Acoustic characteristics of English fricatives. — Source link

Allard Jongman, Ratree Wayland, Serena Wong

Institutions: University of Kansas

Published on: 01 Sep 2000 - Journal of the Acoustical Society of America (Acoustical Society of America)

Topics: Place of articulation and Articulation (phonetics)

Acoustic characteristics of English fricatives

Allard Jongman[a)]
Linguistics Department, Blake Hall, The University of Kansas, Lawrence, Kansas 66045

Ratree Wayland
Program in Linguistics, Yon Hall, University of Florida at Gainesville, P. O. Box 115454, Gainesville, Florida 32611-5454

Serena Wong
Columbia University College of Physicians & Surgeons, 630 West 168th Street, New York, New York 10032

(Received 3 March 2000; accepted for publication 11 June 2000)

This study constitutes a large-scale comparative analysis of acoustic cues for classification of place of articulation in fricatives. To date, no single metric has been found to classify fricative place of articulation with a high degree of accuracy. This study presents spectral, amplitudinal, and temporal measurements that involve both static properties (spectral peak location, spectral moments, noise duration, normalized amplitude, and $F2$ onset frequency) and dynamic properties (relative amplitude and locus equations). While all cues (except locus equations) consistently serve to distinguish sibilant from nonsibilant fricatives, the present results indicate that spectral peak location, spectral moments, and both normalized and relative amplitude serve to distinguish all four places of fricative articulation. These findings suggest that these static and dynamic acoustic properties can provide robust and unique information about all four places of articulation, despite variation in speaker, vowel context, and voicing. [S0001-4966(00)02909-X]

PACS numbers: 43.70.Fq [AL]

I. INTRODUCTION

One of the primary goals of speech research is to characterize the defining properties of speech sounds that occur in natural language, and to determine how the listener extracts these properties in the process of speech perception. Phonetic research of the past 50 years has demonstrated that the identification of acoustic cues which uniquely characterize particular (classes of) speech sounds is a serious challenge. A major obstacle in this endeavor is the variability typically found in the speech signal, often resulting in a defective one-to-one correspondence between acoustic cue and phonetic percept (Liberman *et al.*, 1967). This lack of invariance arises from a variety of sources, including speaker size, phonetic context, and speaking rate (see Pisoni and Luce, 1986, for an overview). The basic problem, then, is how perceptual constancy or invariance is achieved in the presence of such varying information.

Much research has been devoted to the question of whether distinct spectral patterns that correspond to phonetic dimensions, such as place and manner of articulation, can be derived from the acoustic waveform. Early studies failed to find any consistent mapping between acoustic properties and phonetic features (e.g., Cooper *et al.*, 1952; Schatz, 1954; Delattre *et al.*, 1955). Some recent research, however, suggests that stable, consistent acoustic properties may indeed be found in the speech signal, with appropriate analyses (e.g., Stevens and Blumstein, 1981; Kewley-Port, 1983; Lahiri *et al.*, 1984; Forrest *et al.*, 1988; Sussman *et al.*, 1991). Such research has predominantly focused on the search for properties distinguishing place of articulation in (English) stop consonants. In contrast, fricatives have been studied in much less detail. Moreover, it is uncertain whether the classification metrics proposed for stop consonants can be successfully applied to fricatives. The current study contributes to the body of research on the mapping between acoustic properties and phonetic categories by providing a detailed look at this mapping for English fricatives.

Fricatives are produced with a very narrow constriction in the oral cavity. A rapid flow of air through the constriction (the position of which depends on the particular fricative) creates turbulence in the flow, and the random velocity fluctuations in the flow act as a source of sound (e.g., Stevens, 1971, 1998; Shadle, 1990). English fricatives are usually grouped into four classes according to their place of articulation: labiodental /f,v/, (inter)dental /θ,ð/, alveolar /s,z/, and palato-alveolar /ʃ,ʒ/. Most studies of fricatives exclude /h/, since it is considered the voiceless counterpart of the abutting vowel (e.g., Pike, 1943; Ladefoged, 1982), and for that reason /h/ will not be considered in the present study either.

Previous studies of fricatives have concentrated on four attributes: spectral properties of the frication noise, amplitude of the noise, duration of the noise, and spectral properties of the transition from the fricative into the following vowel. In general, these studies have documented acoustic differences between the sibilant (/s,z,ʃ,ʒ/) and nonsibilant (/f,v,θ,ð/) fricatives, which involve spectrum, amplitude, and duration of the frication noise. Additionally, /s,z/ may be distinguished from /ʃ,ʒ/ on the basis of noise spectrum, while there are some data suggesting that /f,v/ may be distinguished from /θ,ð/ on the basis of transition information.

[a)]Electronic mail: jongman@ukans.edu

However, no cue has been identified so far that can uniquely distinguish all four places of articulation.

The present study is a comprehensive comparative analysis of acoustic cues to place of articulation in English fricatives. Spectral parameters include spectral peak location, spectral moments, locus equations, and $F2$ onset. Amplitudinal parameters include overall noise amplitude as well as relative amplitude. Temporal measurements consist of fricative noise durations. The data reported here thus concern both static and dynamic properties. Static properties pertain to acoustic information that is measured at one location of the speech signal, while dynamic properties pertain to changes in acoustic information during the fricative and/or adjacent segments. Spectral peak location, spectral moments, $F2$ onset frequency, noise amplitude, and noise duration are considered static properties. Dynamic properties include locus equations and relative amplitude. Inclusion of both static and dynamic parameters may result in a more comprehensive characterization of fricative acoustics. In particular, the goal of this study is to identify stable acoustic cues to place of articulation, to evaluate the nature of these cues: are they primarily in terms of spectrum, amplitude, or duration, and, finally, to determine their location: are these cues uniformly distributed throughout the fricative, or are some regions more informative than others?

A. Spectral properties

1. Frication noise: Spectral peak location and spectral moments

The overall spectral shape of each fricative is determined by the size and shape of the oral cavity in front of the constriction. The longer this anterior cavity, the more defined the resulting spectrum (e.g., Stevens, 1998). As a result, the alveolar and palato-alveolar fricatives are characterized by well-defined, distinct spectral shapes while labiodental and (inter)dental fricatives display a relatively flat spectrum (e.g., Strevens, 1960; Jassem, 1965; Behrens and Blumstein, 1988a). In particular, /ʃ,ʒ/ typically exhibit a midfrequency spectral peak at around 2.5–3 kHz which often corresponds to $F3$ of the following vowel. Alveolar /s,z/ are produced with a shorter anterior cavity relative to /ʃ,ʒ/ and therefore display a primary spectral peak at higher frequencies, around 4 to 5 kHz. In addition, since for these fricatives the airstream hits the teeth, the high-frequency turbulence is very intense. Both /f,v/ and /θ,ð/ are characterized by a relatively flat spectrum with no clearly dominating peak in any particular frequency region.

Previous studies reveal that the local spectral properties of frication noise serve to distinguish the sibilant fricatives /s,z,ʃ,ʒ/ as a group from the nonsibilants /f,v,θ,ð/. Within the sibilants, /s,z/ can also be distinguished from /ʃ,ʒ/ on the basis of the spectral properties of the noise (e.g., Hughes and Halle, 1956; Strevens, 1960; Heinz and Stevens, 1961; Shadle, 1990; Behrens and Blumstein, 1988a; Evers, Reetz, and Lahiri, 1998). However, the location of the spectral peaks in the frication noise is to some extent speaker dependent (Hughes and Halle, 1956) and vowel dependent (Soli, 1981). Recently, Tabain (1998) obtained high classification rates for sibilants and moderate rates for nonsibilants. Averaged spectra were calculated based on a series of fast Fourier transforms (FFTs) across each fricative. These spectra were then subjected to a classification algorithm based on a Bayesian distance measure. Classification across five male and five female speakers averaged 97% for the sibilants but only 70% for the nonsibilants. Unfortunately, for the nonsibilants, there were no consistent differences in the spectra which correlated with classification accuracy. In other words, it was not clear which acoustic properties contributed to correct classification.

Spectral moments analysis involves a statistical procedure for classifying obstruents, capturing both local (mean frequency) and global (spectral tilt and peakedness) aspects of speech sounds. These analyses may be based on one or multiple regions of the speech signal. In Forrest *et al.* (1988), a series of FFTs was calculated every 10 ms from the onset of the word-initial obstruent. Each FFT was treated as a random probability distribution from which the first four moments (mean, variance, skewness, and kurtosis) were computed. Mean and variance reflect the average energy concentration and range, respectively. Skewness is an indicator of a distribution's asymmetry. A skewness of zero indicates a symmetrical distribution around the mean. Skewness is positive when the right tail of the distribution extends further than the left tail. Likewise, skewness is negative when the left tail of the distribution extends further than the right tail (e.g., Newell and Hancock, 1984). In phonetic terms, skewness refers to spectral tilt, the overall slant of the energy distribution. Positive skewness suggests a negative tilt with a concentration of energy in the lower frequencies. Negative skewness is associated with a positive tilt and a predominance of energy in the higher frequencies. Finally, kurtosis is an indicator of the peakedness of the distribution. Positive kurtosis values indicate a relatively high peakedness (the higher the value, the more peaked the distribution), while negative values indicate a relatively flat distribution. Positive kurtosis thus suggests a clearly defined spectrum with well-resolved peaks, while negative kurtosis indicates a flat spectrum without clearly defined peaks. The spectral moments metric thus incorporates both local (spectral peak) and more global (spectral shape) information.

