Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English

Shari R. Baum, and Sheila E. Blumstein

Citation: The Journal of the Acoustical Society of America **82**, 1073 (1987); doi: 10.1121/1.395382
View online: https://doi.org/10.1121/1.395382
View Table of Contents: https://asa.scitation.org/toc/jas/82/3
Published by the Acoustical Society of America

LETTERS TO THE EDITOR

Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English

Shari R. Baum and Sheila E. Blumstein
Brown University, Department of Cognitive and Linguistic Sciences, Box 1978, Providence, Rhode Island 02912

(Received 13 April 1987; accepted for publication 3 June 1987)

Acoustic analyses were undertaken to explore the durational characteristics of the fricatives [f,θ,s,v,ð,z] as cues to initial consonant voicing in English. Based on reports on the perception of voiced–voiceless fricatives, it was expected that there would be clear-cut duration differences distinguishing voiced and voiceless fricatives. Preliminary results for three speakers indicate that, although differences emerged in the overall mean duration of voiced and voiceless fricatives, contrary to expectations, there was a great deal of overlap in the duration distribution of voiced and voiceless fricative tokens. Further research is needed to examine the role of duration as a cue to syllable-initial fricative consonant voicing in English.

PACS numbers: 43.70.Fq, 43.71.Es

INTRODUCTION

A primary aim of research in speech production is the characterization of the acoustic properties corresponding to certain speech sounds or features of speech segments. A great deal of research has focused on the durational properties of speech sounds. Duration has been shown to be an important cue for the voicing contrast in English. In particular, voice-onset time (VOT) is a parameter distinguishing voiced and voiceless stops in English. Studies of the production of voiced and voiceless stops in initial position in citation form have shown two separate distributions of VOT with minimal, if any, overlap between voiced and voiceless productions. Voiced stops are characterized by prevoicing or short-lag VOTs, and voiceless stops are characterized by long-lag VOTs (Lisker and Abramson, 1964). Similarly, measurements of vowel duration as a cue to postvocalic voiced and voiceless stop consonants have shown two distributions of vowel durations with voiced stops preceded by longer vowel durations than voiceless stops. As with the VOT functions, there is minimal overlap in the distribution of vowel durations in utterances spoken in citation form (Duffy and Gawle, 1984; Raphael, 1972).

Similarly, it has been claimed that the duration of frication noise serves as a cue to the voiced–voiceless contrast in word-initial fricative consonants (Klatt, 1976; Nartey, 1982). To date, only one study has directly examined the duration of the frication noise in the production of fricative consonants. Manrique and Massone (1981) examined the duration of the frication noise in Spanish voiced and voiceless fricatives. These authors found a significant difference between the mean duration of voiced and voiceless fricatives and "little or no overlap between duration ranges" (p. 1147) for the two voicing categories. They conclude that one of the primary characteristics distinguishing voiced and voiceless fricatives in Spanish is the duration of the frication noise. It should be pointed out, however, that of the four voiced and voiceless fricatives studied, only [š ž] and [x ɣ] were homorganic. The remaining fricative pairs [f β] and [s ð] did not share place of articulation. As a result, it is not clear whether the obtained duration measures reflected differences solely in the voicing characteristics of the stimuli or, alternatively, in the overall amplitude of the stimuli. The second alternative is suggested since stimuli with weaker amplitude, such as [θ ð] have been shown to have intrinsically shorter durations than stimuli with greater amplitude, such as [s z], irrespective of the voicing characteristics of the stimuli (Behrens and Blumstein, 1987).

By far, the great majority of studies exploring the role of duration as a cue to voiced and voiceless fricative consonants have concentrated on perceptual analyses (Cole and Cooper, 1975; Massaro and Cohen, 1976). Cole and Cooper (1975) conducted a study on the perception of voicing in English affricates and fricatives. The experimenters cut six successive segments from the end of the frication noise in naturally produced [sa], [fa], and [ča] syllables via a tape-splicing technique. These researchers found that, in a forced-choice paradigm (e.g., [sa] or [za]), subjects consistently labeled the shortest [sa] and [fa] tokens (consisting of 40 and 80 ms of frication noise for [sa] stimuli, and 35 and 69 ms of frication for [fa] stimuli) as [za] and [va], respectively. All other tokens with longer frication durations yielded voiceless percepts. Cole and Cooper conclude that the duration of the frication noise serves as a primary cue to the voiced–voiceless distinction in English. Earlier studies (e.g., Denes, 1955; Raphael, 1972; Soli, 1982) had focused on voicing in syllable-final fricatives, concluding that preceding vowel duration, and to a lesser extent, frication duration, served to signal the voicing contrast in final position.

