31	5.96	5.53	4.82	4.69	5.69	5.88
M3	-0.07	0.10	0.11	-1.20	-0.94	-0.25	0.03	0.72	0.96
M4	-0.03	0.29	1.00	0.92	0.51	0.51	0.05	2.06	1.92
4 yr. old									
NAmp	-7.26	-17.00	-10.35	-3.28	-5.19	-2.78	-2.83	-8.00	-3.89
Slope	2.32	-1.16	15.56	28.10	20.20	5.57	-5.34	-19.97	-26.93
M1	28.99	28.61	30.19	30.01	29.76	29.16	27.93	25.58	26.59
M2	6.81	7.47	6.75	5.55	4.79	5.26	3.21	6.66	5.18
M3	-0.36	-0.51	-0.92	-1.46	-0.32	0.03	1.14	0.82	1.12
M4	-0.45	0.11	0.10	1.44	0.78	1.34	4.78	1.11	2.47
5 yr. old									
NAmp	-6.48	-10.32	-7.63	-2.99	-5.30	-3.54	-1.44	-7.12	-5.58
Slope	-16.95	-12.77	-10.45	16.25	4.18	13.93	-3.02	-29.02	-25.77
M1	26.67	27.53	27.79	30.11	29.88	30.53	28.63	23.79	25.49
M2	5.39	6.95	7.10	5.37	6.25	4.60	2.86	5.60	6.17
M3	0.98	0.26	0.29	-0.72	-0.02	-0.28	1.27	1.68	1.47
M4	1.31	-0.37	-0.19	0.53	-0.33	0.59	3.93	3.85	3.79
Adult									
NAmp	-9.16	-12.66	-12.14	-4.52	-7.62	-3.66	-4.96	-10.52	-8.53
Slope	-20.43	-19.04	-18.58	-2.79	-7.20	-20.43	-22.06	-46.44	-33.20
M1	26.07	26.60	26.74	29.42	28.92	27.33	26.87	22.94	24.24
M2	6.00	6.99	7.81	5.44	5.92	3.50	4.48	3.36	5.15
M3	0.86	0.40	0.32	0.23	0.30	1.50	1.51	1.70	1.44
M4	1.24	0.34	0.13	-0.40	-0.43	3.94	3.41	4.25	2.72

A. Normalized amplitude

For the dependent measure of normalized amplitude (rms amplitude in decibels of the entire stop burst relative to the strongest component in the following vowel), a main effect of place [$F(2,64)=68.99$, $p<0.001$, $\eta^2=0.68$] was obtained. Pairwise comparisons indicated that all three places of articulation were significantly ($p<0.01$) different from each other in terms of normalized amplitude (-10.1 dB for /p/, -4.3 dB for /t/, and -6.1 dB for /k/). A main effect of vowel context was also found to be significant [$F(2,64)=47.86$, $p<0.001$, $\eta^2=0.60$], with differences in the normalized amplitude of the stop depending on the articulation of the following vowel. All three vowel contexts, /i a u/, produced significantly ($p<0.001$) different normalized amplitudes, with mean values of -5.0, -8.9, and -6.7 dB, respectively. Although a place-by-vowel interaction was also found to be significant [$F(4,128)=5.42$, $p<0.01$], the effect size was relatively small ($\eta^2=0.12$).

B. Spectral measures

1. Spectral slope

Significant differences in spectral slopes were found across place of stop articulation [$F(2,64)=116.23$, p

<0.001 , $\eta^2=0.78$]. Subsequent pairwise comparisons ($p<0.001$) demonstrated that the mean spectral slope of all three stops was significantly different from each other (-3.83 dB/kHz for /p/, 18.34 dB/kHz for /t/, and -16.58 dB/kHz for /k/). In addition, a significant effect of vowel context [$F(2,64)=53.56$, $p<0.001$, $\eta^2=0.63$] indicated that the spectral slope values of the stop articulations were different from each other depending on the articulation of the following vowel (the mean slope for /i, a, u/ contexts were 9.03, -5.60, and -5.50 dB/kHz, respectively). The effect of vowel context was mainly due to the significantly increased slope values of stops preceding an /i/ vowel ($p<0.001$). The ANOVA also yielded a significant place-by-vowel interaction [$F(4,128)=18.95$, $p<0.001$, $\eta^2=0.37$], characterized by a significantly elevated mean slope for /t/ and /k/ when immediately followed by an /i/ vowel ($p<0.001$).