Forrest *et al.* (1988) derived spectral moments for a small corpus of syllable-initial fricatives (''see, she, fought, thought, fat'') produced by five females and five males. These moments were then entered into a discriminant analysis for classification in terms of place of articulation. Classification based on the first 20 ms of the fricative was good for sibilants (85% for /s/, 95% for /ʃ/); however, classification of nonsibilants was poor (58% for /θ/, 75% for /f/). Classification rates for individual moments were not reported.

Although promising as a technique to quantify spectral properties of obstruents, surprisingly little research has attempted to replicate or extend the Forrest *et al.* (1988) findings. In a preliminary report, Shadle and Mair (1996) analyzed all eight English fricatives produced by only one female and one male speaker. Moments were computed at the beginning, middle, and end of each fricative. Moments did capture some important fricative characteristics: the second moment (variance) was large for the nonsibilant fricatives, and /ʃ/ was uniquely characterized by a low first moment (mean). Nevertheless, the authors concluded that spectral moments did not reliably differentiate fricative place of articulation.

The most comprehensive study to date is that by Tomiak (1990), who reported all moments for the four voiceless fricatives and /h/ as produced by six American speakers. Although Tomiak (1990) did not subject her measurements to analyses of variance, she reported the following observations: /θ/ displayed a greater standard deviation, skewness, and kurtosis than /f/; /s/ was distinct from /ʃ/, having a higher mean, lower standard deviation, and greater kurtosis. Discriminant analysis yielded poor classification rates for the nonsibilant fricatives (67% for /f/, 44% for /θ/) and high rates for the sibilants (96% for both /s/ and /ʃ/).

Most studies using spectral moments have concentrated on the spectral mean and report that /ʃ/ has a lower mean than /s/ (e.g., Nittrouer *et al.*, 1989; Tjaden and Turner, 1997 for normal controls). Nittrouer (1995) and McFarland *et al.* (1996) have reported that spectral moments 1, 3, and 4 (mean, skewness, and kurtosis, respectively) distinguish /s/ from /ʃ/ across male and female adult speakers and different vowel contexts. Specifically, /ʃ/ was characterized by a lower spectral mean, positive skewness, and smaller kurtosis, indicating a slightly flatter spectrum.

In sum, while some spectral moments distinguished /s/ from /ʃ/, spectral moments have not been shown to reliably differentiate the nonsibilants.

2. Transition information: Locus equations and F2 onset

Locus equations are based on the second formant frequency ($F2$) at vowel onset and at vowel midpoint (e.g., Sussman *et al.*, 1991; Sussman, 1994) and constitute a dynamic representation of speech sounds since they express a relation between $F2$ at different points in the speech signal. Results indicate that the apparent $F2$ starting frequency of a vowel preceded by an obstruent provides information about the articulatory configuration used to generate the consonant. Although locus equations have recently been successful in the classification of place of articulation in voiced stop consonants, researchers have only just begun to apply this method to fricatives (e.g., Wilde, 1993; Fowler, 1994; Sussman, 1994; Sussman and Shore, 1996; Yeou, 1997). At present, there are very few data on fricative locus equations, and the results are contradictory: Fowler (1994) and Yeou (1997) obtained good classification of fricatives, with each place of articulation characterized by a distinct slope and y intercept. Yeou (1997) investigated locus equations for Arabic stops and fricatives. Slope and y-intercept values uniquely distinguished those fricatives that are common to Arabic and English (/f,ð,s,ʃ/). However, overlap occurred between postalveolar /ʃ/ and pharyngeal /h/ and between labial /f/ and uvular /χ/ in terms of both slope and y intercept.

Unfortunately, there is little correspondence between the values observed across these two studies for each place of articulation. The only qualitative agreement is that the labiodental place has the highest slope and lowest y-intercept value (Fowler: /v/ 0.73 and 337 Hz, respectively; Yeou: /f/ 0.92 and 61 Hz, respectively). In two smaller-scale studies, Wilde (1993) and Sussman (1994) did not obtain unique classification. In his analysis of the voiced fricatives /v,ð,z,ʒ/ of four speakers, Sussman (1994) found that only labiodental /v/ was significantly different in terms of slope (0.74) from the other three places of articulation. Similarly, although Wilde did not provide any statistics, only /f,v/ seem different from the other three places of articulation.

Wilde (1993) provides preliminary data suggesting that the onset of $F2$ alone at the fricative–vowel boundary or its range varies systematically as a function of place of articulation. Based on data from two speakers, Wilde (1993) observed that, for a given vowel context, $F2$ onset is progressively higher as the place of constriction moves back in the oral cavity. Studies investigating effects of formant transition information on perception of the /s/–/ʃ/ distinction also typically employ synthetic stimuli in which $F2$ onset frequency for /ʃ/ is substantially higher (approximately 100–300 Hz) than for /s/ (e.g., Mann and Repp, 1980; Whalen, 1981; Nittrouer, 1992). In addition, Wilde (1993) presented data that indicate that the range of $F2$ onset is progressively smaller as place of constriction moves further back, as had been previously reported for stop consonants by Kewley-Port (1982). These findings are also consistent with Recasens' (1985) observation that consonants with a greater degree of tongue-body raising (and thus typically a more posterior place of articulation) are more resistant to coarticulation.

B. Amplitude

1. Overall noise amplitude

Most research concerned with frication amplitude has investigated the overall amplitude of fricatives. These studies (e.g., Strevens, 1960; Behrens and Blumstein, 1988a) have focused on voiceless fricatives and converge on similar findings: sibilant /s,ʃ/ have a substantially greater (10–15 dB) amplitude than nonsibilant /f,θ/. Within each group, however, the two fricatives are not different from each other.

2. Relative amplitude

It has been suggested that overall amplitude may not be the relevant parameter; instead, a change in amplitude of the frication relative to the vowel in a specific frequency region may vary with place of articulation (Stevens, 1985). However, to date, no systematic acoustic study has been conducted to determine the magnitude of differences in relative amplitude as a function of place of articulation. Instead, research on relative amplitude has focused on its role in perception (e.g., Stevens, 1985; Hedrick and Ohde, 1993; Hedrick, 1997; Hedrick and Carney, 1997). For example, in order to create appropriate synthetic stimuli, Hedrick and Ohde (1993) measured relative amplitude for /s,S/ in the context of /ɑ/ produced by a female speaker. Relative amplitude, defined as the difference between fricative and vowel amplitude in the $F3$ region for sibilants, was -17 dB for /s/ and $+16$ dB for /ʃ/, suggesting that relative amplitude may distinguish sibilant fricatives in terms of place. Indeed, relative amplitude was shown to be a cue to perception of the place contrast between /s/ and /ʃ/ (Stevens, 1985; Hedrick and Ohde, 1993). In addition, in an /s–θ/ labeling task, relative amplitude values of -20 to 0 dB were shown to yield /θ/ responses, while values of 10 to 20 dB elicited /s/ responses (Hedrick and Ohde, 1993). Unfortunately, no relative amplitude measures were provided for /θ/, nor has anyone investigated relative amplitude in the labiodental fricatives.

C. Noise duration

Noise duration serves to distinguish sibilant from nonsibilant fricatives, with /s,ʃ/ being longer than /f,θ/ (e.g., Behrens and Blumstein, 1988a). However, Behrens and Blumstein (1988a) found no difference in duration between /s/ and /ʃ/ and only a trend for /θ/ to be shorter than /f/. Noise duration does provide a robust cue to the voicing distinction in syllable-initial position, with voiceless fricatives having longer noise durations than voiced fricatives. This observation holds both for fricatives in isolated syllables (e.g., Behrens and Blumstein, 1988a; Baum and Blumstein, 1987) and in connected speech (Crystal and House, 1988).

In sum, acoustic studies focusing on the frication noise have shown that properties of the spectrum, amplitude, and duration of the noise can all serve to distinguish the sibilant /s,z,ʃ,ʒ/ from the nonsibilant /f,v,θ,ð/ fricatives. In addition, spectral properties serve to distinguish /s/ from /ʃ/, with /s/ having a concentration of energy in higher frequencies than /ʃ/. None of the noise properties alone, however, seems adequate to distinguish /f,v/ from /θ,ð/. More recent metrics such as spectral moments, locus equations, and relative amplitude show some promise for the distinction between labiodental and dental fricatives, although studies examining all eight fricatives with these metrics are few. The present study therefore consists of a comprehensive investigation of English fricatives, incorporating both recent and more traditional approaches with the aim of establishing stable acoustic cues to all four places of fricative articulation.

