PARAMETERIZING SPECTRAL CHARACTERISTICS OF EUROPEAN PORTUGUESE FRICATIVES

Luis M. T. Jesus and Christine H. Shadle
Department of Electronics and Computer Science
University of Southampton, Highfield, Southampton, SO17 1BJ, UK
e-mail: {lmtj97r, chs}@ecs.soton.ac.uk

ABSTRACT

Fricative spectra from four native Portuguese speakers have been parameterized in order to aid comparisons across speaker and across corpus, and to gain insight into the production mechanisms underlying the language-specific variations. The parameters, derived from previous studies, capture source-related changes for the most part as predicted; for the sustained fricative, they also separate fricatives by place.

1 INTRODUCTION

As part of our on-going study of Portuguese fricatives, we have developed corpora that include real words and phonologically-possible nonsense words, and performed acoustic analysis of recordings of four native speakers. Principle findings reported to date include devoicing of voiced fricatives, especially in destressed syllables. Only /ʃ/ appears in final word position, but other fricatives can appear nearly word-finally, followed by only a reduced vowel [3].

The multiple comparisons possible in such a study, across speakers, corpora, place, vowel context, syllable stress, location within fricative, etc., demand a systematic approach, since our interest is primarily in the production mechanisms of the fricatives and the language-specific variation of these mechanisms. We seek a way of parameterizing the fricatives that makes use of our knowledge of the underlying aeroacoustics.

Many ways of parameterizing fricatives exist in the literature. Wilde studied acoustic cues (place and voiced/voiceless categorization) in fricative-vowel boundaries. Results showed that voiceless fricatives are more dependent on vowel context, and that voicing onset time and formant structure provide important place information. She also showed that the amplitudes of fricative noise in restricted frequency regions can distinguish sibilants from nonsibilants [10].

Jongman et al. [4, 5] studied spectral moments, locus equations, the spectral peak location, and noise duration and amplitude, as cues to place of articulation of English fricatives. Spectral peak location and noise duration distinguished sibilants from non-sibilants; spectral peak location separated /s,z/ from /ʃ,ʒ/; the amplitude distinguished all four places of articulation. The slope of locus equations could be used to differentiate labiodental from the other three places of articulation. The first moment and spectral mean distinguished all places of articulation.

Sussman [9] also used locus equations on fricatives but without much success, although this technique worked well on stops.

Evers et al. [1] studied the acoustic characteristics of fricatives /s/ and /ʃ/ produced by two speakers of English, Bengali and Dutch. They used power spectra computed from a single 40 ms window placed mid-fricative, and calculated the slopes of linear regression lines fit to spectra from 0 to 2.5 kHz (S_a) and from 2.5 kHz to 8 kHz (S_b). Their results showed that it was possible to separate /s/ from /ʃ/ by using the difference in slope below and above 2.5 kHz, i.e., $(S_a - S_b)_ʃ > (S_a - S_b)_s$. The slope difference was successful in categorizing the two sibilants within a range of 7-15 dB/kHz across the three languages. Results also showed that there is no vowel influence in the discrimination, and that there is a variation between speakers.

Forrest at al. [2] used spectral moments to characterize normal speech with the intent of using them on disordered speech. Results showed that spectral moments worked well to classify stops but could not distinguish all fricatives. However, the authors used a very limited corpus.

Shadle and Mair [7] used spectral moments (as in [2]) on a large fricative corpus recorded by one American English and one French native speaker. The moments that were the most useful for distinguishing fricatives in [2] proved not to be, when used on multiple tokens, varying effort levels, different vowel contexts, and three different locations within a fricative. Two additional parameters, dynamic amplitude and spectral slope, were defined. These did not distinguish the fricatives completely but did vary with source location and effort level as predicted.

Parameters similar to those used in [7], and $S_{p'}$ similar to S_a used in [1], were used in this study in order to compare fricatives across-speaker, relate the more controlled productions (sustained and nonsense words) to those of real words, and gain insight into the production mechanisms underlying the variations specific to Portuguese.