Interestingly, a main effect was obtained for both the sex of the speaker [$F(1,32)=9.32$, $p<0.01$, $\eta^2=0.23$] and the age group [$F(3,32)=9.19$, $p<0.001$, $\eta^2=0.46$]. In addition, a significant sex-by-age group interaction was also noted [$F(3,32)=3.10$, $p<0.05$, $\eta^2=0.23$]. As can be seen in Fig. 1,

TABLE II. Acoustic measures from female speakers for three classes of voiceless American English stop consonants, grouped as a function of speaker age, place of articulation, and vowel context. Normalized amplitude in dB (NAmplitude), spectral slope in dB/kHz (slope), and the first four spectral moments (mean in ERB, M1; variance in kHz², M2; skewness, M3; and kurtosis, M4).

	Bilabial-/p/			Alveolar-/t/			Velar-/k/		
	/i/	/a/	/u/	/i/	/a/	/u/	/i/	/a/	/u/
3 yr. old									
NAmplitude	-6.89	-13.90	-14.36	-6.01	-7.33	-4.93	-6.42	-8.30	-6.77
Slope	6.04	0.93	2.32	40.40	20.67	24.85	10.68	-26.01	-30.65
M1	28.44	28.93	28.95	31.01	29.86	30.29	29.17	24.39	25.92
M2	7.02	6.31	6.87	4.53	5.53	4.94	3.89	5.17	4.64
M3	-0.22	-0.14	-0.12	-1.84	-0.72	-0.75	0.44	1.39	1.26
M4	0.03	0.01	-0.61	3.13	0.05	0.37	0.85	7.26	4.06
4 yr. old									
NAmplitude	-6.71	-10.67	-5.56	-1.97	-4.12	-3.92	-1.33	-8.80	-3.61
Slope	-4.18	-4.18	-0.46	34.59	21.13	24.38	26.70	-16.95	-19.27
M1	28.20	27.54	28.61	31.17	30.61	29.71	29.91	25.92	27.21
M2	6.06	7.13	6.45	4.99	5.56	5.39	4.96	6.56	4.98
M3	0.06	0.15	0.08	-1.80	-0.51	-0.56	-0.75	1.19	1.08
M4	-0.07	-0.32	-0.27	2.34	0.01	0.22	-0.33	2.80	3.02
5 yr. old									
NAmplitude	-6.96	-9.55	-8.99	-3.45	-5.52	-3.45	-5.20	-7.33	-6.64
Slope	6.97	14.16	17.41	49.92	28.79	32.74	13.70	-20.67	-21.36
M1	28.58	29.85	30.26	31.58	31.48	30.55	29.11	26.56	25.95
M2	6.35	7.86	6.78	3.25	4.04	5.03	3.19	9.16	6.91
M3	-0.24	-1.10	-1.04	-2.11	-0.97	-1.18	0.47	0.15	0.19
M4	-0.32	0.16	0.12	5.17	1.18	1.34	1.08	-0.68	1.00
Adult									
NAmplitude	-9.55	-14.93	-16.12	-4.51	-4.81	-2.24	-6.66	-10.67	-9.35
Slope	-15.56	-13.24	-15.09	31.58	5.57	-0.46	-9.75	-30.19	-35.06
M1	26.94	27.16	27.53	32.24	30.21	29.57	27.73	24.37	24.54
M2	6.49	7.57	6.16	3.56	4.54	3.91	4.10	5.15	4.66
M3	0.34	-0.06	0.40	-1.32	-0.14	0.53	1.27	1.54	1.83
M4	-0.15	-0.70	0.12	2.58	0.66	0.57	3.08	4.14	4.73

sex-specific differences ($p < 0.001$) in spectral slope began with the 5-year-old speakers and extended to the adults, with male speakers showing a significant decrease in spectral slope when compared to female speakers of similar age.

