QUANTIFYING TIME-VARYING SPECTRA OF ENGLISH FRICATIVES

Lorin Wilde
Research Lab of Electronics, Dept. Electrical Engineering & Computer Science
MIT, Cambridge, MA 02139, USA
email: wilde@speech.mit.edu

ABSTRACT

Noise characteristics of fricatives were quantified with respect to adjacent vowel spectra. The weak and strong fricatives were well-separated: the maximum amplitudes above 2 kHz in the fricative, normalized relative to vowel amplitude, were 15-20 dB more for the alveolars and palato-alveolars than for the labiodentals and dentals. In addition, spectral changes during the consonantal interval were calculated.

1. INTRODUCTION

The acoustic consequences during fricative production include continuous spectral variations over time. First, the articulation and aerodynamics in noise generation during a particular fricative are continuous. In addition, the acoustics of fricatives produced in connected speech are influenced by concurrent coarticulatory movements.

Recent studies have provided additional evidence that the kinematics of fricative articulation create an acoustic signal that is inherently non-static [1] [2] [3]. Difficulty in the analysis of fricatives also arises from the nature of random noise generation in fricative production. The nature of a noise source complicates the accurate measurement of spectral properties associated with the articulatory movement.

An automatic analysis system for quantifying fricative noise spectra was developed. The objective was to reduce the dimensionality of the data while measuring essential spectral properties. Spectral changes during the consonant were examined. In addition, the attribute of stridency, signaled by greater energy in the high frequencies in the consonant relative to the vowel, was examined. The following questions motivated the choice of acoustic measures: 1) How much greater energy? and 2) In which frequency regions?

2. METHODS

A database was collected in order to examine in detail the acoustic attributes of fricative consonants in the front, back and back-rounded vowel contexts. Three normal speakers of American English, one male and two female, recorded 'CVCV'CVC nonsense syllables. The consonant was one of the eight English fricatives: /f, v, θ, ð, s, z, š, ž/ and the stressed vowels were /i, ε, ɑ, ʌ, o, u/. The first and third vowels in an utterance were the same. Two repetitions of each fricative in pre-stressed position were analyzed in this study.

The speech was recorded in a sound-treated room using an omnidirectional microphone which was located approximately 25 cm in front of the speaker and 5 cm above the speaker's mouth. The recordings were low-pass filtered at 7.5 kz and digitized at 16 kHz. One additional male speaker, previously low-pass filtered at 4.8 kHz and digitized at 10 kHz by Klatt [4], was also studied. The combined database was used to develop an automatic analysis system for quantifying fricative noise spectra.

Fricative noise characteristics were considered with respect to adjacent vowel spectra, with measures made relative to the consonant-vowel (CV) boundary. Digitized waveforms labeled with acoustic landmarks, i.e. fricative-vowel boundaries, are the inputs to the analysis system. Averaged spectrograms were computed by advancing a 6.4 msec window in 1 msec steps and averaging overlapping windows. A 20 ms averaging interval was empirically chosen: long enough to reduce error due to random fluctuations and short enough to quantify time variations in individual tokens. The maximum power was calculated in five frequency bands: 1) 0-0.5 kHz, 2) 0.5-1 kHz, 3) 1-2 kHz, 4) 2-4 kHz and 5) 4-8 kHz. The amplitudes and frequencies of spectral peaks, occurring relative to landmark times, are the outputs. Further details are provided in Wilde [5].

3. RESULTS

The following results are reported here: 1) variation in the noise over the duration of the consonant and 2) quantification of the feature [strident]. All measures are made relative to the consonant-vowel (CV) boundary. The following results are reported for the voiceless fricatives, in order to restrict our discussion to utterances for which the CV landmark could be accurately identified to within 4 ms.

3.1 Time-varying Noise Spectra

A measure of spectral variation over time was calculated by subtracting the amplitude value at the right edge of the fricative ($CV - 20$ ms) from the amplitude value at the temporal center of the fricative. A negative difference means that the amplitude at the edge is greater than the amplitude at the midpoint. The results for the three highest frequency bands are shown for all four speakers in Figure 1. The main, not unexpected finding is that there is considerable variation in noise spectra over time. That is, the noise amplitude is not constant and, from the interquartile ranges of all subjects, appears to vary from about -13 to +8 dB over the interval from the fricative midpoint to just before the fricative-vowel boundary.

