Acoustic Characteristics of English Fricatives: I. Static Cues

Allard Jongman, Ratree Wayland, and Serena Wong

To date, no single metric has been able to classify place of articulation for all English fricatives with a high degree of accuracy. This paper constitutes a first report on a large-scale study of acoustic cues for classification of place of articulation in fricatives and focuses on static cues: spectral peak location, noise duration, and noise amplitude. While all three static properties of fricatives investigated in this study served to distinguish sibilant from non-sibilant fricatives, the present results indicate that a measure of noise amplitude serves to distinguish all four places of fricative articulation. This finding suggests that static acoustic properties can provide robust information about all four places of articulation, despite variation in speaker, vowel context, and voicing.

1 Introduction

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. Fricatives are produced with a very narrow constriction in the oral cavity. A rapid flow of air through the constriction creates turbulence in the flow, and the random velocity fluctuations in the flow can act as a source of sound. Air turbulence produced in this way, by various kinds of constrictions in the vocal tract (the position of which depends on the particular fricative), is the typical sound source for all fricatives (e.g., Stevens 1971; Shadle 1985).

Most research on fricatives has investigated fricatives in syllable-initial position in carefully produced speech, and that is also the focus of the current project, although syllable-medial fricatives will be investigated at a later stage. 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, acoustic analyses of fricatives suggest that properties of the spectrum, amplitude, and duration of the frication noise all play a role in the distinction between the sibilant (/s, z, ʃ, ʒ/) and non-sibilant (f, v, θ, ð/) fricatives. Additionally, /s, z/ may be distinguished from /ʃ, ʒ/ on the basis of noise spectrum, while there are some (mixed) data suggesting that /f, v/ may be distinguished from /θ, ð/ on the basis of transition information.

The present paper is a first report on the acoustic properties of fricatives and focuses on static cues, defined here as acoustic information which is measured at one location of the

fricative. A second report (Jongman, Sereno, Wayland, and Wong 1998; see also Jongman and Sereno 1995) will detail our investigations of dynamic characteristics, defined here as acoustic information measured at more than one location in the fricative. These dynamic characteristics include relative amplitude (e.g., Hedrick and Ohde 1993), locus equations (e.g., Sussman, McCaffrey, and Matthews 1991; Sussman 1994; Fowler 1994), and spectral moments (Forrest, Weismer, Milenkovic, and Dougall 1988). The static cues considered in this paper are spectral peak location, noise duration, and noise amplitude. The remainder of this section summarizes previous research on these cues, followed by Methods in section 2, Results in section 3, and Discussion in section 4.

1.1   Spectral properties of frication noise

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. 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; Behrens and Blumstein 1988a). In particular, /ʃ, ʒ/ typically exhibit a mid-frequency 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.

These studies reveal that the local spectral properties of frication noise serve to distinguish the sibilant fricatives /s, z, ʃ, ʒ/ as a group from non-sibilant /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 1985; Behrens and Blumstein 1988a; Klatt ms.) 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).

1.2 Noise duration

Similar to amplitude, noise duration serves to distinguish sibilant from non-sibilant fricatives, with /s, ʃ, ʒ/ being longer than /f, θ/ (e.g., Behrens and Blumstein 1988a; Klatt ms.). As for distinctions within each of these two groups, the literature reports conflicting results: Behrens and Blumstein (1988a) found no difference in duration between /s/ and /ʃ/ and a trend for /θ/ to be shorter than /f/. Klatt, however, found that /f/ is shorter than /θ, s, ʃ/ and that /θ, s, ʃ/ have similar durations.

Noise duration provides 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).

1.3 Noise amplitude

Most studies which have investigated amplitude of the frication noise have focused on voiceless fricatives and converge on similar findings: sibilant /s, ʃ/ have a substantially greater (10-15 dB) amplitude than non-sibilant /f, θ/, and within each group, the two fricatives are not very different from each other (e.g., Strevens 1960; Behrens and Blumstein 1988a).