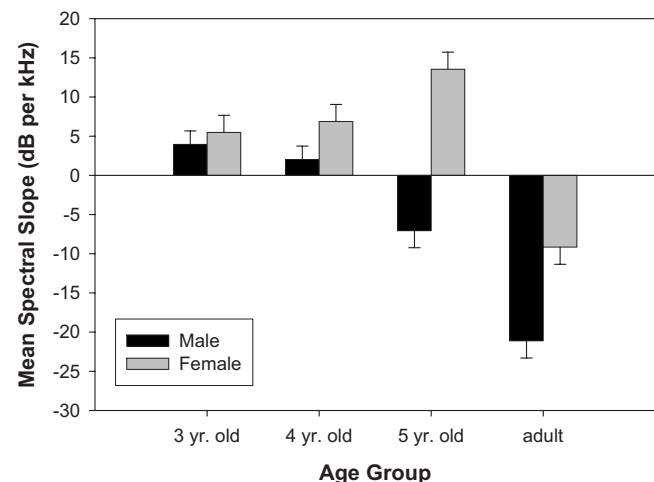
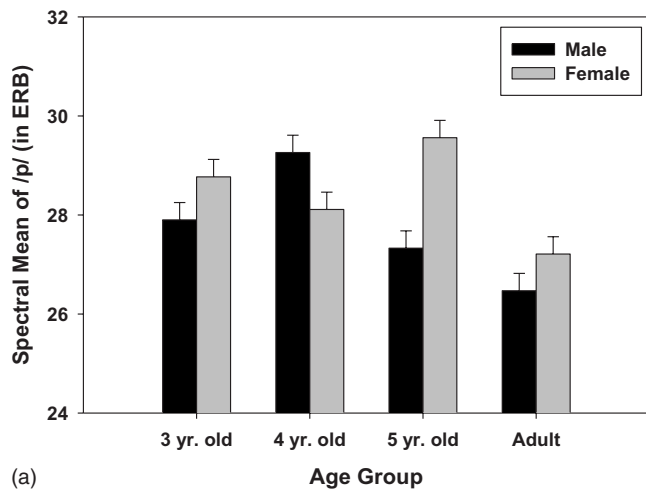


FIG. 1. Spectral slope as a function of the sex of the speaker and age group.

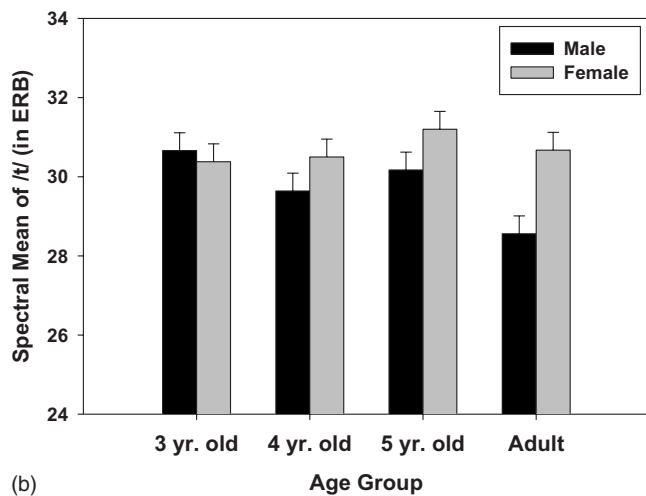
2. Spectral mean

As expected, the statistical analysis revealed a significant main effect for place of articulation [$F(2, 64) = 191.93$, $p < 0.001$], characterized by a strong effect size ($\eta^2 = 0.86$). *Post hoc* analyses indicated significant differences ($p < 0.001$) between all three places of articulation. Collapsed across speaker and vowel context the stop consonants, /p t k/, exhibited spectral means of 28.08, 30.22, and 26.45 ERB, respectively. There was also a significant effect of vowel context [$F(2, 64) = 36.74$, $p < 0.001$, $\eta^2 = 0.53$] and a significant place-by-vowel interaction effect [$F(4, 128) = 30.11$, $p < 0.001$, $\eta^2 = 0.48$]. Pairwise comparisons indicated that the main effect of vowel context was attributed primarily to the elevated spectral mean of the stop burst preceding an /i/ vowel ($p < 0.001$), whereas the differences between /a/ and /u/ were not found to be statistically significant. The /i/ vowel context effect was increased in velar stops (/k/) and relatively reduced in bilabial (/p/) and alveolar stops (/t/).