However, results of a recent study suggest that listeners may not use duration as a primary cue distinguishing voiced and voiceless fricatives in English. Jongman (1987) examined the duration of frication noise as a cue to place and manner of articulation. He explored perception of the voicing contrast as well, for the fricatives [f,θ,s,v,ð,z]. Results showed that overall mean performance was at an accuracy level of 83%, even when subjects were presented with only the first 20 ms of frication noise. Thus, despite the fact that duration differences in the frication noise were nullified, subjects were able to accurately perceive voiced and voiceless fricatives. Moreover, perception errors (voiceless for voiced) occurred more often for voiced tokens, suggesting that, contrary to Cole and Cooper's (1975) results, shortening voiceless fricatives to as little as 20 ms does not effect a change in the percept of voicelessness. In addition, in measuring the original voiced and voiceless fricative stimuli, Jongman (1987) found that voiceless fricatives were not always longer than their voiced counterparts.

On the basis of both the perception and production studies described above, it is not certain that duration is a primary cue to voicing in fricative consonants. As a first step in addressing this question, it would be useful to determine whether consistent duration differences emerge in the production of voiced and voiceless fricatives. If they do not, then other properties must serve as the primary, or at least additional, cues to this distinction. The present study represents a preliminary step in that direction. Its focus is on the durational characteristics of syllable-initial English fricative homorganic pairs ([f–v, θ–ð, s–z]) produced in citation form in various vocalic contexts.

I. METHODS

A. Subjects

Three adult male speakers of American English served as subjects for this study. All were phonetically trained graduate students in linguistics.

B. Stimuli

Stimuli consisted of the six fricative consonants [f,θ,s,v,ð,z] in the environment preceding the five vowels [i,e,a,o,u]. Each syllable was printed in orthographic form on 3 × 5 cards and presented to the subjects five times each in random order. Each subject read the CV syllables in citation form and was instructed to produce the stimuli clearly and carefully, but as naturally as possible. In all, each subject produced a total of 150 fricative-vowel syllables.

C. Procedure

Stimuli were recorded onto magnetic tape in a sound-treated room using a Nagra 4.2 recorder and Shure SM81 microphone. The fricative-vowel syllables were then digitized onto disk via a PDP 11/34 computer at a sampling rate of 20 kHz with a 9.0-kHz low-pass filter and 10-bit quantization.

D. Durational analyses

For all of the test stimuli, durations of the frication noise segments were measured from the waveform display. For the voiceless fricatives, a cursor was placed at the start of frication noise, as defined by a greater than 10-dB increase in spectral energy (measured by discrete Fourier transform) at or above 2 kHz as compared to the background noise level. The offset of the fricative was also demarcated by a cursor placed at the end of the noise segment which usually coincided with the onset of periodicity corresponding to the vowel transition. The onset of voiced fricatives was readily determined by visual inspection of the waveform; its offset was defined as a decrease or loss of high-frequency noise in the region between 4 and 6.5 kHz (determined by Fourier analysis).

II. RESULTS

Table I shows the mean durations and ranges of the voiced and voiceless fricatives. As the results show, the overall mean value of the voiceless fricatives was longer than that of the voiced fricatives for all fricative pairs. Mean duration of [f] frication noise segments was 149 ms, and mean duration of frication noise for [v] was 116 ms. Mean [θ] duration was 134 ms as compared to a mean of 107 ms for [ð] tokens. Mean durations for [s] and [z] frication noise were 174 and 152 ms, respectively.

Looking at the duration differences as a function of fricative and vowel across the subjects, it is clear that for most fricative pairs, the voiceless fricatives are longer than the voiced fricatives. Exceptions to this pattern do occur for individual subjects, particularly for [θ–ð] in the environment of [i] (subject BS), [a] (subject BS), and [u] (subjects BS and WB), and to a lesser extent for [f]. In the case of [f], the frication noise produced by subject BS was longer for [v] than for [f] in the environment of [e] and [a]. Nevertheless, on balance, most subjects produced fricatives with longer frication noise durations for voiceless fricatives compared to voiced fricatives.