2 METHOD

We have used part of a large corpus of Portuguese fricatives /f, v, s, z, ʃ, ʒ/, which includes sustained fricatives preceded by vowels /ɨ, ɐ, u/ (Corpus 1a), and fricatives sustained at different effort levels (Corpus 1b). There is also a corpus of nonsense words (Corpus 2), /pV_1FV_2/, where V_i is one of /ɨ, ɐ, u/, and a corpus of real words (Corpus 3) produced in a frame sentence. The subjects used in this study were two male (LMTJ and CFGA) and two female (ACC and ISSS) adult Portuguese native speakers. Recordings were made in a sound-treated booth (B & K 4165 1/2 inch mic. 1 m from the subject and B & K 2636 measurement amplifier), onto a Sony TCD-D7 DAT recorder. Averaged power spectra were computed using nine 10 ms Hamming windows and time-averaging for Corpus 1a, 1b and 3; ensemble-averaging was used at the beginning, middle or end of the fricatives in the repeated words in Corpus 2 (more detail is given in [3]).

3 PARAMETERIZATION

The parameters used were defined first from mechanical model results [6] and further developed as a potential tool for classifying fricatives using real speech [7]. They consist of measures of the dynamic range of the spectrum, and spectral slope, and are applied to the spectrum of the far-field acoustic signal.

The far-field acoustic signal is the result of the excitation of the tract transfer function by the source (for unvoiced) or sources (for voiced fricatives). The transfer function consists of poles, which are the resonances of the entire vocal tract, and zeros, which are antiresonances of the part of the tract upstream of the noise source. If the noise source is distributed, zero frequencies will be correspondingly smeared. An intermediate source location (as for all fricatives) always produces a low-frequency zero. Poles and zeros corresponding to back-cavity resonances tend nearly to cancel. Uncancelled poles correspond to front cavity resonances; their spectral prominence will depend on both the losses (especially radiation losses) and the noise source strength at their respective frequencies.

The noise source spectrum depends on the shape of the constriction, the tract downstream of it, and the flow velocity through it. The noise source spectrum envelope has its highest amplitude at low frequencies and falls off smoothly. If the tract geometry remains the same and flow velocity is increased, the noise spectral envelope increases in amplitude at all frequencies, *but more so at higher frequencies* [6]. The noise source is weaker in voiced fricatives than their unvoiced counterparts.

If our goal is identification of the fricative spoken regardless of its context or the way in which it was spoken, we are then interested in the transfer function, since the peak frequencies offer clues to the place, and in the source type, since that not only differentiates voiced and unvoiced versions, but, in indicating whether the source is localized or distributed, again offers clues to the place. If our goal instead is to describe the acoustic variation caused by the context or the way in which a particular fricative is spoken, we are then interested in the source spectrum, since it offers clues to the source variations across subject and corpus. In this study we are primarily interested in the latter goal.

Figure 1 illustrates the four parameters that we consider in this paper. F is the frequency of the spectral peak between 0.5 and 20 kHz having maximum amplitude. The dynamic amplitude, A_d, is the difference in amplitude between the minimum amplitude occurring between 0 and 2 kHz, and the amplitude at F. Two linear regression lines are fit to the spectrum; $S_{p'}$ is the slope of the line fit to all the points from 500 Hz to F, and S_p is the slope of the line fit to all the points from F to 20 kHz.