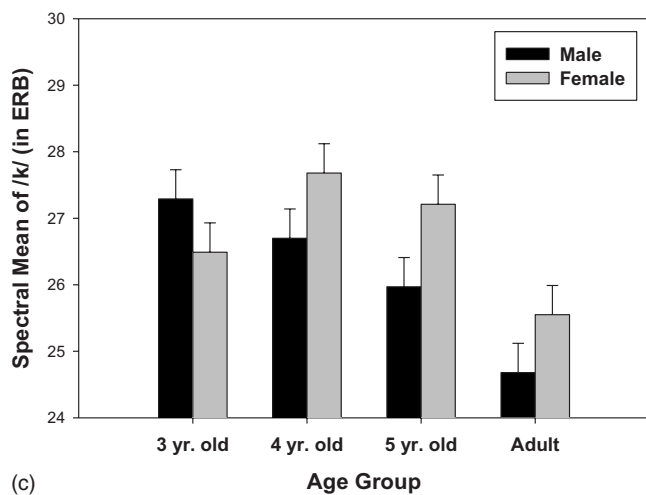
As expected, significant differences in spectral mean were found as a function of both the sex of the speaker [$F(1, 32) = 8.03$, $p = 0.008$, $\eta^2 = 0.20$] and the age group



(a)



(b)



(c)

FIG. 2. [(a)–(c)] Spectral mean as a function of place of stop articulation, speaker sex, and age group. Linear measures in hertz were converted to an ERB scale (Glasberg and Moore, 1990; Moore, 1997) prior to analysis.

[$F(3, 32)=7.65, p=0.001, \eta^2=0.42$], with female and child speakers exhibiting higher overall mean values than the male and adult speakers. Interestingly, the data also contained a significant speaker sex-by-age group [$F(3, 32)=4.31, p=0.012, \eta^2=0.29$] and place-by-speaker sex-by-age group interactions [$F(6, 64)=3.35, p=0.006, \eta^2=0.24$]. As shown in Fig. 2, significant sex-specific differences ($p<0.001$)

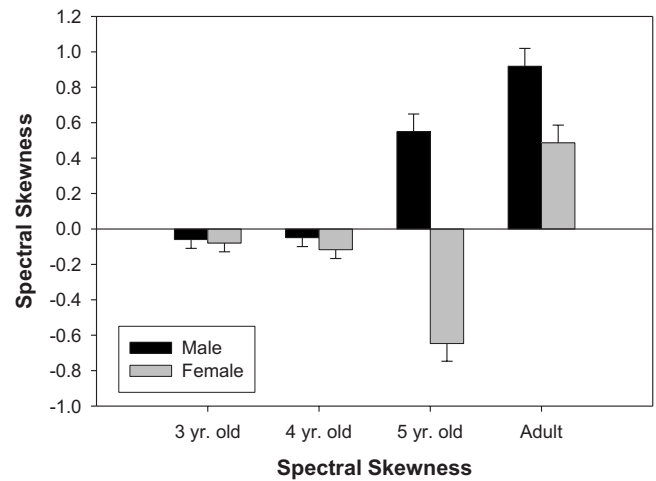


FIG. 3. Spectral skewness as a function of the sex of the speaker and age group.

were exhibited for /p/ and /t/ productions starting at age 4 and by age 5 for /k/. Within these two age groups, stop productions from male speakers were found to have a significant decrease in spectral mean when compared to female productions of the same age group, except for the 4-year old groups' productions of /p/, in which the spectral mean was significantly elevated.

3. Spectral variance

A significant main effect of place of stop articulation [$F(2, 64)=31.86, p<0.001, \eta^2=0.50$] and vowel context [$F(2, 64)=18.17, p<0.001, \eta^2=0.36$] was also found for the measure of spectral variance. The spectral variance of the bilabial stops was significantly higher ($p<0.001$) than the other two places of articulation (the variances for /p t k/ were 6.7, 4.9, and 5.1 kHz², respectively). In addition, the variance of the stop burst was significantly lower ($p<0.001$) when followed by an /i/ vowel. However, this pattern was reversed for alveolar stops (/t/), which explains a small (an effect size of $\eta^2=0.12$) yet significant place-by-vowel interaction effect [$F(4, 128)=4.51, p<0.01$].

4. Spectral skewness

A main effect of place [$F(2, 64)=87.47, p<0.001$], with a large effect size ($\eta^2=0.73$) was obtained for spectral skewness. *Post hoc* analyses ($p<0.001$) indicated that the mean spectral skewness of all three stops was significantly different from each other (−0.017 for /p/, −0.640 for /t/, and 0.997 for /k/). In addition, an effect of vowel context [$F(2, 64)=15.42, p<0.001, \eta^2=0.32$], as well as a significant place-by-vowel interaction [$F(4, 128)=14.97, p<0.001, \eta^2=0.32$] were obtained from the ANOVA. The effect of vowel context and associated interaction were largely due to the fact that the spectral skewness of the alveolar stop burst decreased significantly ($p<0.001$) when preceding an /i/ vowel.