However, the duration ranges suggest that, while there may be a difference in the means for voiced and voiceless fricatives, there is considerable overlap in the distribution of responses between the two categories. For example, note that, for [fi] and [vi] produced by WB, the duration ranges are 131–179 ms for [f] and 63–185 ms for [v], and, for [si] and [zi], they are 185–207 ms and 165– 197 ms, respectively. Reviewing these ranges across vowels and speakers suggests that the distribution of voiced and voiceless tokens may not be consistently bimodal.

To address this issue directly, the distribution of duration of frication noise was examined for each subject. Figures 1–3 show the results for the fricative pairs [f–v], [s–z], and [θ–ð] for each subject. As can be seen, while, in general, voiceless fricatives show longer durations than voiced fricatives, there was a great deal of overlap between the voicing categories. No clear boundary emerges between the voiced and voiceless fricative durations for any consonant pair for any subject. These graphs clearly suggest that, although a mean difference may result between the durations of frication noise of voiced and voiceless fricatives, the duration distinction is not consistently produced by these subjects. It

TABLE I. Mean duration (in ms) of the frication noise for voiced and voiceless segments. Duration ranges are indicated below each mean.

	S	[f]	[v]	[θ]	[ð]	[s]	[z]
[i]	BS	183	109	106	120	196	175
		146–226	62–149	60–146	119–121	185–207	165–197
	SD	143	92	153	124	173	175
		150–202	44–147	108–209	82–157	159–197	166–191
	WB	150	138	138	93	178	167
		131–179	63–185	90–204	46–149	169–189	135–188
	$\bar{X}$	159	113	138	112	182	172
[e]	BS	141	147	130	90	169	156
		123–159	121–160	76–158	56–121	153–191	149–164
	SD	131	79	159	72	163	133
		84–155	44–115	112–221	45–96	156–173	120–139
	WB	132	96	181	68	164	158
		69–172	52–133	109–292	39–148	146–198	139–168
	$\bar{X}$	135	107	157	77	165	149
[a]	BS	128	145	110	145	191	143
		60–172	114–186	54–155	112–161	165–205	122–164
	SD	121	94	113	89	160	117
		71–172	55–124	110–117	41–137	135–179	99–132
	WB	161	116	96	58	175	157
		130–179	39–170	67–134	24–135	154–196	131–185
	$\bar{X}$	137	118	106	97	175	139
[o]	BS	166	148	149	132	171	148
		153–177	114–168	143–153	99–152	159–189	144–152
	SD	149	79	147	99	163	156
		139–167	29–119	101–176	36–136	150–184	137–175
	WB	146	122	124	86	164	112
		104–167	49–149	110–144	36–129	142–186	66–144
	$\bar{X}$	153	116	140	106	166	139
[u]	BS	172	130	134	152	199	154
		158–195	118–146	107–165	128–174	186–216	125–197
	SD	167	126	146	114	178	159
		140–183	83–165	92–169	87–137	157–194	122–192
	WB	143	120	114	143	167	164
		129–159	64–140	99–184	42–186	171–197	151–187
	$\bar{X}$	162	125	131	136	181	159
$\bar{X}$	BS	158	136	126	130	185	156
		60–226	62–186	54–165	56–174	153–216	122–197
	SD	142	95	144	100	167	148
		71–202	29–165	92–221	36–157	135–197	99–192
	WB	147	118	134	90	170	151
		69–179	39–185	67–292	24–186	142–198	66–188
	$\bar{X}$	149	116	134	107	174	152

should be noted that the tokens were perceived (by the experimenter) as good exemplars of their voicing categories.

III. DISCUSSION

Overall, as expected from the literature, voiceless fricatives were found to be longer than their voiced counterparts. Nevertheless, distinctions between voiced and voiceless fricatives did not emerge across all vowel environments or across the three speakers. More importantly, in examining the duration distribution of the frication noise for voiced and voiceless fricatives, while there is a tendency toward a bimodal distribution, there is considerable overlap between the two duration distributions for all fricatives and for all speakers. In effect, this suggests that speakers are not consistently producing a duration distinction as a cue to syllable-initial fricative voicing. It is important to note that these results obtain in citation form. Presumably, were fricatives produced in more natural contexts, an even greater degree of category overlap would result due, in part, to changes in rate of speech and to precision of articulation. Often in running speech, both vowels and consonants are reduced in overall duration and even clear dichotomies such as those found for VOT become diminished (Lisker and Abramson, 1967). The results of these preliminary observations suggest that speakers do not maintain a clear dichotomy in the duration of frication noise as a cue to voicing. While the perceptual literature may suggest that listeners are sensitive to such attributes, the fact that this distinction is not consistently made by speakers suggests that duration of frication noise may not be a primary attribute distinguishing voiced and voiceless fricative consonants.