In sum, acoustic studies focusing on properties of the frication noise show that properties of the spectrum, amplitude, and duration of the noise can all serve to distinguish the sibilant / s, z, ʃ, ʒ/ from the non-sibilant /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 seem adequate to distinguish /f/ from /θ/.

2 Methods

2.1 Participants

Twenty speakers (10 females and 10 males) were recruited from the Cornell undergraduate and graduate 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.

2.2 Stimulus materials

The eight English fricatives /f, v, θ, ð, s, z, ʃ, ʒ/ were recorded in CVC syllables in the carrier phrase "Say ____ again". The fricatives were in initial position, followed by each of six vowels /i, e, æ, a, 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 x 6 vowels x 3 repetitions).

2.3 Recordings

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-degree angle and 15 cm away from the corner of the speaker's mouth, to prevent turbulence from direct airflow from impinging on the microphone. Once recording levels had been set, a calibration tone was recorded to facilitate constant access to the original Sound Pressure Level of the speaker's speech.

2.4 Analysis and further processing

All recordings were sampled at 22 kHz with 16 bit quantization on a Sun SPARCstation 5. All measurements were made using Entropics Systems' Waves+/ESPS software. Fricative segmentation involved the simultaneous consultation of waveform and wide-band 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).

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 has shown 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 match those of the FFT spectra. Spectral peak is defined here as the highest-amplitude peak of the FFT spectrum.

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 minus vowel 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 (cf. Behrens and Blumstein 1988b).

3 Results

3.1 Spectral peak location

Table 1 shows mean spectral peak location as a function of place of articulation. A four-way ANOVA (Place x Voicing x Vowel x Gender) revealed a main effect for Place of articulation [$F(3, 2498) = 327.69$, p <.0001]. Bonferroni post-hoc tests indicated that spectral peak location of /ʃ, ʒ/ was significantly different from that of the other three places of articulation. In addition, spectral peak location of /s, z/ was significantly different from that of /f, v/ and /θ, ð/. Finally, there was no significant difference in peak location between /f, v/ and /θ, ð/ (p>.208).

Place of articulation	Spectral peak location (in Hz)
labiodental /f, v/	7678
interdental /θ, ð//	7503
alveolar /s, z/	6882
palato-alveolar /ʃ, ʒ/	3712

Table 1. Mean spectral peak location (averaged across vowels, voiced and voiceless fricatives, and male and female speakers) as a function of place of articulation.

A main effect of Gender [$F(1, 2498) = 55.06$, p <.0001] indicated that, as expected, mean spectral peak location was significantly higher for female (6809 Hz) than for male (6066 Hz) speakers. There was no main effect for Voicing or Vowel.

3.2 Noise duration

Table 2 shows mean frication duration as a function of place of articulation. A four-way ANOVA (Place x Voicing x Vowel x Gender) revealed a main effect for Place of articulation [F(3, 3094) = 327.69, p <.0001]. Bonferroni post-hoc tests indicated that noise duration of the non-sibilant fricatives was significantly shorter than that of the sibilant fricatives.

Place of articulation	Frication duration (in ms)
labiodental /f, v/	123
interdental /θ, ð//	126
alveolar /s, z/	148
palato-alveolar /ʃ, ʒ/	150

Table 2. Mean frication duration (averaged across vowels, voiced and voiceless fricatives, and male and female speakers) as a function of place of articulation.

A main effect of Voicing [F(1, 3094) = 2478.0, p <.0001] indicated that voiceless fricatives (171 ms) were significantly longer than their voiced counterparts (102 ms). A main effect of Gender [F(1, 3094) = 8.33, p <.004] indicated that fricatives produced by female speakers (139 ms) were slightly longer than those produced by male speakers (135 ms). Finally, a main effect of Vowel [F(5, 3094) = 23.63, p <.0001] obtained. Bonferroni post-hoc tests indicated the following hierarchy: noise duration preceding /u/ was 149 ms, /i/: 143 ms, /o/: 138 ms, /e/: 136 ms, /ɑ/: 129 ms, and /æ/: 126 ms. All differences were significant except that between /ɑ/ and /æ/.