II. EXPERIMENT

A. Method

1. Participants

Twenty speakers (ten females and ten males) were recruited from the Cornell University student population. All were native speakers of American English, representing a variety of regional backgrounds. No participants reported any known history of either speech of hearing impairment. Participants were paid for their participation.

2. Materials

The eight English fricatives /f,v,θ,ð,s,z,ʃ,ʒ/ were recorded in consonant–vowel–consonant (CVC) syllables in the carrier phrase ''Say — again.'' The fricatives were in initial position, followed by each of six vowels /i,e,æ,ɑ,o,u/. The final consonant was always /p/. Each CVC token was repeated three times, yielding a total of 144 tokens per subject (8 fricatives$\times$6 vowels$\times$3 repetitions).

3. Procedure and analysis

Speakers were recorded in the Cornell Phonetics Laboratory, in a soundproof booth (IAC) with a high-quality microphone (Electro-Voice RE20), microphone pre-amp (Gaines Audio MP-1), and cassette deck (Carver TD1700). The microphone was placed at approximately a 45-deg angle and 15 cm away from the corner of the speaker's mouth, to prevent turbulence due to direct airflow from impinging on the microphone.

All recordings were sampled at 22 kHz (16-bit quantization, 11-kHz low-pass filter) on a Sun SPARCstation 5. All measurements were made using Entropics Systems' WAVES +/ESPS software. Fricative segmentation involved the simultaneous consultation of waveform and wideband spectrogram. Fricative onset was defined as the point at which high-frequency energy first appeared on the spectrogram and/or the point at which the number of zero crossings rapidly increased. Frication offset for voiceless fricatives was defined as the intensity minimum immediately preceding the onset of vowel periodicity. For voiced fricatives, the earliest pitch period exhibiting a change in the waveform from that seen throughout the initial frication was identified. The zero crossing of the preceding pitch period was then designated as the end of the voiced fricative (see Yeni-Komshian and Soli, 1981). Word duration was defined as the interval between fricative onset and the syllable-final /p/ release burst.

Spectral peak location of the fricatives was examined using a 40-ms full Hamming window placed in the middle of the frication noise. This larger window size yields better resolution in the frequency domain, at the expense of resolution in the temporal domain. Since fricatives are characterized by a relatively stationary articulatory configuration, the advantage of increased frequency resolution outweighs the disadvantage of decreased temporal resolution. A previous comparison of spectral properties of fricatives as measured at onset, midpoint, and offset of the frication noise showed that these properties are relatively stable throughout the noise portion, with high-frequency peaks more likely to emerge in the middle and end of the noise (Behrens and Blumstein, 1988a). Spectral peak estimation was based on spectra generated by means of FFT (fast Fourier transform) and LPC (linear predictive coding). For both FFT and LPC, a 40-ms full Hamming window was used, with a pre-emphasis factor of 98%. For LPC, 24 poles were used. LPC spectra were computed to examine if their peaks matched those of the FFT spectra. Spectral peak is defined here as the highest-amplitude peak of the FFT spectrum.

Spectral moments were computed following the procedures described by Forrest *et al.* (1988) with a few modifications. FFTs were calculated using a 40-ms full Hamming window (as compared to Forrest *et al.*'s 20-ms window) at four different locations in the fricative: onset, middle, and end, as well as centered over fricative offset. For example, the first window included the first 40 ms of the fricative, while the last window spanned the final 20 ms of the fricative and the first 20 ms of the following vowel. Each FFT was treated as a random probability distribution from which the first four moments were calculated. Moments were calculated from both linear and bark-transformed spectra. Only moments based on linear spectra are reported here, since there was no substantial difference between them and bark-transformed spectra.

Locus equations were derived using the procedure described by Sussman and Shore (1996) for fricatives. For both voiced and voiceless fricatives, $F2$ was measured at vowel onset and midway in the vowel. Specifically, $F2$ at vowel onset was estimated by means of FFT spectra, with a 23.3-ms full Hamming window (similar to Sussman and Shore, 1996, and Fowler, 1994) starting at the first glottal pulse following cessation of the fricative. (These data were also used in the analysis of $F2$ onset.) Similarly, $F2$ at vowel nucleus was estimated by placing a 23.3-ms window at the vowel's midpoint. In the case of the diphthongized vowels /e/ and /o/, data points from the vowel offglide were excluded. In addition to FFT spectra, wideband spectrograms and LPC spectra were also consulted.

Root-mean-square (rms) amplitude in dB was measured for the entire noise portion of each fricative token. In order to normalize for intensity differences among speakers, a difference of fricative amplitude minus vowel amplitude (''normalized amplitude'') was calculated, where vowel amplitude was defined as rms amplitude (in dB) averaged over three consecutive pitch periods at the point of maximum vowel amplitude (see Behrens and Blumstein, 1988b).

Relative amplitude in dB was measured as described in Hedrick and Ohde (1993). Briefly, for the vowel, a discrete Fourier transform (DFT) was derived at vowel onset, using a 23.3-ms Hamming window. The amplitude (in dB) of the component at $F3$ for /s,z,ʃ,ʒ/ and at $F5$ for /f,v,θ,ð/ was measured. For the fricative, a DFT was then derived at the center of the fricative, using a 23.3-ms Hamming window. The amplitude (in dB) of the component in the same frequency region as that selected for the vowel was measured. Relative amplitude was then expressed as the difference between fricative amplitude and vowel amplitude.

B. Results

1. Spectral properties

a. Spectral peak location. A four-way analysis of variance (ANOVA) (place×voicing×vowel×gender) revealed a main effect for place of articulation [$F(3,2876)=1083.72$, $p<0.0001$; $\eta^2=0.512$]. Averaged across speakers, voicing, and vowel context, spectral peak location for the labiodentals was 7733 Hz, for dentals 7470 Hz, for alveolars 6839 Hz, and for palato-alveolars 3820 Hz. Spectral peak location thus decreases in frequency as place of articulation moves further back in the oral cavity. Bonferroni *post hoc* tests indicated that all four places of articulation were significantly different from each other in terms of spectral peak location ($p<0.003$ for the contrast between labiodentals and dentals, $p<0.0001$ for all other contrasts).

A main effect of voicing [$F(1,2876)=30.65$, $p<0.0001$; $\eta^2=0.01$] indicated that voiceless fricatives had spectral peaks at a significantly higher frequency (6612 Hz) than voiced fricatives (6310 Hz). A place×voicing interaction [$F(3,2876)=12.14$, $p<0.0001$; $\eta^2=0.012$] and subsequent *post hoc* tests revealed that the difference in spectral

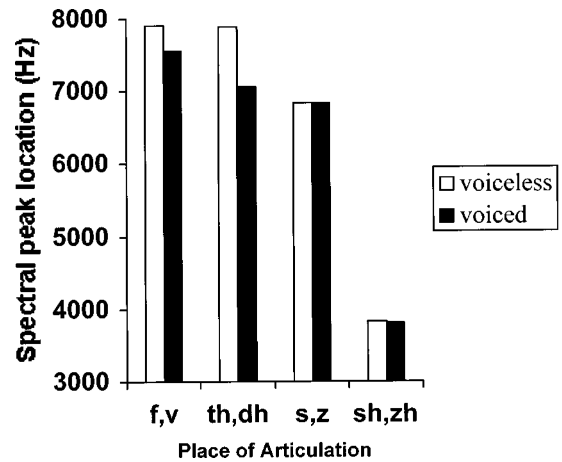

FIG. 1. Mean spectral peak location as a function of place of articulation and voicing (in Hz, averaged across vowels, and male and female speakers). Spectral peak location was computed over a 40-ms window placed in the middle of the fricative noise.

peak location between voiceless and voiced fricatives was carried by the nonsibilant fricatives. As shown in Fig. 1, while there was no difference between /s/ and /z/ and between /ʃ/ and /ʒ/, the differences in spectral peak between /θ/ and /ð/ (832 Hz) and between /f/ and /v/ (340 Hz) were significant.

A main effect of gender [$F(1,2876)=154.15$, $p<0.0001$; $\eta^2=0.047$] indicated that, as expected, mean spectral peak location was significantly higher for female (6800 Hz) than for male (6122 Hz) speakers. A place×gender interaction [$F(3,2876)=34.61$, $p<0.0001$; $\eta^2=0.032$] and subsequent *post hoc* tests revealed that the pattern of males and females was not entirely parallel. As shown in Fig. 2, male speakers show a pattern in which spectral peak frequency decreases as place moves back; however, female speakers are different in that their dentals have a higher spectral peak frequency than their labiodentals.