The ANOVA also revealed main effects for both the sex [$F(1, 32)=8.96, p<0.01, \eta^2=0.22$] and age [$F(3, 32)=7.23, p<0.001, \eta^2=0.40$] of the speaker, as well as a significant sex-by-age group interaction [$F(3, 32)=3.59, p<0.03, \eta^2=0.25$], as shown in Fig. 3. Similar to the pattern of results

obtained for spectral mean, when looking at skewness there were distinct sex-specific differences which began with the 5-year-old age group and were extended to the adults. *Post hoc* tests demonstrated that within these two age groups the sex-specific differences were significant ($p < 0.001$).

5. Spectral kurtosis

Analysis revealed only one significant effect for the measure of kurtosis, that being place of articulation [$F(2, 64) = 20.24, p < 0.001, \eta^2 = 0.39$], with subsequent comparisons indicating that all three places were significantly different ($p < 0.01$). The spectral kurtosis was found to increase as the stop articulation moved posterior in the oral cavity (0.06 for /p/, 1.13 for /t/, and 2.72 for /k/).

IV. DISCUSSION

Findings indicated that the dependent measures of normalized amplitude, spectral slope, and all four spectral moments varied significantly as a function of place of articulation. Subsequent pairwise comparisons revealed that, with the exception of spectral variance, the acoustic and spectral measures varied significantly across all three places of stop articulation. These results are similar to previous findings (i.e., Forrest *et al.*, 1988; Nittrouer, 1995), which indicate that the spectral moments measures of mean and skewness differentiate alveolar and velar stops. In addition, subsequent analysis demonstrated that all three places of stop articulation were significantly different from each other in terms of normalized amplitude and spectral slope; measures not reported by Forrest *et al.* (1988) and Nittrouer (1995). Unlike Forrest *et al.* (1988), this study found significant differences between stops in terms of the second spectral moment (variance), a spectral measure that has not been included in a number of spectral moments studies of child productions (e.g., Forrest *et al.*, 1990; Nittrouer, 1995). The results of this study found that spectral variance was useful in distinguishing /p/ from /t/ and /k/ across all age groups.

Not surprisingly, the results of this study indicate that the acoustic structure of the stop burst also varied as a function of the following vowel. Significant vowel context effects were found for the measures of spectral slope and the first three spectral moments. Nittrouer (1995) postulated that similar vowel context effects for spectral mean were primarily the result of changes in the acoustic parameters of stop productions when followed by a high front vowel. In theory, this anticipatory action would produce a shortening of the anterior resonating cavity and thereby result in an increase in the spectral mean. Due to the relative position of the tongue during production, this theory is a possible explanation for differences in velar productions, but unlikely for alveolar stops.

Findings from locus equation studies (Gibson and Ohde, 2007; Sussman *et al.*, 1992) may provide some evidence regarding this interpretation. These studies examined locus equations as a measure of coarticulation across voiced stop productions in young children. Results indicated that voiced alveolar stops exhibited less coarticulation than voiced bilabial or velar stops. However, comparisons between slope val-

ues from locus equations and spectral moments studies should be interpreted with caution considering the differences in analysis procedures.

Relevant to the reported differences in normalized amplitude in this study, previous research has indicated that the amplitude of the stop burst does have an effect on the perception of the place of articulation for both voiceless and voiced stop consonants (e.g., Blumstein and Stevens, 1980; Hedrick and Jesteadt, 1996; Ohde and Stevens, 1983; Repp, 1984). It is unclear if the spectral amplitude and shape of the stop burst provide an invariant and independent cue to the perception of place of articulation (Blumstein and Stevens, 1979) or if the cue of relative burst amplitude is context dependent, determined in part by the acoustic components of the following vowel (Dorman *et al.*, 1977). In support of the latter approach, the findings of the current study found that the normalized amplitude of the stop burst was significantly affected by the following vowel context. This result may provide evidence that children at an earlier stage of development, as young as 3 years of age, exhibit phoneme specific patterns of anticipatory coarticulation (Gibson and Ohde, 2007; Sussman *et al.*, 1996). Within the scope of this study, these results also indicate that male and female children appear to produce normalized amplitude in a similar manner.

