Quantifying Spectral Characteristics of Fricatives

Christine H. Shadle [1,2] *and Sheila J. Mair* [2]

[1] ATR Human Information Processing Research Laboratories,
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan;
[2] Department of Electronics and Computer Science,
University of Southampton, Southampton SO17 1BJ, United Kingdom.

ABSTRACT

In a search for spectral parameters that can be used to distinguish and to model fricatives, spectral moments, dynamic amplitude, and slope above maximum amplitude were computed for a fricative corpus including sustained fricatives at different effort levels, and fricatives in vowel context. Moments varied significantly by frequency range used in computation. M3 appeared to vary the least across fricative, contrasting with Forrest et al.'s 1988 study. Dynamic amplitude separated sibilants and non-sibilants, as predicted; slope above the maximum amplitude varied significantly with effort level.

1 Introduction

Many attempts have been made to define quantitative parameters describing fricatives. In some cases the purpose has been to find characteristics that will distinguish one fricative from another, across vowel context and speaker. These are descriptive parameters, such as the frequency range of the high-energy part of the spectrum and overall sound pressure level (e.g.[1]), and spectral moments [2,3,4]. In other cases the purpose has been to find values appropriate for source-filter models, and include parameters such as SPL and spectral tilt, to describe the source, and pole and zero frequencies, to describe the filter [5,6,7]. The results for both approaches, however, have generally been disappointing.

In this study a third approach is taken, which is first to consider the parameters that have been used in light of knowledge of the acoustic mechanisms obtained from speech and mechanical model studies, and then to define new parameters that are more closely related to those mechanisms. First, it is clear that fricatives have formants whose frequencies are related to the vocal tract configuration [8]. Thus, parameters specifying high-amplitude regions of the spectrum in terms of absolute frequency (e.g. [1]) will fail when used on subjects with different length vocal tracts. Second, fricative place can shift according to vowel context, which shifts formants and also antiresonances [9, 10]. This will likewise affect measures specified in terms of absolute frequency. Third, both small and large differences in spectral shape can be acoustically significant, yet may not be equally well incorporated in spectral measures. For instance, the presence of well-defined troughs indicate zeros, and their frequency can be a sensitive indicator of source location. Yet any measure based on the LPC spectrum (e.g. by [4, 11]), which cannot model zeros, will obscure troughs. Spectral moments based on the DFT will include the effect of troughs, but may not be sensitive to small, though significant, shifts in frequency. Fourth, in some cases vowel context affects source characteristics, by resulting in whistling or simply a less efficient noise source [9,10]. Measures of the source are difficult to obtain from the radiated spectrum alone; measures of spectral tilt are confounded by shifting formants, and in any case need to be extended to frequencies well above 10 kHz.

Jassem [2] used many measures of spectral shape, including moments, on an extensive corpus of Polish fricatives. The sibilants had the best prediction rate; the weak fricatives, the poorest. Forrest et al. [3] used many speakers and investigated moments more thoroughly. However, the fricative part of their corpus was relatively small, using only word-initial position: 'see, she, fought, thought, fat'. Wrench [11] used moments and locus equations, but only on two allophones of [s], and [ʃ]. Finally, none of these studies has used spectral analysis above 10 kHz, and some have been considerably more restrictive [12]. Some aspects of spectral shape differences between the front fricatives occur mainly above 10 kHz; spectral tilt, a prime indicator of noise source characteristics, is less obscured by formants at higher frequencies.

It is our assertion that spectral moments do not make full use of our existing knowledge of the acoustics of fricatives. We wish to develop more physically-based parameters that will serve both to distinguish fricatives and to characterize differences induced by vowel context, effort level, and speaker characteristics. Therefore, a dual approach was taken to arrive at a set of quantitative parameters. First, an extensive fricative corpus that has already been intensively studied has been employed here. Second, both spectral moments (as defined by Forrest et al.) and some more physically-based parameters have been computed for this corpus, in an effort both to more fully explore the potential usefulness of moments for fricatives, and as a control on the newer

parameters.

2 Method

The speech corpuses consisted of the fricatives [f, θ, s, ʃ, v, ð, z, ʒ] in two environments: (1) preceded by the vowel [a] and sustained for 3 s, and (2) inserted into the nonsense words [pV_1FV_2] and repeated 10 times on a single breath, for V_1, V_2 chosen from [a,i,u]. Two subjects were recorded: a man speaker of French (PB; recorded twice, three years apart) and a woman speaker of American English (CHS; recorded once). In the first set of recordings (by PB and CHS), six tokens were recorded of each sustained fricative, in randomized order. In the second set of recordings (PB only), each sustained fricative token was produced at three effort levels: normal, soft and loud, in that order.