Finally, no main effect was observed for vowel ($p>0.878$). A significant place×vowel interaction [$F(15,2876)=3.67$, $p<0.001$; $\eta^2=0.017$] and *post hoc* tests indicated that spectral peak location of only /s,z/ varied

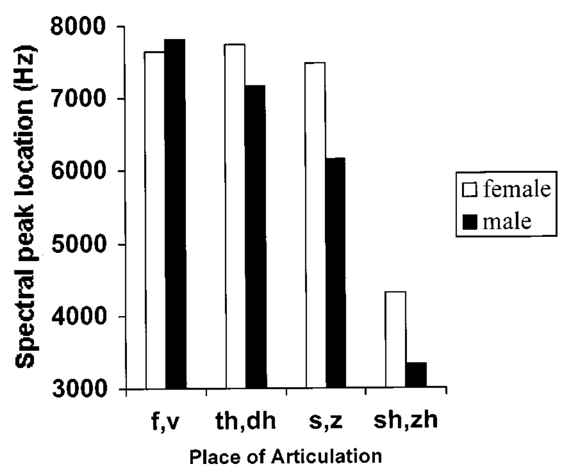

FIG. 2. Mean spectral peak location for male and female speakers as a function of place of articulation (in Hz, averaged across vowels, and voiced and voiceless tokens).

TABLE I. Mean spectral moment values for each place of articulation, averaged across speakers, window location, voiced and voiceless tokens, and vowel context.

Place of articulation	Spectral mean (Hz)	Variance (MHz)	Skewness	Kurtosis
/f,v/	5108	6.37	0.077	2.11
/T,D/	5137	6.19	−0.083	1.27
/s,z/	6133	2.92	−0.229	2.36
/S,Z/	4229	3.38	0.693	0.42

as a function of vowel context: spectral peak for /s,z/ was significantly lower in the context of the back-rounded vowels /o,u/.

b. Spectral moments. One-way ANOVAs were conducted for place, voicing, and gender across window locations with the four moments as dependent variables. For spectral mean, a main effect obtained for place of articulation [$F(3,11520)=488.16$, $p<0.0001$; $\eta^2=0.113$]. As shown in Table I, spectral mean was highest for /s,z/ (6133 Hz) and lowest for /ʃ,ʒ/ (4229 Hz), and this difference was significant ($p<0.0001$). Spectral mean values for /f,v/ (5108 Hz) and /θ,ð/ (5137 Hz) fell in between and were not significantly different from each other ($p>0.9$). For spectral variance, a main effect obtained for place of articulation [$F(3,11520)=1216.02$, $p<0.0001$; $\eta^2=0.241$]. Variance was low for the sibilant fricatives and high for the nonsibilants. Differences among all places were highly significant ($p<0.0001$) except that between /f,v/ and /θ,ð/ which was only marginally so ($p>0.066$). A main effect for skewness [$F(3,11520)=332.24$, $p<0.0001$; $\eta^2=0.080$] and subsequent *post hoc* tests revealed that skewness distinguished all four places of articulation ($p<0.0001$). Skewness was highest for /ʃ,ʒ/, indicating that the palato-alveolars had the strongest concentration of energy in the lower frequencies. Finally, there was a main effect for kurtosis [$F(3,11520)=90.69$, $p<0.0001$; $\eta^2=0.023$]. Kurtosis failed to distinguish /f,v/ from /s,z/ ($p>0.293$), both of which had high values indicating peaked spectra. All other comparisons were significant ($p<0.0001$).

A main effect was obtained for voice for all four moments. Effect size was rather small, with η^2 ranging from 0.001 for kurtosis to 0.069 for variance. Voiceless fricatives were characterized by higher values for spectral mean (5267 Hz), skewness (0.238), and kurtosis (1.70) than voiced fricatives (5036 Hz; −0.009; and 1.38, respectively). Thus, compared to voiced fricatives, the spectra of voiceless fricatives had a concentration of energy towards slightly lower frequencies and slightly better defined peaks. In addition, voiced fricatives had a significantly greater variance (5.56 MHz) than voiceless ones (3.87 MHz).

Finally, a main effect for gender indicated that females exhibited significantly higher values than males for spectral mean (5286 vs 5018 Hz), variance (4.9 vs 4.5 MHz), and kurtosis (1.64 vs 1.44), while female skewness values were significantly lower than those of males (0.084 vs 0.145). Thus, compared to males, the spectra of female speakers had clearer peaks and a concentration of energy towards higher frequencies. It must be noted that effect size was very small, with η^2 ranging from 0.001 for skewness to 0.004 for spectral mean. A table with values for each moment at each window location for voiced and voiceless tokens and female and male speakers can be found in the Appendix.

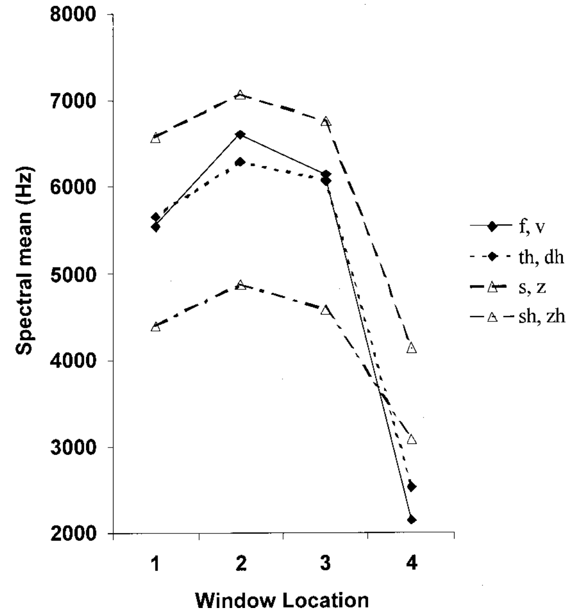

FIG. 3. Spectral mean (moment 1) in Hz (averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation. Window locations 1, 2, and 3 refer to the first, middle, and last 40 ms of the fricative noise, respectively; window location 4 includes the final 20 ms of the fricative and the first 20 ms of the following vowel.

In order to assess the importance of acoustic information at different positions in the speech signal, four-way ANOVAs (place×voicing×vowel×gender) and subsequent Bonferroni *post hoc* tests were conducted for each moment at each window location. Figures 3 through 6 show moment values for each place of articulation as a function of window location, for moments 1 through 4, respectively. Results of the statistical tests are summarized in Table II. This table shows the number of places of articulation differentiated by a given moment at a given window location. It is clear that spectral moments distinguish at least three places of articulation at all window locations, and four places in the majority of cases. All but two confusions involved a lack of differentiation between /f,v/ and /θ,ð/.

M1 (spectral mean) (Fig. 3) distinguishes all four places of articulation at the second and fourth window locations. In general, /s,z/ have the highest spectral mean, and /ʃ,ʒ/ the lowest. The nonsibilants' spectral means fall in between. M2 (variance) (Fig. 4) distinguishes all places at all but the second window location. Variance is low for the sibilant fricatives and high for the nonsibilants. M3 (skewness) (Fig. 5) distinguishes all places at all but the third window location. Skewness is always positive for /ʃ,ʒ/, indicating a concentration of energy in the lower frequencies. Skewness increases substantially at the fricative–vowel transition (window 4) for the nonsibilants, reflecting the predominance of low-frequency over high-frequency energy as the vowel begins. M4 (kurtosis) (Fig. 6) distinguishes all places at only the first window location. Kurtosis is highest for /s,z/, indi-

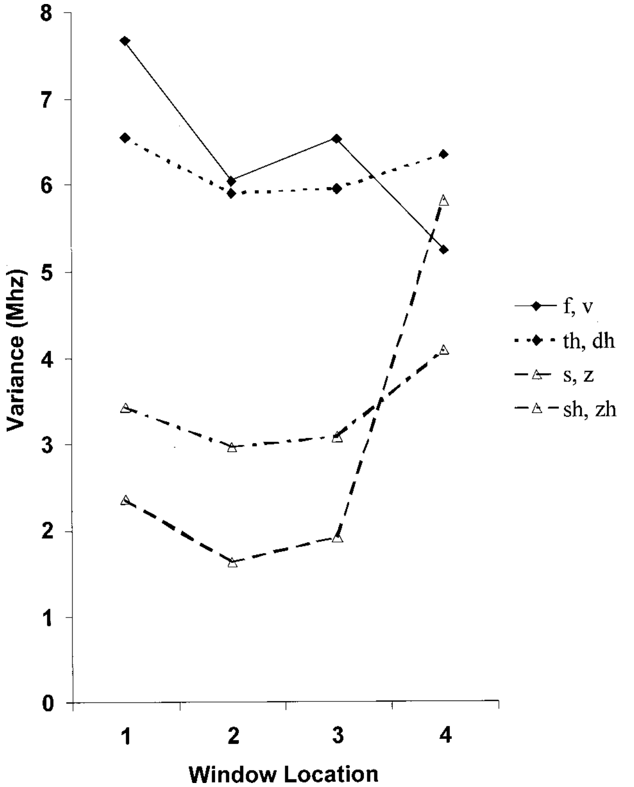

FIG. 4. Spectral variance (moment 2) in MHz (averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation.

cating a spectrum with clearly defined peaks. Kurtosis yields the only confusions that do not involve /f,v/ and /θ,ð/; instead, /f,v/ and /ʃ,ʒ/ are nondistinct at fricative offset while /s,z/ and /ʃ,ʒ/ are not differentiated in the transition region between fricative and vowel. The effect for each moment is quite sizable at nearly every window location. For spectral mean, η^2 ranges from 0.296 to 0.387, for variance from 0.103 to 0.545, and for skewness from 0.321 to 0.380. Finally, effects were somewhat weaker for kurtosis, with η^2 ranging from 0.066 to 0.281.