3.3 Noise amplitude

Table 3 shows mean noise amplitude, vowel amplitude, and the difference between the two as a function of place of articulation. A four-way ANOVA (Place x Voicing x Vowel x Gender) with Difference as the dependent variable revealed a main effect for Place of articulation [F(3, 3094) = 1489.51, p <.0001]. Bonferroni post-hoc tests indicated that all four places of articulation were significantly different from each other in terms of amplitude difference.

Place of articulation	Noise amplitude	Vowel amplitude	Difference
Labiodental /f, v/	59.4	76.4	-17
Interdental /θ, ð/	58.7	76.7	-18
Alveolar /s, z/	66.3	76.3	-10
Palato-alveolar /ʃ, ʒ/	67.3	76.4	-9.1

Table 3. Mean noise and vowel amplitude in dB (averaged across vowels, voiced and voiceless fricatives, and male and female speakers) as a function of place of articulation. Difference refers to noise amplitude minus vowel amplitude in dB.

A main effect of Voicing [$F(1, 3094) = 1644.06$, $p < .0001$] 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 [$F(5, 3094) = 11.94$, $p < .0001$] obtained. Amplitude difference preceding /o/ was -14 dB, /u/: -13.8 dB, /e/: -13.8 dB, /a/: -13.6 dB, /æ/: -13 dB, /i/: -12.7 dB. Bonferroni post-hoc tests indicated that only the amplitude difference for /i/ and /æ/ differed from that for all other vowels. There was no main effect of Gender.

4   Discussion

The present results indicate that there are several acoustic dimensions along which fricative places of articulation can be distinguished. Spectral peak location distinguishes sibilant /s, z/ and /ʃ, ʒ/ from non-sibilant /f, v/ and /θ, ð/. On average, sibilants had a maximum-amplitude peak at a lower frequency than non-sibilants. In addition, peak location distinguishes alveolar /s, z/ from palato-alveolar /ʃ, ʒ/, the latter having a spectral maximum at a much lower frequency. This finding is consonant 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 1985; Behrens and Blumstein 1988a; Klatt ms.). These studies all support the distinction between sibilants and non-sibilants and between /s, z/ and /ʃ, ʒ/ even though all differed from each other and from the present one in terms of their criteria for picking spectral peaks.

Research on the perceptual role of spectral properties of the frication noise points to the same conclusion. Heinz and Stevens (1961) synthesized fricatives by systematically varying the location of the spectral peak from 2-8 kHz. Listeners' responses showed that fricatives with a peak below 3 kHz were identified as /ʃ/; fricatives with a peak in between

approximately 4.5 and 6.5 kHz were identified as /s/, and those with a peak above 6.5 kHz were perceived as /f/ or /θ/. Listeners were unable to distinguish /f/ from /θ/.

Noise duration also served to separate the sibilants from the non-sibilants. Sibilant fricatives were longer than non-sibilants, supporting similar findings by Behrens and Blumstein (1988a). 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).

Perceptually, no studies seem to have directly manipulated the duration of a spectrally ambiguous frication noise to determine the role of duration as a cue to place of articulation. However, given that, in natural fricatives, duration and spectrum cannot be separated, noise duration is unlikely to be a primary cue to place of articulation. Jongman (1989) showed that identification of place of articulation is already quite accurate based on the first 20 ms of natural fricatives. Noise duration does not seem to be an important cue to fricative voicing either (e.g., Jongman 1989, for syllable-initial position; Stevens, Blumstein, Glicksman, Burton, and Kurowski 1992, for intervocalic fricatives).