As expected, female and child speakers exhibited higher overall spectral mean values. However, significant sex and age group differences were also found for the spectral measures of slope and skewness. Of particular interest, significant speaker sex-by-age interaction effects were found for spectral slope, mean, and skewness. In general, sex-specific differences for these measures began with the 5-year-old speakers and extended to the adults. For some spectral measures, the age at which these differences emerged was also dependent on the specific type of stop articulation. For example, sex-specific differences in spectral mean for bilabial stop productions were noted in the 5-year-old speakers, whereas such differences for alveolar and velar stops were found to emerge at 4 years of age. It is unclear why sex-specific differences in spectral mean were delayed in bilabial stops relative to alveolar and velar stops or why the 4-year old bilabial productions were elevated. Normative research examining the developmental age of stop consonant acquisition has indicated that bilabial stop consonants are typically acquired at the same time or several months earlier than alveolar and velar stops (Arlt and Goodban, 1976; Smit *et al.*, 1990); thus it is unlikely that developmental age is a factor in the differences noted in this study. Sex-differences across stop type may be due to the manner of the articulations required, namely, a bilabial closure as compared to a tongue-to-palate contact patterns.

Acoustic differences in the speech of male and female adult speakers can be largely explained by sex-related variation in characteristics such as fundamental frequency and formant ratios. For example, typical male speakers generally exhibit lower fundamental and formant frequencies than female speakers when producing vowel segments (Hillenbrand *et al.*, 1995; Mattingly, 1966; Peterson and Barney, 1952). In adults, these acoustic differences are in part the result of

sexual dimorphism of anatomical factors such as vocal tract length and shape, as well as vocal fold size (Fitch and Giedd, 1999; Titze, 1989).

However, findings from studies with children indicate that sex-related acoustic differences may be more complex than can be reasonably explained by nonuniform variation in vocal tract morphology. Sexual dimorphism of the vocal tract in children is generally considered to begin at peri- and post-pubertal stages of physical maturation. Although anatomical studies of vocal tract morphology have historically involved a relatively small number of adult subjects (Baer *et al.*, 1991; Dang *et al.*, 1994; Moore, 1992; Story *et al.*, 1996; Sulter *et al.*, 1992), the increased availability and decreased health risks associated with MRI have provided the opportunity for more accurate morphometric research on the vocal tract anatomy of larger numbers of children. Recent large-scale MRI studies (Fitch and Giedd, 1999; Vorperian *et al.*, 2005, 2009) have provided a greater understanding of the anatomical development of the vocal tract in children. Findings from these studies indicate that the oral and pharyngeal structures of the vocal tract have ongoing, and at times, accelerated periods of growth through childhood. However, results of both studies indicated no appreciable sexual dimorphism in the vocal tract structures of younger prepubescent children.

In view of these anatomical data, it is reasonable to postulate that sex-related differences in acoustic properties of speech in children are in part the result of factors other than anatomical variation in the vocal tract. Data from this study may support the conclusion that articulatory development for some aspects of stop production follow different patterns in girls as opposed to boys, patterns which may be based on male-female archetypes present in adult production patterns. Evidence from perceptual experiments designed to evaluate listeners ability to identify the sex of a speaker from only auditory information have also provided some support for the notion that the acoustic signatures of children's speech may be the result of learned characteristics. In a perceptual study, Sachs *et al.* (1973) found that adult listeners were able to accurately identify (81%) a young speaker's sex from short passages of speech. Interestingly, an acoustic analysis of the recordings indicated that the male children participating in the study exhibited a higher average F0 than the young female speakers, yet lower formant frequencies. From these results, the authors concluded that the listeners' identifications may have been based in part on sex-related acoustic differences (e.g., formant frequency patterns, voice quality, intonation patterns) which arise from learned articulatory patterns: patterns which adhere to culturally determined articulatory patterns viewed as appropriate for each sex.

The findings of this study indicate that in terms of speech production the spectral slope, mean, and skewness differed as a function of the speaker's sex. However, the perceptual relevance and physiologic mechanisms for these acoustic differences remains unclear. Research has indicated that listeners' are able to identify gender from the speech of children as young as 4 years of age (Perry *et al.*, 2001; Sachs *et al.*, 1973). Although the perceptual distinction of gender may primarily be determined by formant characteristics of

vowel segments (Bennett, 1981; Whiteside, 2001), it is unclear if additional acoustic cues, such as those included in this study, might also contribute to such distinctions.