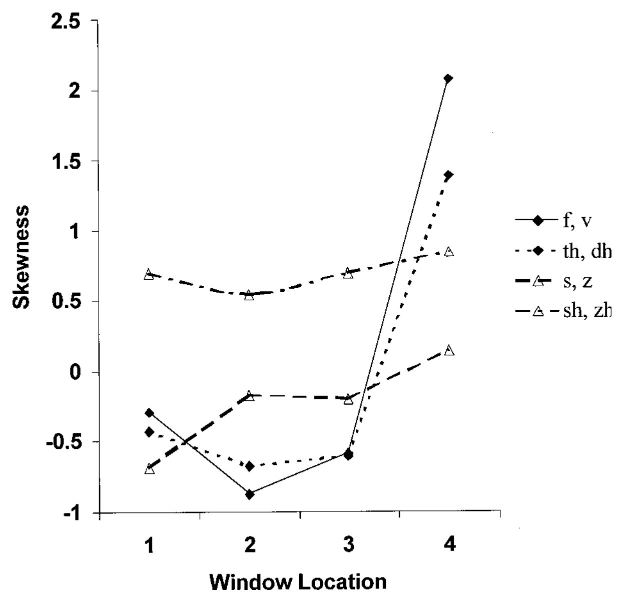

FIG. 5. Spectral skewness (moment 3; averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation.

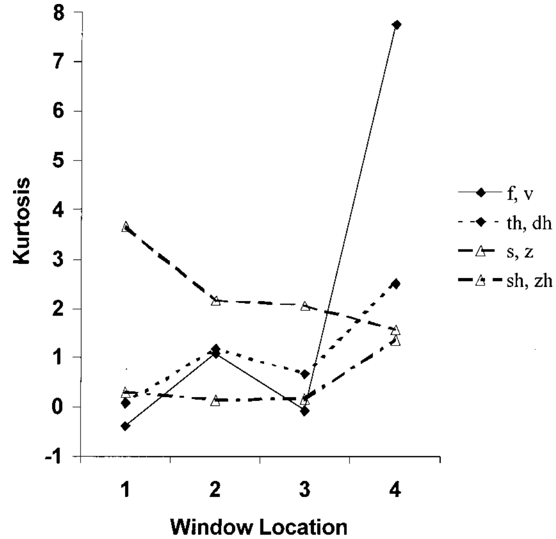

FIG. 6. Spectral kurtosis (moment 4; averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation.

2. Transition information

a. Locus equations. Following Sussman *et al.* (1991), slope and *y*-intercept values were derived for each place of articulation for each speaker, averaged across vowel context. Table III presents slope and *y*-intercept values for each place of articulation for females and males, averaged across all vowel contexts.[1] A two-way ANOVA (place×gender) for slope revealed a main effect for place of articulation [$F(3,72)=33.25$, $p<0.0001$; $\eta^2=0.581$]. *Post hoc* tests indicated that only the slope value for /f,v/ was significantly different from that of the other three places of articulation. For the *y* intercept, a main effect was observed for place [$F(3,72)=51.32$, $p<0.0001$; $\eta^2=0.681$], with subsequent *post hoc* tests revealing that, while *y*-intercept values were distinct for /f,v/ and /ʃ,ʒ/, they were not for /θ,ð/ and /s,z/. A main effect was also observed for gender [$F(1,72)=19.79$, $p<0.0001$; $\eta^2=0.216$], indicating that the *y* intercept was significantly higher for females (900 Hz) than for males (708 Hz).

b. F2 onset values. Table IV presents *F*2 onset values for each place of articulation, averaged across all speakers

TABLE II. Number of places of articulation (out of 4) distinguished by each moment for each window location. A score of 3 was in all but two cases due to confusion of /f,v/ with /θ,ð/. The exceptions were the confusion of /f,v/ with /ʃ,ʒ/ for kurtosis at fricative offset, and of /s,z/ with /ʃ,ʒ/ for kurtosis at the fricative–vowel transition.

	Window			
Moment	Onset	Middle	Offset	Transition
Spectral mean	3	4	3	4
Variance	4	3	4	4
Skewness	4	4	3	4
Kurtosis	4	3	3	3

TABLE III. Mean slope (k) and y intercept (c in Hz) (averaged across voiced and voiceless tokens and vowels) as a function of place of articulation and speaker gender.

	/f,v/		/θ,ð/		/s,z/		/ʃ,ʒ/	
	k	c	k	c	k	c	k	c
Females	0.766	413	0.530	940	0.501	1004	0.452	1242
Males	0.770	299	0.529	819	0.533	825	0.557	887
Mean	0.768	356	0.530	879	0.517	914	0.505	1065

and vowel contexts. A four-way ANOVA (place×voicing ×vowel×gender) revealed a main effect for place [$F(3,2876)=147.25$, $p<0.0001$; $\eta^2=0.133$]. $F2$ onset values generally increased as place of articulation moved further back in the vocal tract. However, Bonferroni *post hoc* tests indicated that the difference between /θ,ð/ and /s,z/ was not significant.

A main effect of vowel [$F(5,2876)=481.74$, $p<0.0001$; $\eta^2=0.456$] obtained: $F2$ onset was 2334 Hz in the context of /i/, 2010 Hz before /e/, 1820 Hz before /æ/, 1710 Hz before /u/, 1526 Hz before /o/, and 1512 Hz before /A/. *Post hoc* tests indicated that $F2$ onset values were higher for front vowels compared to back vowels and that $F2$ onset values significantly increased as a function of increasing vowel height. All differences among vowels were significant except that between /o/ and /A/. There was no main effect of voicing. A place×vowel interaction [$F(15,2876)=22.52$, $p<0.0001$; $\eta^2=0.105$] and *post hoc* tests revealed that while $F2$ onset differed significantly with each vowel for /f,v/ and /s,z/, the vowel context effects for /θ,ð/ and /ʃ,ʒ/ were restricted to /i,e/. A place×voicing interaction [$F(3,2876)=6.85$, $p<0.0001$; $\eta^2=0.007$] and *post hoc* tests revealed that while there was no difference in $F2$ onset between voiced and voiceless tokens of the labiodental, dental, and alveolar fricatives, $F2$ onset was significantly higher for /ʒ/ (2040 Hz) than for /ʃ/ (1925 Hz). Finally, as expected, there was a main effect for gender [$F(1,2876)=563.9$, $p<0.0001$; $\eta^2=0.164$]; $F2$ onset was significantly higher for females (1967 Hz) than for males (1689 Hz).

3. Amplitude

a. Overall amplitude. Table V shows mean noise amplitude, vowel amplitude, and the difference between the two (''normalized amplitude'') as a function of place of articulation. Using normalized amplitude as the dependent variable, a four-way ANOVA (place×voicing×vowel×gender) revealed a main effect for place [$F(3,2876)=1489.51$, $p<0.0001$; $\eta^2=0.591$]. Bonferroni *post hoc* tests indicated

TABLE IV. Mean $F2$-onset values (Hz) (averaged across voiced and voiceless tokens, and vowels) as a function of place of articulation and speaker gender.

	/f,v/	/θ,ð/	/s,z/	/ʃ,ʒ/
Females	1815	1969	1967	2115
Males	1509	1701	1697	1849
Mean	1661	1833	1832	1982

TABLE V. Mean noise amplitude, vowel amplitude (in dB, averaged across speakers and vowels), and normalized amplitude for each fricative. Normalized amplitude refers to noise amplitude minus vowel amplitude in dB. Mean normalized amplitude refers to normalized amplitude for each place of articulation.