More research needs to be done exploring durational contrasts in fricative consonants. Measurements should be taken in multiple word positions and in running speech, as produced by a number of speakers, in order to determine the extent to which duration is used by speakers as a cue to fricative voicing. An important finding of the present preliminary

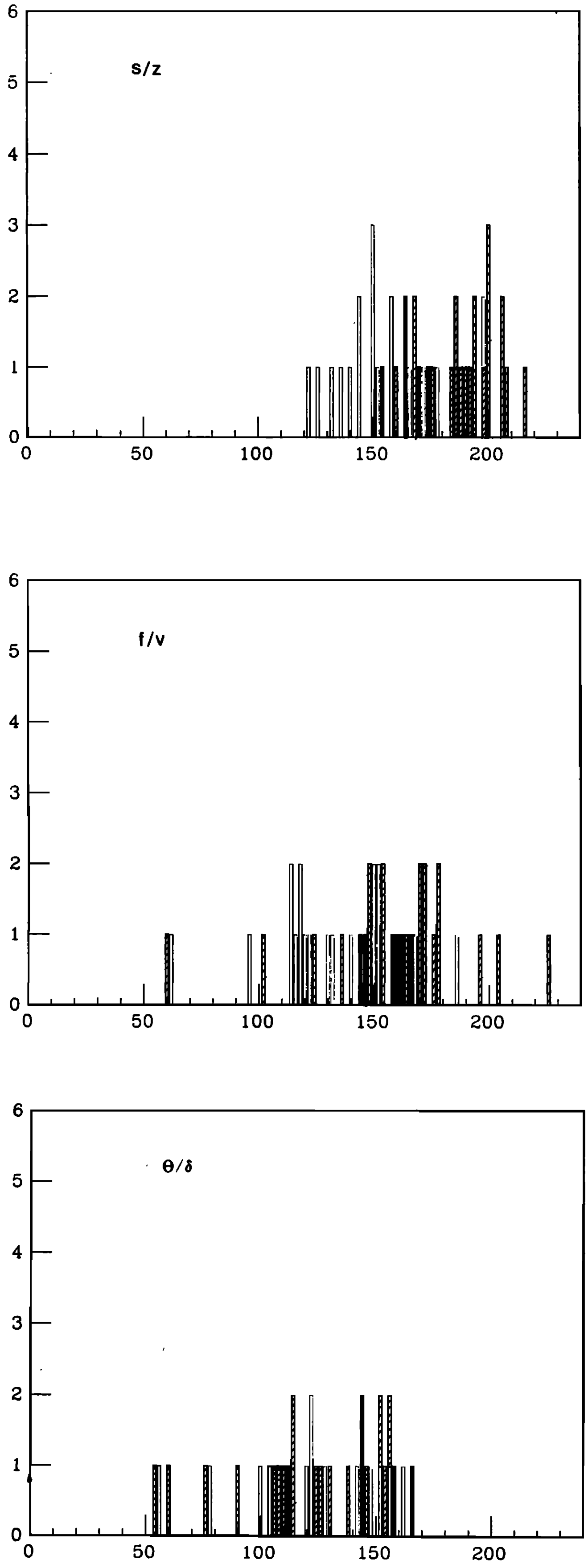

FIG. 1. Distribution (# tokens) of frication noise durations (in ms) for subject BS. Voiced tokens are indicated by white bars, voiceless by striped; solid black indicates regions of complete overlap between voiced and voiceless tokens.