Results from the current study indicate that sex-specific differences in the spectral speech patterns of young children may be associated with learned or behavioral factors that affect articulatory development, beginning at approximately 5 years of age. This conclusion is similar to previous research examining children's obstruent productions (Nissen and Fox, 2005; Whiteside and Marshall, 2001), vowel formant frequencies (Bennett, 1981; Whiteside, 2001), and the perceptual sex identification of children's voices (Perry *et al.*, 2001; Sachs *et al.*, 1973), which have also indicated that sex-specific acoustic and spectral differences in children's speech are not fully explained by anatomic differences alone, but likely the result of cultural or sociophonetic factors.

V. CONCLUSIONS

The findings of this study were based on acoustic measures collected from children in discrete age groups; however, it would be of interest, in future studies, to examine children's speech development in a longitudinal manner. Although more difficult to conduct, this type of study would provide a more comprehensive understanding of possible sex-specific differences in children's speech, while having greater control of inter-speaker variation. Future studies are needed to understand if such measures are perceptually salient, in isolation or in conjunction with other acoustic cues. It would also be of interest to conduct additional research examining the physiologic basis of these spectral characteristics acoustic dimension and thereby correlate spectral moment measures with articulatory movement or vocal fold shape. Furthermore, the acoustic measures utilized in the present study were limited to static acoustic and spectral cues. Future studies might examine how children develop dynamic speech cues, such as the change in spectral slope or mean over the duration of the stop burst. Despite these limitations, it is hoped the findings of this study will contribute to a greater understanding of speech development in prepubescent children and provide additional insight into possible sex-specific patterns of stop articulation, as well as the developmental stage such differences might typically emerge.

ACKNOWLEDGMENTS

This research was supported by a Graduate Student Alumni Research Association grant from The Ohio State University and a McKay School of Education Grant from Brigham Young University.

Adobe Systems Incorporated. (2003). ADOBE AUDITION (Version 1.3) (Computer software), San Jose, CA.

Arlt, P. B., and Goodban, M. T. (1976). "A comparative study of articulation acquisition as based on a study of 240 normals, aged three to six," *Language, Speech, and Hearing Services in Schools* 7, 173-180.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (1991). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *J. Acoust. Soc. Am.* 90, 799-828.