The acoustic recordings in all cases were made under the 'High-Fidelity conditions' reported previously [9]. Averaged power spectra from 0 to 17 kHz were computed at beginning, middle and end of the steady-state portion of the fricatives in vowel context, using ensemble averaging [13], and in the center of the sustained fricatives, using time averaging. In both cases eight Discrete Fourier Transforms, each computed from a 20-ms Hanning windowed portion of the speech signal, were averaged to form the averaged power spectrum. For ensemble averages, each window was located at the same 'event' (mid-fricative, or beginning of steady state) in eight successive tokens of the same [pVFV] item; for time averages, the windows were placed adjacent to each other (i.e. with no overlap) in the center of the fricative.

Spectral moments were computed as defined in [3], with some slight changes appropriate to the corpus. First, the moments were computed from the magnitude-squared averaged spectral amplitudes, excluding the dc value. Third and fourth moments were normalized by powers of the variance to generate dimensionless M3 and M4; however, the factor 3 was not subtracted from M4 since the values ranged so widely that this threshold appeared rather arbitrary. These moments were computed for the time- and ensemble-averaged spectra described above, for 6 frequency ranges (abbreviated F.R.):

1. 50 – 16950 Hz, maximum range of our data
2. 50 – 10 kHz, used by Forrest et al [3]
3. 50 – 5 kHz, used by Wilde [12]
4. 200 – 16950 Hz, voiced fricatives and maximum range of our data
5. 200 – 10 kHz, voiced fricatives and upper limit of [3]
6. 100 – 6 kHz, used by [7] for spectral tilt

In addition, the dynamic amplitude A_d was defined for unvoiced fricatives as the difference in amplitude, expressed in dB's, between the minimum amplitude value between 0 and 2 kHz, and the maximum amplitude occurring between 0.5 and 17 kHz. This parameter is a simple measure of the noise source characteristics, and has been shown to be effective in separating spectra of mechanical models with two noise source types, the 'obstacle' and 'no-obstacle' cases [5]; it was thus expected that it might differentiate between sibilants and non-sibilants, and perhaps between effort levels, where non-sibilants and lower effort would be expected to produce lower values of A_d.

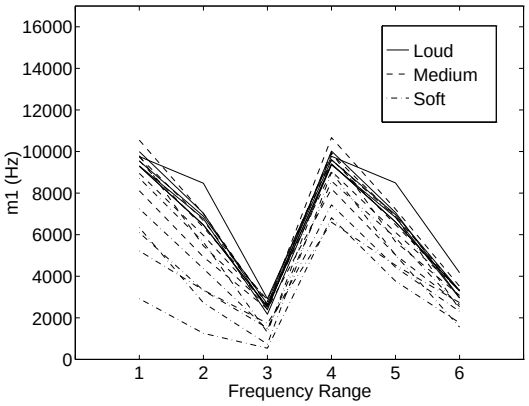

Figure 1: First moment for sustained [f], plotted against frequency range. Frequency ranges 1, 2, and 3 are, respectively, 0.05 – 16.95, – 10, –5 kHz; ranges 4, 5, 6 are .2 – 16.95, 0.2 – 10, 0.1 – 6 kHz.

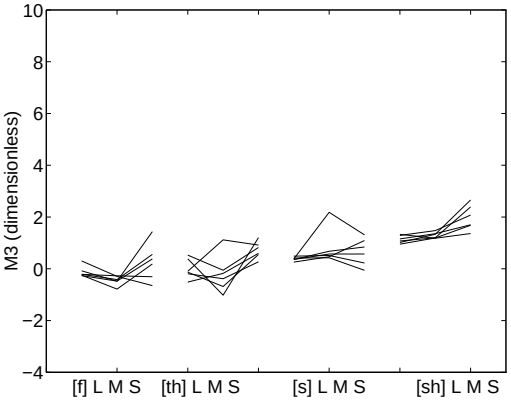

Figure 2: Normalized 3rd moment for sustained fricatives, 6 tokens at each of 3 effort levels, loud (L), medium (M) and soft (S). F.R. = 1.

Spectral slopes were computed by fitting a line to the spectrum over various frequency ranges: for all fricatives, from 11 – 16.95 kHz; for all fricatives, over all of the ranges used to compute spectral moments; for unvoiced fricatives, from the frequency of the spectral maximum amplitude, f_{max}, up to 16.95 kHz; for voiced fricatives, from each of the spectral maxima defined to exclude the voicing-excited formants, to 16.95 kHz.

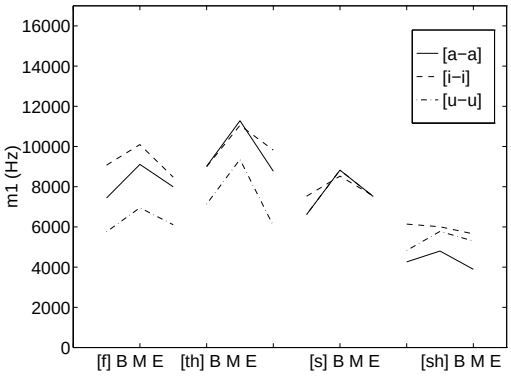

Figure 3: First moment for VCV Corpus, for 3 vowel contexts and 3 locations within the fricative of spectral analysis: beginning (B), middle (M) and end (E). F.R. = 1.