The individual results for each band suggest a trend for differences between the weak and strong fricatives. For Band 3 (1-2 kHz) the clear trend is that there is greater amplitude difference for the labiodental and dental fricatives, grouped here as nonstrident. Band 4 (2-4 kHz) shows the same trend, although the ranges are more similar. In the 1-4 kHz range, the differences for all fricatives are negative, i.e., the edge is stronger than the middle. However, for the highest frequencies in Band 5 (4-8 kHz) there is a contrast in trends between the nonstri-

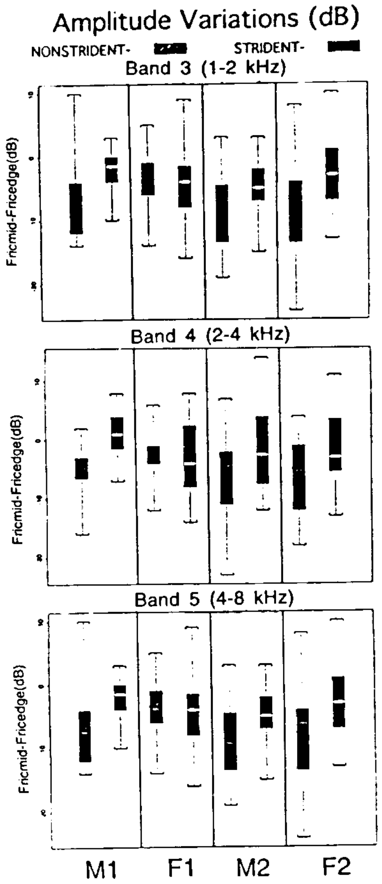

Figure 1: *The medians (lines) and interquartiles ranges (IQR=box height) illustrate the magnitude of amplitude variations (dB) in Bands 3 (top), 4 (middle) and 5 (bottom). Each box represents 24 data points, each calculated by subtracting the amplitude at the right edge ($CV - 20$) from the amplitude in the middle ($CV - duration/2$), with nonstrident vs. strident voiceless fricatives are shown separately for each speaker. The dotted lines extend to the extreme values of the data or a distance 1.5xIQR from the center, whichever is less.*

dents, which are clearly negative, and the stridents, which are clearly positive.

3.2 Quantifying Stridency

The feature [strident] was quantified by subtracting the F1 amplitude in the following vowel from the maximum amplitude peak above 2000 Hz in the consonant. As expected, the weak fricatives are well-separated from the strong fricatives. For example, the mean amplitude differences between /θ/ and /s/, which have the closest relative location of supraglottal constrictions, range from 13.7 to 19.9 dB for measures made at fricative midpoint and from 12 to 21.5 dB for measures made at the right edge of the fricative.

We can also compare these normalized amplitudes averaged separately for the weak voiceless fricatives (/f, θ/) and for the strong voiceless fricatives (/s, š/), which we have grouped as nonstrident and strident fricatives, respectively. The average normalized amplitudes for nonstrident and strident fricatives, measured at the edge of the fricative (at relative time = $CV - 20$) and normalized with respect to F1 amplitude (at relative time = $CV + 20$) are shown in Table 1. The difference between the grand average means for nonstrident and strident fricatives, computed as the average of the means of individual subjects, is 17 dB.

4. DISCUSSION

In quantifying the time-varying spectra of fricatives, we asked the following questions: How big a change and in which frequency regions? The observed amplitude changes of -13 to +8 dB from the fricative midpoint to the fricative-vowel boundary exceed the expected error from the noise source, and presumably reflect movements of the major articulators in forming and releasing the supraglottal constriction.