Fricative	Noise amplitude	Vowel amplitude	Normalized amplitude	Mean norm. ampl.
/f/	55.7	76.5	−20.8	−17
/v/	63.2	76.3	−13.1	
/θ/	54.7	76.6	−21.9	−18
/ð/	62.7	76.7	−14.0	
/s/	64.9	75.9	−11.0	−10
/z/	67.7	76.7	−9.0	
/ʃ/	66.4	76.3	−9.9	−9
/ʒ/	68.2	76.5	−8.3	

that all four places of articulation were significantly different from each other in terms of normalized amplitude.

A main effect of voicing [$F(1,2876)=1644.06$, $p<0.0001$; $\eta^2=0.347$] indicated that voiced fricatives (−15.9 dB) had a significantly smaller amplitude relative to the vowel than their voiceless counterparts (−11.1 dB). A main effect of vowel [$F(5,2876)=11.94$, $p<0.0001$; $\eta^2=0.019$] was obtained. The normalized amplitude preceding /o/ was −14 dB, /u/: −13.8 dB, /e/: −13.8 dB, /æ/: −13.6 dB, /ɑ/: −13 dB, /i/: −12.7 dB. Bonferroni *post hoc* tests indicated that only the amplitude difference for /i/ and /ɑ/ differed from that for all other vowels. There was no main effect of gender. Finally, a place×voicing interaction [$F(3,2876)=214.15$, $p<0.0001$; $\eta^2=0.172$] and *post hoc* tests indicated that the difference between voiced and voiceless fricatives was much greater for the nonsibilants than for the sibilants.

b. Relative amplitude. Figure 7 presents relative amplitude values for each place of articulation for voiced and voiceless tokens. A four-way ANOVA (place×voicing ×vowel×gender) revealed a main effect for place [$F(3,2876)=458.27$, $p<0.0001$; $\eta^2=0.308$]. Bonferroni *post hoc* tests indicated that all four places of articulation were significantly different ($p<0.0001$ for all comparisons).

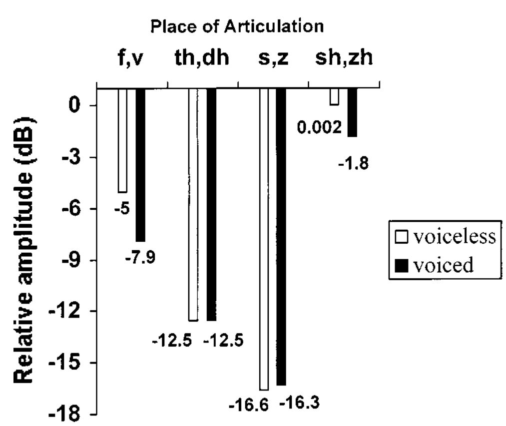

FIG. 7. Relative amplitude (dB) as a function of place of articulation and voicing (see Sec. II A 3 for calculation).

TABLE VI. Mean frication duration, total word duration (in ms, averaged across speakers and vowels), and normalized duration for each fricative. Normalized duration refers to the ratio of frication duration over word duration. Mean normalized duration refers to normalized duration for each place of articulation.

Fricative	Frication duration	Word duration	Normalized duration	Mean norm. dur.
/f/	166	395	0.420	
				0.333
/v/	80	326	0.245	
/θ/	163	393	0.415	
				0.340
/ð/	88	333	0.264	
/s/	178	406	0.438	
				0.382
/z/	118	362	0.326	
/ʃ/	178	397	0.448	
				0.393
/ʒ/	123	364	0.338	

A main effect of voicing $[F(1,2876)=14.03, p<0.0001; \eta^2=0.005]$ indicated that voiceless fricatives (−8.5 dB) had a significantly greater relative amplitude than their voiced counterparts (−9.6 dB). A main effect of vowel $[F(5,2876)=6.36, p<0.0001; \eta^2=0.01]$ was also obtained. Relative amplitude preceding /e/ was −10.4 dB, /o/: −9.1 dB, /i/, /æ/: −8.9 dB, /A/: −8.6 dB, /u/: −7.6 dB. Bonferroni *post hoc* tests indicated that only the relative amplitude for /e/ and /u/ differed from that for all other vowels. Finally, a main effect of gender $[F(1,2876)=28.73, p<0.0001; \eta^2=0.009]$ indicated that relative amplitude values were smaller for females (−9.8 dB) than for males (−8.1 dB).

A place by vowel interaction $[F(15,2876)=4.95, p<0.0001; \eta^2=0.023]$ revealed that while vowel-intrinsic differences in relative amplitude were similar across most places of articulation, labiodental /f,v/ deviated from this pattern, showing much lower values for the back vowels /ɑ,o,u/ as compared to the front vowels /i,e,æ/. A place by voicing interaction $[F(3,2876)=4.82, p<0.002; \eta^2=0.005]$ was due to the fact that while there was no difference in relative amplitude between /θ/ and /ð/ or between /s/ and /z/, the difference in relative amplitude between /f/ and /v/ (2.9 dB) and that between /ʃ/ and /ʒ/ (1.8 dB) was significant. A place by gender interaction $[F(3,2876)=6.01, p<0.0001; \eta^2=0.006]$ indicated that the gender difference in relative amplitude was most pronounced for /θ,ð/. Finally, a voicing by gender interaction $[F(1,2876)=13.74, p<0.0001; \eta^2=0.004]$ was obtained because the voicing difference in relative amplitude was mostly due to the male speakers.

4. Noise duration

Table VI shows mean frication duration, word duration, and their ratio as a function of place of articulation. Analyses involving duration have typically focused on absolute frication duration (e.g., Behrens and Blumstein, 1988a). An initial four-way ANOVA (place×voicing×vowel×gender) with fricative duration as the dependent variable revealed a main effect for place $[F(3,2876)=327.69, p<0.0001; \eta^2=0.092]$. However, Bonferroni *post hoc* tests indicated that all four places were not significantly different but that noise duration of the nonsibilant fricatives was significantly shorter than that of the sibilant fricatives. Because absolute duration may vary as a function of speaking rate, a four-way ANOVA (place×voicing×vowel×gender) was conducted using ''normalized duration,'' defined as the ratio of fricative duration over word duration, as the dependent variable. A main effect for place $[F(3,2876)=236.56, p<0.0001; \eta^2=0.187]$ and subsequent *post hoc* tests indicated that only the difference between /f,v/ and /θ,ð/ was not significant. All other comparisons were significant at the $p<0.0001$ level, except that between /s,z/ and /ʃ,ʒ/ $(p<0.001)$. A main effect of voicing $[F(1,2876)=4547.30, p<0.0001; \eta^2=0.595]$ indicated normalized duration was significantly greater for voiceless fricatives (0.429) than for voiced ones (0.293). A place×voicing interaction $[F(3,2876)=58.28, p<0.0001; \eta^2=0.053]$ and *post hoc* tests indicated that the effect of voicing was more pronounced for the nonsibilants than for the sibilants. A main effect of gender $[F(1,2876)=66.32, p<0.0001; \eta^2=0.021]$ indicated that fricatives produced by female speakers (0.351) had slightly smaller normalized durations than those produced by male speakers (0.368). Finally, a main effect of vowel $[F(5,2876)=138.04, p<0.0001; \eta^2=0.182]$ was obtained. Bonferroni *post hoc* tests indicated that normalized duration decreased with decreasing vowel height: normalized duration preceding /i/ was 0.390, /u/: 0.400, /e/: 0.356, /o/: 0.357, /æ/: 0.324, and /ɑ/: 0.324. Differences between vowels of different heights were all significant $(p<0.0001)$; differences between vowels of the same height were not significant (/i/ vs /u/: $p>0.098$; /e/ vs /o/ and /æ/ vs /ɑ/: $p>0.90$).

5. Discriminant analysis

Discriminant analysis was performed to evaluate the extent to which the acoustic parameters reported here could categorize the fricatives in terms of place of articulation. All acoustic parameters discussed above, except for locus equations because they are not a property of individual productions, were entered as predictors. For the moments, each moment at each window location was entered. A stepwise linear discriminant analysis was conducted with 21 predictors (spectral peak location, 4 moments×4 window locations, $F2$ onset, normalized amplitude, relative amplitude, and normalized duration). Classification results are based on the jackknife method, whereby each speaker in turn was used as the testing speaker with training being done on the 19 remaining speakers. Final classification scores were then averaged across the 20 testing speakers.

Classification scores for each place of articulation based on the jackknife method are shown in Table VII. Overall classification accuracy was 77%. While classification of all four places of articulation was significantly above chance, it was clearly better for the sibilants (88%) than for the nonsibilants (66%). Classification errors rarely crossed the sibilant/nonsibilant distinction. That is, labiodentals and dentals were mostly confused with each other, and the same was true of alveolars and palato-alveolars.