3 Results

First, the values of the moments were strongly affected by the frequency range used. Figure 1 shows an example of the first moment of sustained [f] for PB. This amount of variation is typical across the unvoiced fricatives, all moments, although the first moment is relatively constant with frequency range for [s, ʃ]. For [s], the third frequency range produces more extreme values of M3 and M4 (respectively, very small and very large) than any other frequency range. The order of the tokens, and particularly where the different effort levels occur, changes sometimes with frequency range.

Second, moments of multiple tokens of sustained fricatives showed generally more variation across tokens within a fricative than across fricatives. Figure 2 shows M3 for the unvoiced case. Forrest et al. indicated that m1, M3 and M4 were the most useful in distinguishing between fricatives, but for this corpus, m1, sd and M4 varied more across fricative.

In general, the moments did show the expected relationships. Standard deviation was large for the weak, relatively flat fricatives; M3, skewness, changed the most with frequency range.

Moments m1, sd and M4 computed for the VCV Corpus are shown for frequency range 1 in Figs. 3-5. The most striking change due to vowel context in this corpus was the [u-u] context for [s], which resulted in a fairly narrow high-amplitude peak with 'shoulders' considerably lower in amplitude than the [a-a, i-i] contexts [9]. This effect is apparent in all moments except the first, and is most striking in the high value of M4 (kurtosis). Standard deviation is higher for [f,θ], as in the sustained corpus, but drops mid-fricative, when the spectra become less flat. M3 is relatively constant across fricatives for frequency ranges 1 and 2 (not shown). m1 varies with vowel context and position; all m1 values for [ʃ] are significantly different from those of the other fricatives.

Figure 6 shows dynamic amplitude for the sustained corpus.

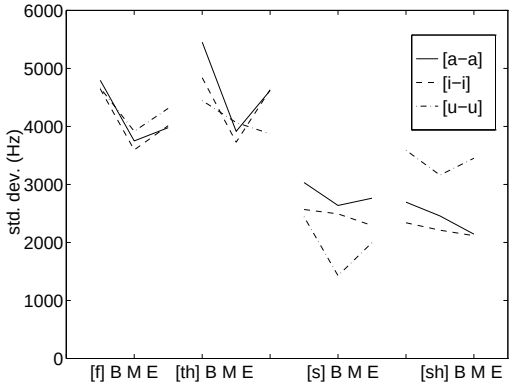

Figure 4: Standard deviation for VCV Corpus, for 3 vowel contexts and 3 locations within the fricative of spectral analysis: beginning (B), middle (M) and end (E). F.R. = 1

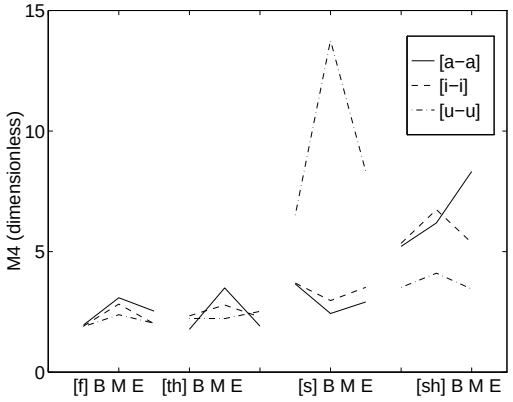

Figure 5: Normalized fourth moment for VCV Corpus, for 3 vowel contexts and 3 locations within the fricative of spectral analysis: beginning (B), middle (M) and end (E). F.R. = 1.

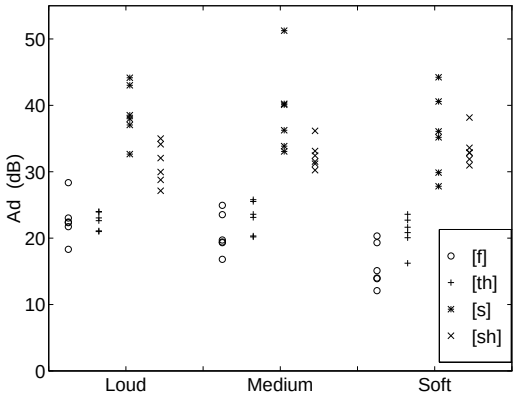

Figure 6: Dynamic amplitude for sustained fricatives, 6 tokens of each of three effort levels (loud (L), medium (M) and soft (S)).