Behrens, S., and Blumstein, S. E. (1988a). "Acoustic characteristics of Eng-

- lish 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.
- Bennett, S. (1981). "Vowel formant frequency characteristics of preadolescent males and females," *J. Acoust. Soc. Am.* **69**, 231–238.
- Blumstein, S. E., and Stevens, K. N. (1980). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," *J. Acoust. Soc. Am.* **67**, 648–662.
- Busby, P., and Plant, G. (1995). "Formant frequency values of vowels produced by preadolescent boys and girls," *J. Acoust. Soc. Am.* **97**, 2603–2606.
- Byrd, D. (1992). "Preliminary results on speaker-dependent variation in the TIMIT database," *J. Acoust. Soc. Am.* **92**, 593–596.
- Byrd, D. (1994). "Relations of sex and dialect to reduction," *Speech Commun.* **15**, 39–54.
- Dang, J., Honda, K., and Suzuki, H. (1994). "Morphological and acoustic analysis of the nasal and the paranasal cavities," *J. Acoust. Soc. Am.* **96**, 2088–2100.
- Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (1977). "Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues," *Percept. Psychophys.* **22**, 109–122.
- Fant, G. (1966). "A note on vocal tract size factors and nonuniform F-pattern scaling," *Speech Sounds and Features* (MIT, Cambridge, MA), pp. 84–93.
- Fant, G., Kruckenberg, A., and Nord, L. (1991). "Prosodic and segmental speaker variation," *Speech Commun.* **10**, 521–531.
- 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.
- Fitzsimmons, M., Sheahan, N., and Staunton, H. (2001). "Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies," *Brain Lang* **78**, 94–108.
- 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. Phonetics* **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. Phonetics* **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. (2005). "Sex-related acoustic changes in voiceless English fricatives," *J. Speech Lang. Hear. Res.* **48**, 753–765.
- Gibson, T., and Ohde, R. N. (2007). "F2 locus equations: Phonetic descriptors of coarticulation in 17–22 month old children," *J. Speech Lang. Hear. Res.* **50**, 97–108.
- 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).
- Graddol, D. (1986). "Discourse specific pitch behavior," in *Intonation in Discourse*, edited by C. Johns-Lewis (Croom Helm, London), pp. 221–237.
- Hedrick, M. S., and Jesteadt, W. (1996). "Effect of relative amplitude, presentation level, and vowel duration on perception of voiceless stop consonants by normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **100**, 3398–3407.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.
- Lee, S., Potamianos, A., and Narayanan, S. (1999). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *J. Acoust. Soc. Am.* **105**, 1455–1468.
- Mattingly, I. (1966). "Speaker variation and vocal-tract size," *J. Acoust. Soc. Am.* **39**, S1219A.
- Mendoza, E., Valencia, N., Munoz, J., and Trujillo, H. (1996). "Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS)," *J. Voice* **10**, 59–66.
- Miccio, A. W. (1996). "A spectral moments analysis of the acquisition of word-initial voiceless fricatives in children with normal and disordered phonologies," Ph.D. thesis, Indiana University, Bloomington, IN.
- Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing* (Academic, New York).
- Moore, C. A. (1992). "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images," *J. Speech Hear. Res.* **35**, 1009–1023.
- Nissen, S. L., and Fox, R. A. (2005). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," *J. Acoust. Soc. Am.* **118**, 2570–2578.
- Nittrouer, S. (1992). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," *J. Phonetics* **20**, 351–382.
- 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 Lang. Hear. Res.* **32**, 120–132.
- Ohde, R. N., and Stevens, K. N. (1983). "Effect of burst amplitude on the perception of stop consonant place of articulation," *J. Acoust. Soc. Am.* **74**, 706–714.
- Perry, T. L., Ohde, R. N., and Ashmead, D. H. (2001). "The acoustic bases for gender identification from children's voices," *J. Acoust. Soc. Am.* **109**, 2988–2998.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 585–594.
- Repp, B. H. (1984). "Closure duration and release burst amplitude cues to stop consonant manner and place of articulation," *Lang Speech* **27**, 245–254.
- Ryalls, J. H., Zipprer, A., and Baldauff, P. (1997). "A preliminary investigation of the effects of gender and race on voice onset time," *J. Speech Lang. Hear. Res.* **40**, 642–645.
- Sachs, J., Liberman, P., and Erickson, D. (1973). "Anatomical and cultural determinants of male and female speech," in *Language Attitudes*, edited by R. W. Shuy and R. W. Fasold (Georgetown University Press, Washington DC), pp. 74–84.
- Smit, A. B., Hand, L., Feilinger, J. J., Bernthal, J. E., and Bird, A. (1990). "The Iowa articulation norms project and its Nebraska replication," *J. Speech Hear. Disord.* **55**, 779–798.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.* **100**, 537–554.
- Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyart, E. L. (1992). "On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI," *Magn. Reson. Imaging* **10**, 365–373.
- Sussman, H., Hoemeke, K., and McCaffrey, H. (1992). "Locus equation as an index of coarticulation for place of articulation distinctions in children," *J. Speech Hear. Res.* **35**, 769–781.
- Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C., and Smith, J. (1996). "Consonant-vowel interdependencies in babbling and early words: preliminary examination of a locus equation approach," *J. Speech Hear. Res.* **39**, 424–433.
- Swartz, B. L. (1992). "Gender differences in voice onset time," *Percept. Mot. Skills* **75**, 983–992.
- Sweeting, P. M., Baken, R. J. (1982). "Voice onset time in a normal-ages population," *J. Speech Hear. Res.* **25**, 129–134.
- Titze, I. R. (1989). "Physiologic and acoustic differences between male and female voices," *J. Acoust. Soc. Am.* **85**, 1699–1707.
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (2005). "Development of vocal tract length during childhood: A magnetic resonance imaging study," *J. Acoust. Soc. Am.* **117**, 338–350.
- Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., and Gentry, L. R. (2009). "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," *J. Acoust. Soc. Am.* **125**, 1666–1678.

- Whiteside, S. P. (1996). "Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences," *J. Int. Phonetic Assoc.* **26**, 23–40.
- Whiteside, S. P. (2001). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," *J. Acoust. Soc. Am.* **110**, 464–478.
- Whiteside, S. P., and Hodgson, C. (2000). "Speech patterns of children and adults elicited via a picture-naming task: An acoustic study," *Speech Commun.* **32**, 267–285.
- Whiteside, S. P., and Irving, C. J. (1997). "Speakers' sex differences in voice onset time: Some preliminary findings," *Percept. Mot. Skills* **85**, 459–463.
- Whiteside, S. P., and Marshall, J. (2001). "Developmental trends in voice onset time: Some evidence of sex difference," *Phonetica* **58**, 196–210.