The edges were stronger than the middle for all fricatives in the mid-frequency bands (1-4 kHz), consistent with observed excitation of the second and third formants near the vowel boundary, and with the presence of aspiration in the vicinity of the fricative-vowel boundary. It should be noted that back cavity excitation can also reflect incomplete pole-zero cancellation

Table 1: *Means and standard deviations of normalized amplitudes (dB) for the nonstrident and strident fricatives for all four subjects. The normalized amplitudes were found by subtracting the amplitude of the first formant (at $CV + 20$ ms) from the maximum peak above 2000 Hz (at $CV - 20$ ms).*

Normalized Amplitudes (dB)			
Speaker		Means	S.D.
F1	Nonstrident	-41	4.74
	Strident	-23	5.58
F2	Nonstrident	-31	3.88
	Strident	-16	5.46
M1	Nonstrident	-33	3.71
	Strident	-18	4.93
M2	Nonstrident	-41	8.11
	Strident	-19	5.31
Average	Nonstrident	-36	7.07
	Strident	-19	5.90

which can occur when there is coupling between the front and back cavities. Often, there is a short (less than 20 ms) gap, where neither the frication nor aspiration noise is very strong. A short gap in energy at the boundary between a voiceless fricative and the following vowel could be interpreted as reflecting that the supraglottal constriction is released before the glottis is closed. Presumably this reflects the mistiming between turning off the noise source for the fricative and turning on the voicing source for the following vowel.

Significant spectrum amplitude differences were observed at higher frequencies (4-8 kHz) between the nonstrident and strident fricatives. For the strident fricatives, the highest frequencies are strongest in the middle of the consonant, when the cross-sectional area of the supraglottal constriction may reach its minimum. This finding

is consistent with a previous study [1], an LPC analysis of voiceless fricatives preceding five vowels, in which high-frequency peaks to tended to appear more often in the midpoint of a fricative than in the initial or final 15 ms.

The nonstrident fricatives in English show greater overall variability in amplitude than the stridents. Results of Utman and Blumstein [6] suggest that the realization of an acoustic property is influenced by the linguistic role its associated feature plays in a particular language's sound inventory.

The normalized amplitudes of the weak and strong fricatives in English were well-separated: the maximum amplitude above 2 kHz in the fricative, normalized relative to vowel amplitude, is 15-20 dB more for /s/ and /š/ than for /f/ and /θ/. Spectral differences between strident and nonstrident fricatives suggested that models of the filtering of the noise source by the front cavity might be improved if the losses in the vocal tract were better represented and if better estimates could be made of the source location.

5. SUMMARY

In the present analysis, the amplitudes in restricted frequency regions of fricative noise were examined with respect to the neighboring vowel. It was hypothesized that relative measures could be found to capture important characteristics of the time-varying noise and reduce the dimensionality of the data. Studying noisy speech sounds yields inherently noisy findings. We observe noise variations over time, variations from one token to another and inter-speaker variability. Our calculations of the amplitude variations in selected frequency bands for English fricatives guide understanding of the considerable variability.

ACKNOWLEDGEMENTS

This research was partially supported by NIH.

REFERENCES

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

[2] Shadle, C. H., Moulinier, A., Dobelke, C. U., and Scully, C. (1992), "Ensemble averaging applied to the analysis of fricative consonants," in *Proceedings of the International Conference on Spoken Language Processing*, Vol. 1, (Banff, Alberta, Canada), 53-56.

[3] Xu, Y. and Wilde, L. (1994), "Combining time-averaging and ensemble-averaging in analyzing voiceless fricatives in Mandarin", *J. Acoust. Soc. Am.* 96(5), Pt. 2, 3230. features", *Proc. ICSLP*, 1992, pp. 499-502.

[4] Klatt, D. H. (Chapter 6, unpublished manuscript), "Fricative Consonants."

[5] Wilde, L. F. "Analysis and synthesis of fricative consonants", *Ph.D. Thesis*, Dept. Electrical Engineering and Computer Science, MIT, Feb 1995.

[6] Utman, J. A. and Blumstein, S. E. (1994), "The influence of language on the acoustic properties of phonetic features: A study of the Feature [strident] in Ewe and English." *Phonetica*, 51(4): 221-238.