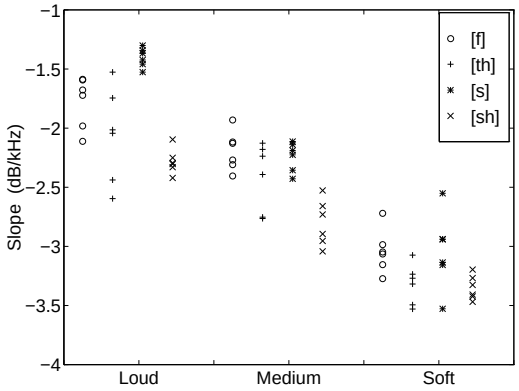

Figure 7: Slope of line fit from frequency of maximum spectral amplitude to 16.95 kHz, plotted for sustained fricatives, 6 tokens of each of three effort levels (loud (L), medium (M) and soft (S)).

Two effects are noticeable: [s,ʃ] have A_d higher than [f,θ] at all effort levels, and A_d decreases slightly with effort level for all fricatives except [ʃ]. Acting in nearly the opposite manner was spectral slope, which does not discriminate between fricatives, but did change noticeably with effort level, in a consistent way for all fricatives. (See Fig. 7.)

4 Conclusion

Spectral moments and other parameters describing spectral shape have been computed for an extensive fricative corpus that has already been studied for the acoustic effects of vowel context, effort level, repetition and window location. Spectral moments do not distinguish reliably between fricatives, but the variations were consistent with acoustic analysis. The moments proved to be quite sensitive to frequency range, and relatively insensitive to position within the fricative (in VCV's) even though significant acoustic changes with position are well-documented. Unlike [3], 2nd moment appeared to vary more across fricative than did M3.

The dynamic amplitude parameter worked well to separate [s,ʃ] from [f,θ] across effort levels and repetitions. Spectral slope from f_{max} up showed a much stronger effect with effort level, and did not show a significant effect with fricative. We do not so far have a parameter that will distinguish the front fricatives from each other.

Many more relationships have been investigated than could be presented here, in particular with the voiced fricatives in the corpus, and additional parameters have been suggested by the findings to date. The search for parameters that can be automatically measured, that use our knowledge of the acoustics of production of fricatives, and that either reliably distinguish fricatives or deliver parameters useful for modelling continues.

Acknowledgment

The fricative corpuses used were developed with the support of an EC SCIENCE grant CEC-SCI*0147C(EDB), and a European CEC-ESPRIT project Speech MAPS.

5 References

1. Strevens, P. "Spectra of fricative noise in human speech." *Lang. and Speech* 3, 32-49, 1960.
2. Jassem, W. "Classification of fricative spectra using statistical discriminant functions." In B. Lindblom and S. Ohman, eds., *Frontiers of Speech Communication Research*, New York: Academic Press, 77-91, 1979.
3. Forrest, K., Weismer, G., Milenkovic, P. and Dougall, R.N. "Statistical analysis of word-initial voiceless obstruents: Preliminary data." *J. Acoust. Soc. Am.* 84, 115-123, 1988.
4. Jongman, A. and Sereno, J.A. "Acoustic properties of non-sibilant fricatives." *Proc. of ICPhS* v.4, Stockholm, 432-435, 1995.
5. Shadle, C.H. *The Acoustics of Fricative Consonants.* M.I.T. PhD thesis, R.L.E. Tech. Rpt. 506, 1985.
6. Badin, P., Shadle, C.H., Pham Thi Ngoc, Y., Carter, J.N., Chiu, W.S.C., Scully, C. and Stromberg, K. "Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis." *Proc. of ICSLP-94*, Yokohama, 163-166, 1994.
7. Badin, P., Mawass, K. and Castelli, E. "A model of frication noise source based on data from fricative consonants in vowel context." *Proc. of ICPhS* Stockholm, v.2, 202-205, 1995.
8. Shadle, C.H., Badin, P. and Moulinier, A. "Towards the spectral characteristics of fricative consonants." *Proc. ICPhS*, Aix-en-Provence, v.3, 42-45, 1991.
9. Shadle, C.H. and Scully, C. "An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences." *J. Phonetics* 23, 53-66, 1995.
10. Shadle, C.H., Mair, S.J. and Carter, J.N. "The Acoustic characteristics of the front fricatives [f, v, θ, ð]." *Proc. ETRW*, Autrans, 1996.
11. Wrench, A. "Analysis of fricatives using multiple centres of gravity." *Proc. ICPhS* 4, 460-463, 1995.
12. Wilde, L. "Quantifying time-varying spectra of English fricatives." *Proc. ICPhS* v.4, Stockholm, 120-123, 1995.
13. Shadle, C.H., Moulinier, Dobelke, C.U., and Scully, C. "Ensemble averaging applied to the analysis of fricative consonants," *Proc. of ICSLP-92*, Banff, v. 1, 53-56, 1992.