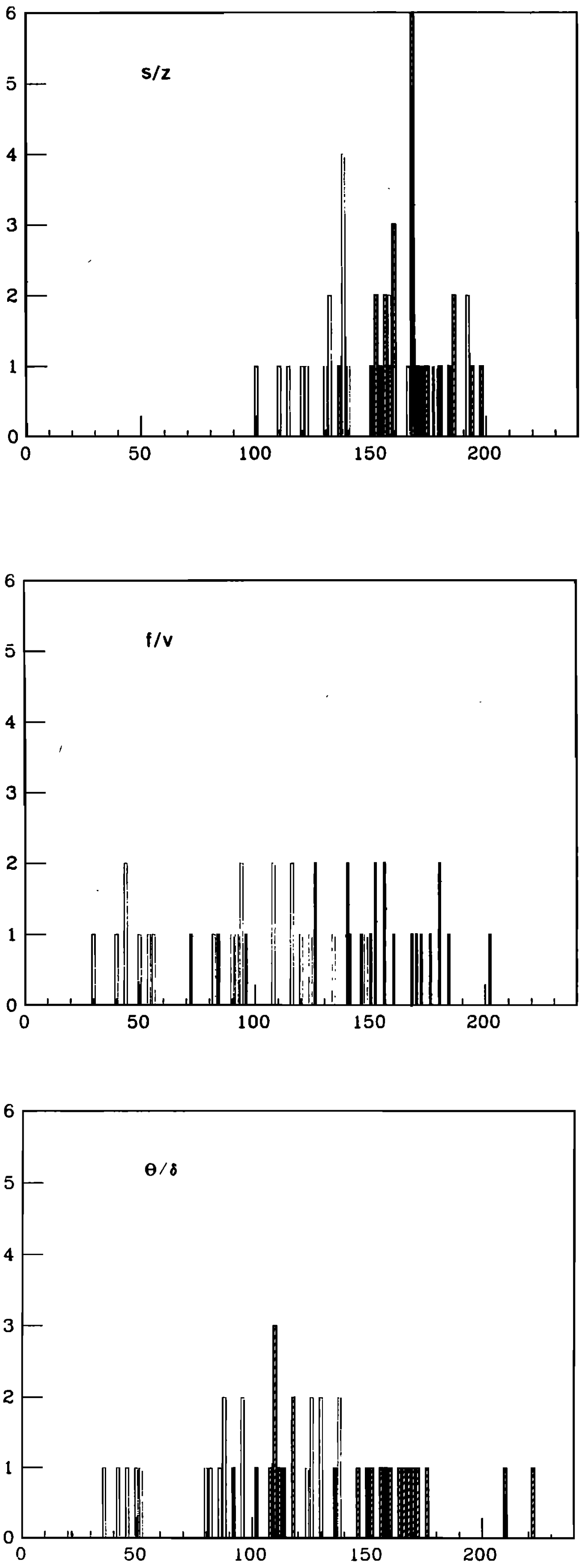

FIG. 2. Distribution (# tokens) of frication noise durations (in ms) for subject SD. Voiced tokens are indicated by white bars, voiceless by striped; solid black indicates regions of complete overlap between voiced and voiceless tokens.