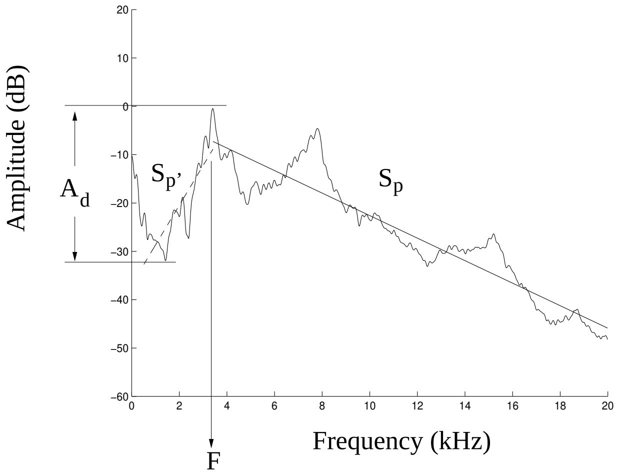

Figure 1: Dynamic amplitude A_d, and regression lines used to calculate low frequency (500 Hz to F kHz) slope $S_{p'}$ (dashed line) and high frequency (F kHz to 20 kHz) slope S_p (solid line). Sustained fricative /ʃ/ (Corpus 1a) produced by Speaker ISSS.

Given these definitions, we can make the following predictions. The parameter F should be related to place of the fricative, decreasing as place moves posteriorly. The parameter A_d should be maximized for a localized source, and for higher source strength, as in sibilants, and unvoiced fricatives. The parameter S_p should be related to the source strength. Although the resonance peaks will affect the line fit, they should affect the fit in the same way for within-fricative comparisons. Thus, for a given fricative where transfer function will vary only slightly from token to token, S_p should increase, i.e. become less negative, as flow velocity through the constriction increases. Effort level and syllable stress should be correlated with increased flow velocity; the velocity should also be at a maximum mid-fricative, when constriction area is smallest and pressure across the constriction highest. The parameter $S_{p'}$ should be similar to A_d. For a fricative with a localized source and posterior place, $S_{p'}$ will be the largest. Within a fricative, increased $S_{p'}$ should be correlated with either more posterior place (due, for instance, to a more rounded vowel context) or greater source strength.

We therefore predict that on an $S_{p'}$ vs. S_p plot, each place will cluster separately, with voiced tokens having lower $S_{p'}$ but similar S_p relative to their unvoiced counterparts.

4 RESULTS

4.1Sustained Fricatives

Figure 2 shows average regression line fits (from F to 20 kHz) to the spectra of the sustained fricatives in Corpus 1b. Each graph corresponds to a single place, and shows lines for three effort levels, voiced and unvoiced. Clearly, each place has a different "family" of nearly-parallel lines; higher effort level increases amplitude significantly and slope slightly, as predicted. The families of lines for the voiced and unvoiced fricatives always overlap, with the voiced cases mostly lower in amplitude and occupying a smaller range of amplitudes than the unvoiced cases.

Figure 3 shows S_p vs. effort level for subject ACC. Again, slope generally increases with increased effort level, though this pattern is much more consistent for unvoiced fricatives. This is consistent with results in [7].

For all subjects, Corpus 1a and 1b, /s, z, ʃ, ʒ/ have a higher A_d than /f, v/, as predicted; this parameter also differentiates between voiced fricatives and their unvoiced counterparts. On an A_d vs. S_p plot, sustained fricatives form two distinct clusters, of sibilants and /f,v/.

Figure 4 shows $S_{p'}$ vs. S_p values plotted for corpus 1a, subject CFGA. Results for speakers LMTJ, CFGA and ISSS, confirm the findings of Evers et al. [1], i.e., that it is possible to separate /s/ from /ʃ/. In fact, for these subjects, fricatives /f,v/, /s,z/ and /ʃ,ʒ/ form clusters in the feature space, i.e., they are separated by place; as predicted, the voiced tokens of each had lower $S_{p'}$ and similar S_p than their unvoiced correlates. For ACC the voicing relationship was maintained, but /s,z/ tokens fell inbetween the /ʃ/ and /ʒ/ tokens.

4.2 Fricatives in Context

In Figure 5, A_d and S_p are plotted vs. location of the analysis window within the fricative for Corpus 2, subject LMTJ. For /f,v/ there is no consistent pattern; results in [8] indicate that the vowel context may play more of a role.