Finally, all four places of articulation could be distinguished in terms of noise amplitude. Sibilant fricatives had a greater noise amplitude than non-sibilants; moreover, within the group of sibilants, palato-alveolar /ʃ, ʒ/ had a greater noise amplitude than alveolar /s, z/, while for the non-sibilants labiodental /f, v/ had a greater amplitude than interdental /θ, ð/. Previous research supports the role of noise amplitude in the sibilant/non-sibilant distinction (e.g., Strevens 1960; Behrens and Blumstein 1988a, b). In particular, in their study of /f, θ/ and /s, ʃ/ Behrens and Blumstein (1988a, b) reported amplitude differences of similar magnitude as the present study. However, contrary to these studies, the present study also indicates that noise amplitude can distinguish place of articulation *within* these two groups.

Most research on the role of noise amplitude as a perceptual cue to fricative place of articulation has involved cross-splicing of both natural and synthetic speech. In such experiments (e.g., McCasland 1979), the frication noise of either /s, ʃ/ or /f, θ/ was cross-spliced with transitions corresponding to /f, θ/ or /s, ʃ/, respectively. Overall noise amplitude was then systematically manipulated ranging from low values typical of /f, θ/ to high values typical of /s, ʃ/. Results of such cross-splicing experiments typically show that noise intensity is a primary characteristic distinguishing sibilant /s, ʃ/ from non-sibilant /f, θ/.

However, as Behrens and Blumstein (1988b) point out, it is impossible to draw any firm conclusions from such experiments since they involve spectrally incompatible

information between frication noise and vocalic transitions: if the noise amplitude of /s/ is reduced and appended to vocalic transitions appropriate for /f/ and listeners report hearing /f/, it is impossible to establish whether they based their judgment on the frication noise or on the vocalic transitions (or both). In fact, these researchers showed that when noise amplitude is manipulated for stimuli in which the spectral properties of the noise and the transitions *are* compatible, noise amplitude has relatively little effect on fricative perception. In sum, more research is needed to evaluate the importance of noise amplitude as a robust cue in fricative perception.

While all three static properties of fricatives investigated in this study (spectral peak, duration, and amplitude) served to distinguish sibilant from non-sibilant fricatives, the present results indicate that a measure of noise amplitude serves to distinguish all four places of fricative articulation. This finding suggests that, contrary to earlier reports, static acoustic properties can provide robust information about all four places of articulation, despite variation in speaker, vowel context, and voicing.

We are currently in the process of determining the extent to which both the static cues reported here and the dynamic cues reported in Jongman et al. 1998 contribute to fricative perception. We have collected data on listeners' identification of English fricatives based on fricative-vowel syllables and on isolated fricative noise portions. Regression of these perceptual data on our acoustic measurements should allow for a ranking of acoustic properties as cues to fricative identification.

5 Acknowledgements

6 References

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.

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.

Crystal, T. and House, A. (1988) Segmental durations in connected-speech signals: Current results. *J. Acoust. Soc. Am.* 83, 1553-1573.

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. *Perception & Psychophysics* 55, 597-611.

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

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.

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. *Language and 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., and Sereno, J.A. (1995) Acoustic properties of non-sibilant fricatives. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 432-435.

Jongman, A., Sereno, J.A., Wayland, R., and Wong, S. (1998) Acoustic properties of English fricatives. In P. Kuhl and L. Crum (eds.), *Joint Proceedings of the International Congress on Acoustics and the Acoustical Society of America*, 2935-2936.

Klatt, D.H. (ms). Voiceless fricatives. Unpublished manuscript.

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

McCasland, G.P. (1979) Noise intensity and spectrum cues for spoken fricatives. *J. Acoust. Soc. Am.* Suppl 1 65, S78-79.

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

Shadle, C. (1985) The acoustics of fricative consonants. *MIT RLE Tech. Rep.* 506.

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 noise for fricative and stop consonants: Static considerations. *J. Acoust. Soc. Am.* 50, 1182-1192.

Stevens, K.N., Blumstein, S.E., Glicksman, L., Burton, M., and Kurowski, K. (1992) Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *J. Acoust. Soc. Am.* 91, 2979-3000.

Strevens, P. (1960) Spectra of fricative noise in human speech. *Language and 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.

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.