In order to assess the contribution of each predictor variable to the discriminant functions, the standardized canonical discriminant function coefficients were analyzed (Klecka,

TABLE VII. Predicted group membership (%) in terms of fricative place of articulation. Classification is based on a stepwise linear discriminant analysis with all acoustic measures as predictors (see the text). Bold percentages indicate correct classification rates. Overall correct classification was 77%.

	Predicted group membership			
	/f,v/	/θ,ð/	/s,z/	/ʃ,ʒ/
/f,v/	**68**	27	3	2
/T,D/	26	**64**	6	4
/s,z/	1	4	**85**	9
/S,Z/	4	0	5	**91**

1980). These coefficients suggested that spectral peak location, normalized amplitude, relative amplitude, and spectral mean at fricative onset and midpoint were the main parameters used for fricative classification. A subsequent discriminant analysis with only those five predictors yielded an overall classification rate of 69%. Exclusion of the spectral mean at onset and midpoint only slightly decreased classification accuracy to 67%. Combinations of only two predictors yielded substantially lower rates, below 60% accuracy. Overall, then, spectral peak location, normalized amplitude, and relative amplitude served to distinguish the fricatives in terms of place of articulation with reasonable accuracy. Classification rates for this analysis were as follows: /f,v/: 53%, /θ,ð/: 48%, /s,z/: 81%, and /ʃ,ʒ/: 88%.

C. Discussion

The present results from 20 speakers indicate that spectral and amplitudinal information provide the most critical information to place of articulation in fricatives. In agreement with previous research on spectral properties of the frication noise (e.g., Hughes and Halle, 1956; Strevens, 1960; Heinz and Stevens, 1961; Jassem, 1965; Shadle, 1990; Behrens and Blumstein, 1988a), spectral peak location distinguishes sibilants from nonsibilants, and alveolar /s,z/ from palato-alveolar /ʃ,ʒ/. Importantly, however, the present results indicate that spectral peak location also distinguished /f,v/ from /θ,ð/. Thus, contrary to previous reports, spectral peak location does distinguish all four places of articulation.

Spectral moments also served to distinguish all four places of articulation. If the success of a moment is measured in terms of the number of places it distinguished at each location in the fricative, m2 (variance) and m3 (skewness) performed best (see Table II). Across moments, a comparison of window locations suggests that window locations 1 and 4 (noise onset and fricative–vowel transition region, respectively) contain the most distinctive information (see Table II).

Few studies report all four moment values or values for all fricatives. Most studies have focused on the spectral mean of /s/ and /ʃ/ (e.g., Nittrouer *et al.*, 1989; Baum and McNutt, 1990; Waldstein and Baum, 1991). Interestingly, those studies that did include more moments usually excluded spectral variance, perhaps because Forrest *et al.* (1988) excluded it from their original analysis since it did not appear to distinguish among any of the obstruents in their study. The fact that variance turns out to be a robust cue to place in the present study may be the result of sampling a larger and more representative number of speakers and tokens, as compared to the rather small database of Forrest *et al.* (1988), consisting of only five target words and no voiced fricatives.

Generally, in those cases in which direct comparisons could be made, there is good agreement in terms of spectral mean between the present study and previous research (Tomiak, 1990; Nittrouer, 1995; Avery and Liss, 1996; McFarland *et al.*, 1996; Tjaden and Turner, 1996). In terms of spectral variance, there is good agreement with Tomiak (1990), the only other study reporting values for the second moment.

With respect to the third moment, the present finding of a negative skewness for /s/ and a positive skewness for /ʃ/ is supported by previous findings by Nittrouer (1995) and McFarland *et al.* (1996) but differs from Tomiak (1990) and Avery and Liss (1996), who reported a greater positive skewness for /s/ than for /ʃ/. Shadle and Mair (1996) did report that variance was a more reliable indicator of fricative place than skewness, although the authors report no overall analysis of place of articulation for these moments and only one female and one male speaker were included. Finally, our finding of a large positive kurtosis for /s/ and a small positive or negative kurtosis for /ʃ/ is in agreement with Tomiak (1990), Nittrouer (1995), Avery and Liss (1996), and McFarland *et al.* (1996). In general, the present data clearly show that four places of articulation were distinguished by most moments at most window locations.

Both normalized and relative amplitude properties were also found to be consistent cues to fricative place of articulation. In terms of normalized amplitude, sibilant fricatives had a greater noise amplitude than nonsibilants; moreover, within the group of sibilants, palato-alveolar /ʃ,ʒ/ had a greater noise amplitude than alveolar /s,z/, while for the nonsibilants labiodental /f,v/ had a greater amplitude than interdental /θ,ð/. Previous research supports the role of noise amplitude in the sibilant/nonsibilant distinction (e.g., Strevens, 1960; Behrens and Blumstein, 1988a, b). In particular, in their study of /f,θ,s,ʃ/, Behrens and Blumstein (1988a, b) reported overall amplitude differences of similar magnitude as the present study. However, contrary to these studies, the present study also indicates that normalized amplitude can distinguish place of articulation *within* these two groups. One of the three speakers analyzed by Behrens and Blumstein (1988a) showed a significantly greater amplitude for /f/ compared to /θ/, suggesting that the difference in the present study may be due to our larger sampling of speakers and tokens.

Relative amplitude also distinguished all four places of articulation. Relative amplitude was small for the palato-alveolars, indicating that /ʃ,ʒ/ has a major concentration of energy in the region corresponding to $F3$ of the following vowel. For the other places, relative amplitude was seen to decrease as place moved further back in the oral cavity. Fricative amplitude in the $F5$ region is smaller for /θ,ð/ than /f,v/. In addition, the large difference between fricative and vowel amplitude in the $F3$ region for /s,z/ supports the notion that these fricatives have their major energy in a frequency region well above $F3$. The present findings are qualitatively in line with those of Hedrick and Ohde (1993), who also reported a much greater relative amplitude for /ʃ/ than for /s/ for their speaker. The present value for /s,z/ (−16.5 dB) is very similar to that used by Hedrick and Ohde (−17 dB), while that for /ʃ,ʒ/ (−0.9 dB) is much lower than theirs (16 dB). As mentioned previously, research on relative amplitude has exclusively focused on perception, which makes it impossible to compare the present findings to earlier work in any detail. However, the present acoustic data are corroborated by perceptual data on relative amplitude. Perceptually, the crossover boundary between /s/ and /ʃ/ has been shown to correspond to a relative amplitude of approximately −7.5 dB (Hedrick and Ohde, 1993), which is also halfway in between the relative amplitude measurements for /s/ and /ʃ/ reported here.

A comparison of the results from ANOVA and discriminant analysis reveals a high degree of agreement. Using η^2 to select those acoustic parameters from the ANOVAs that contribute most to distinguishing all four places of articulation, normalized amplitude ($\eta^2=0.591$), spectral peak location ($\eta^2=0.512$), relative amplitude ($\eta^2=0.308$), and spectral variance (M2) ($\eta^2=0.287$) were identified as the primary contributors. Results from the discriminant analyses also identified three of these parameters (spectral peak location as well as normalized and relative amplitude) as the strongest predictors of group membership.

In the present study, a number of measures were shown not to distinguish fricative place of articulation. These include $F2$ transition properties and noise duration. Properties specific to the $F2$ transition failed to distinguish all fricative places of articulation. Both the combination of slope and intercept values of the locus equations and the $F2$ onset values could only single out the labiodental and palato-alveolar fricatives. As for normalized noise duration, sibilant fricatives were longer than nonsibilants, supporting similar findings by Behrens and Blumstein (1988a) based on absolute noise duration. In addition, voiceless fricatives were substantially longer than their voiced counterparts (see also Baum and Blumstein, 1987; Behrens and Blumstein, 1988a; Crystal and House, 1988; Jongman, 1989). A new finding in the current study is that normalized duration also distinguished /s,z/ from /ʃ,ʒ/, which may be due to the use of normalized duration rather than absolute duration. However, even normalized duration failed to distinguish /f,v/ from /θ,ð/. Given the present findings with other parameters and the extent of the present database, it must be concluded that $F2$ transition properties and noise duration do not reliably distinguish place of articulation in fricatives.

In sum, the present study indicates that several acoustic properties serve to distinguish all four places of fricative articulation. These properties include both spectral (spectral peak location, spectral moments) and amplitudinal (normalized and relative amplitude) parameters, as well as both static (spectral peak location, spectral moments, normalized amplitude) and dynamic (relative amplitude) information. This finding suggests that, contrary to earlier reports, acoustic properties can provide robust information about all four places of articulation, despite variation in speaker, vowel context, and voicing. Future research will need to address the extent to which the properties identified here contribute to perception of place of articulation in fricatives.