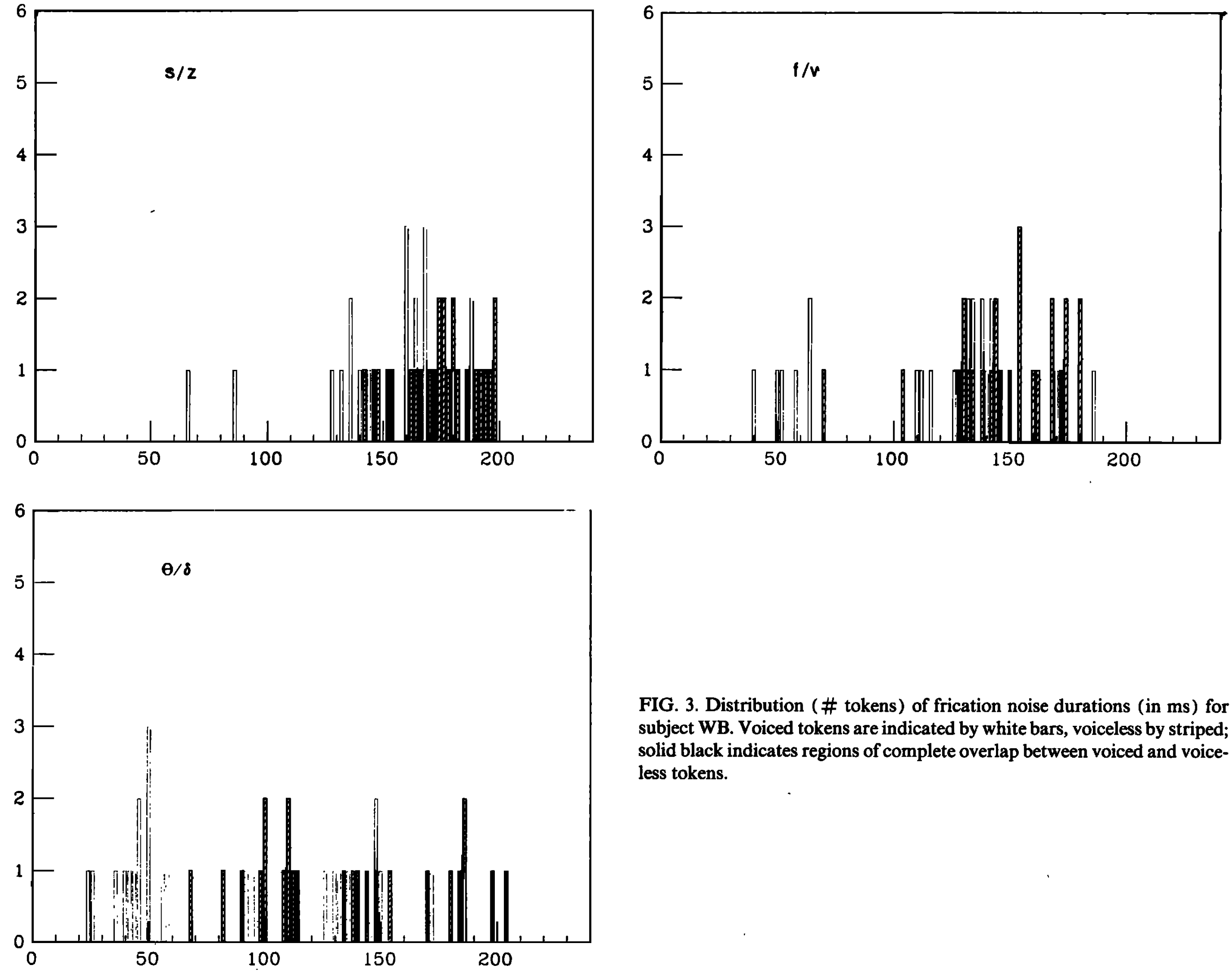

FIG. 3. Distribution (# tokens) of frication noise durations (in ms) for subject WB. Voiced tokens are indicated by white bars, voiceless by striped; solid black indicates regions of complete overlap between voiced and voiceless tokens.

work is that, in contrast to what has been generally assumed in the literature, a voicing distinction based on frication noise duration may not be consistently produced by normal speakers of English.

ACKNOWLEDGMENTS

Many thanks to Susan Behrens, Allard Jongman, and Kathleen Kurowski for making available the recordings used in this study. This research was supported, in part, by Grants NS 22282 and NS 15123 to Brown University.

Behrens, S., and Blumstein, S. (**1987**). "Acoustic characteristics of English voiceless fricatives: a descriptive analysis," submitted.

Cole, R., and Cooper, W. (**1975**). "The perception of voicing in English affricates and fricatives," J. Acoust. Soc. Am. **58**, 1280–1287.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Duffy, J., and Gawle, C. (**1984**). "Apraxic speakers' vowel duration in consonant–vowel–consonant syllables," in *Apraxia of Speech*, edited by J. Rosenbek, M. McNeil, and A. Aronson (College-Hill, San Diego, CA).

Jongman, A. (**1987**). "Minimum amount of frication noise required for identification of English fricatives," submitted to J. Acoust. Soc. Am.

Klatt, D. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.

Lisker, L., and Abramson, A. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lisker, L., and Abramson, A. (**1967**). "Some effects of context on voice onset time in English stops," Lang. Speech **10**, 1–28.

Manrique, A. M. B., and Massone, M. I. (**1981**). "Acoustic analysis and perception of Spanish fricative consonants," J. Acoust. Soc. Am. **69**, 1145–1153.

Massaro, D., and Cohen, M. (**1976**). "The contribution of fundamental frequency and voice onset time to the /zi/–/si/ distinction," J. Acoust. Soc. Am. **60**, 704–717.

Nartey, J. (**1982**). "On fricative phones and phonemes: measuring the phonetic differences within and between languages," UCLA working papers in Phonetics, **55**.

Raphael, L. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Soli, S. (**1982**). "Structure and duration of vowels together specify fricative voicing," J. Acoust. Soc. Am. **72**, 366–378.