ACKNOWLEDGMENTS

Portions of this research were conducted while the first author was at Cornell University. This research was supported (in part) by Research Grant Number 1 R29 DC 02537-01A1 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors thank Scott Gargash, Eric Evans, Michelle Spence, and Yue Wang for assistance, and Anders Lofqvist, Marija Tabain, Joan Sereno, Roel Smits, and an anonymous reviewer for comments on a previous version of this paper. Portions of this research were presented at the 135th meeting of the Acoustical Society of America and at the 1999 meeting of the American Speech Language and Hearing Association.

APPENDIX

Table AI gives values of the four spectral moments for each window location, as a function of voicing and speaker gender. Moments 1, 2, 3, and 4 refer to spectral mean (Hz), variance (MHz), skewness, and kurtosis, respectively. Window locations 1, 2, and 3 refer to the first, middle, and final 40 ms of the frication noise, respectively. Window location 4 refers to a window spanning the last 20 ms of the fricative and the first 20 ms of the following vowel.

TABLE AI. Values of the four spectral moments as a function of voicing and speaker gender.

		Females				Males			
		Window location				Window location			
Moment		1	2	3	4	1	2	3	4
1	voiceless	6149	6858	6320	2457	5822	6426	5862	2244
	voiced	5230	5883	5763	3629	4957	5652	5606	3573
2	voiceless	3.77	3.10	3.67	5.37	3.92	2.94	3.70	4.53
	voiced	6.26	5.68	5.44	5.93	6.06	4.80	4.67	5.64
3	voiceless	−0.1882	−0.3798	−0.2111	1.5576	−0.1064	−0.3139	−0.1081	1.6543
	voiced	−0.2624	−0.2580	−0.1906	0.6060	−0.1600	−0.2337	−0.2026	0.6286
4	voiceless	0.6238	0.9031	0.6125	4.619	0.2943	1.0144	0.6045	4.9541
	voiced	2.2613	1.0994	0.7272	2.3101	1.1438	1.2629	0.9329	1.3235

[1]These data were previously reported in a brief commentary (Jongman, 1998). They are repeated here in more detail and for the sake of completeness.

Avery, J. D., and Liss, J. M. (**1996**). ''Acoustic characteristics of less-masculine-sounding male speech,'' J. Acoust. Soc. Am. **99**, 3738–3748.

Baum, S. R., and Blumstein, S. E. (**1987**). ''Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English,'' J. Acoust. Soc. Am. **82**, 1073–1077.

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

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

Behrens, S. J., 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.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (**1952**). ''Some experiments on the perception of synthetic speech sounds,'' J. Acoust. Soc. Am. **24**, 597–606.

Crystal, T., and House, A. (**1988**). ''Segmental durations in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1955**). ''Acoustic loci and transitional cues for consonants,'' J. Acoust. Soc. Am. **27**, 769–773.

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

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**, 115–124.

Fowler, C. A. (**1994**). ''Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation,'' Percept. Psychophys. **55**, 597–611.

Hedrick, M. (**1997**). ''Effect of acoustic cues on labeling fricatives and affricates,'' J. Speech Lang. Hear. Res. **40**, 925–938.

Hedrick, M. S., and Carney, A. E. (**1997**). ''Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants,'' J. Speech Lang. Hear. Res. **40**, 1445–1457.

Hedrick, M. S., and Ohde, R. N. (**1993**). ''Effect of relative amplitude of frication on perception of place of articulation,'' J. Acoust. Soc. Am. **94**, 2005–2027.

Heinz, J. M., and Stevens, K. N. (**1961**). ''On the properties of voiceless fricative consonants,'' J. Acoust. Soc. Am. **33**, 589–596.

Hughes, G. W., and Halle, M. (**1956**). ''Spectral properties of fricative consonants,'' J. Acoust. Soc. Am. **28**, 303–310.

Jassem, W. (**1965**). ''Formants of fricative consonants,'' Lang. Speech **8**, 1–16.

Jongman, A. (**1989**). ''Duration of fricative noise required for identification of English fricatives,'' J. Acoust. Soc. Am. **85**, 1718–1725.

Jongman, A. (**1998**). ''Are locus equations sufficient or necessary for obstruent perception?,'' Behav. Brain Sci. **21**, 271–272.

Kewley-Port, D. (**1982**). ''Measurement of formant transitions in naturally produced stop consonant–vowel syllables,'' J. Acoust. Soc. Am. **72**, 379–389.

Kewley-Port, D. (**1983**). ''Time-varying features as correlates of place of articulation in stop consonants,'' J. Acoust. Soc. Am. **73**, 322–335.

Klecka, W. (**1980**). *Discriminant Analysis*, Sage University Paper Series, Quantitative Applications in the Social Sciences, Series no. 07-019 (Sage, Beverly Hills, CA and London).

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich College Publishers, New York).

Lahiri, A., Gewirth, L., and Blumstein, S. E. (**1984**). ''A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study,'' J. Acoust. Soc. Am. **76**, 391–404.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.

Mann, V. A., and Repp, B. H. (**1980**). ''Influence of vocalic context on perception of the [S]–[s] distinction,'' Percept. Psychophys. **28**, 213–228.

McFarland, D. H., Baum, S. R., and Chabot, C. (**1996**). ''Speech compensation to structural modifications of the oral cavity,'' J. Acoust. Soc. Am. **100**, 1093–1104.

Newell, K. M., and Hancock, P. A. (**1984**). ''Forgotten moments: A note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions,'' J. Motor Behav. **16**, 320–335.

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.

Pike, K. (**1943**). *Phonetics* (University of Michigan Press, Ann Arbor).

Pisoni, D., and Luce, P. (**1986**). ''Speech perception: Research, theory, and the principal issues,'' in *Pattern Recognition by Humans and Machines*, edited by E. Schwab and H. Nusbaum (Academic, New York), pp. 1–50.

Recasens, D. (**1985**). ''Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences,'' Lang. Speech **28**, 97–114.

Schatz, C. D. (**1954**). ''The role of context in the perception of stops,'' Language **30**, 47–56.

Shadle, C. H. (**1990**). ''Articulatory-acoustic relationships in fricative consonants,'' in *Speech Production and Speech Modeling*, edited by W. Hardrastle and A. Marchal (Kluwer, Dordrecht, 1990), pp. 187–209.

Shadle, C. H., and Mair, S. J. (**1996**). ''Quantifying spectral characteristics of fricatives,'' Proc. Int. Conf. Spoken Lang. Proc. (ICSLP), pp. 1521–1524 (unpublished).

Soli, S. D. (**1981**). ''Second formants in fricatives: Acoustic consequences of fricative–vowel coarticulation,'' J. Acoust. Soc. Am. **70**, 976–984.

Stevens, K. N. (**1971**). ''Airflow and turbulence for noise for fricative and stop consonants: Static considerations,'' J. Acoust. Soc. Am. **50**, 1182–1192.

Stevens, K. N. (**1985**). ''Evidence for the role of acoustic boundaries in the perception of speech sounds,'' in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, New York), pp. 243–255.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Stevens, K. N., and Blumstein, S. E. (**1981**). ''The search for invariant acoustic correlates of phonetic features,'' in *Perspectives of the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ).

Strevens, P. (**1960**). ''Spectra of fricative noise in human speech,'' Lang. Speech **3**, 32–49.

Sussman, H. M. (**1994**). ''The phonological reality of locus equations across manner class distinctions: Preliminary observations,'' Phonetica **51**, 119–131.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). ''An investigation of locus equations as a source of relational invariance for stop place categorization,'' J. Acoust. Soc. Am. **90**, 1309–1325.

Sussman, H. M., and Shore, J. (**1996**). ''Locus equations as phonetic descriptors of consonantal place of articulation,'' Percept. Psychophys. **58**, 936–946.

Tabain, M. (**1998**). ''Nonsibilant fricatives in English: Spectral information above 10 kHz,'' Phonetica **55**, 107–130.

Tjaden, K., and Turner, G. S. (**1997**). ''Spectral properties of fricatives in Amyotrophic Lateral Sclerosis,'' J. Speech Lang. Hear. Res. **40**, 1358–1372.

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

Waldstein, R. S., and Baum, S. R. (**1991**). ''Anticipatory coarticulation in the speech of profoundly hearing-impaired and normally hearing children,'' J. Speech Lang. Hear. Res. **34**, 1276–1285.

Whalen, D. (**1981**). ''Effects of vocalic formant transitions and vowel quality on the English /s–S/ boundary,'' J. Acoust. Soc. Am. **69**, 275–282.

Wilde, L. (**1993**). ''Inferring articulatory movements from acoustic properties at fricative vowel boundaries,'' J. Acoust. Soc. Am. **94**, 1881.

Yeni-Komshian, B., and Soli, S. (**1981**). ''Recognition of vowels from information in fricatives: Perceptual evidence of fricative–vowel coarticulation,'' J. Acoust. Soc. Am. **70**, 966–975.

Yeou, M. (**1997**). ''Locus equations and the degree of coarticulation of Arabic consonants,'' Phonetica **54**, 187–202.