As for the sustained fricatives, A_d separates sibilants from /f, v/. A_d is higher on average at the middle of the fricative than at the beginning and end for /s, z, ʃ, ʒ/, as predicted.

Preliminary comparisons of stressed and destressed fricatives indicate no or little change in A_d and S_p, not as predicted. We note, though, that syllable stress is strongly correlated with the amount of devoicing, and since Portuguese fricatives devoice in over one half of words [3], there may be some interaction of these parameters.

5 CONCLUSIONS

The parameters spectral slope, frequency of maximum amplitude, and dynamic amplitude, were developed to characterize fricative spectra, and applied to corpora recorded by four native Portuguese speakers. The parameters behaved as predicted for changes in effort level, voicing, and location within the fricative. Some combinations were also useful for separating the fricatives by place or by sibilance.

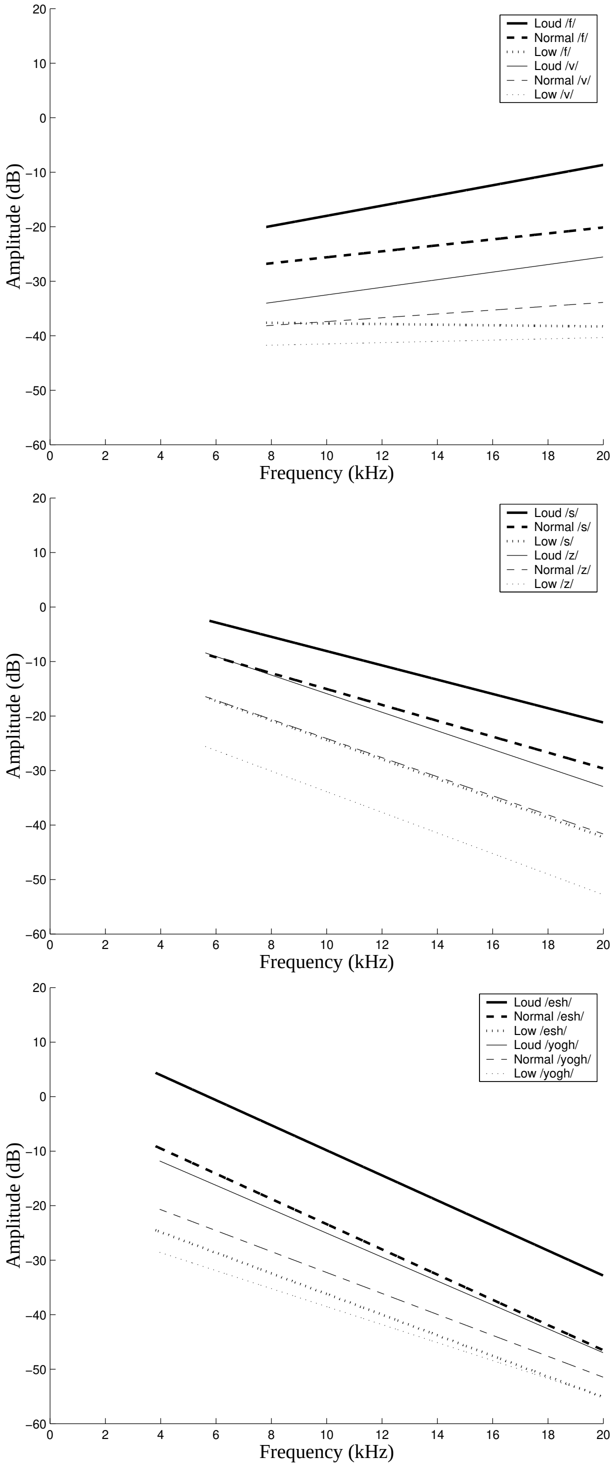

Figure 2: Average regression line of sustained labiodental (top), alveolar (middle) and postalveolar (bottom) fricatives from Corpus 1b at loud, medium and soft effort levels. Speaker ISSS.

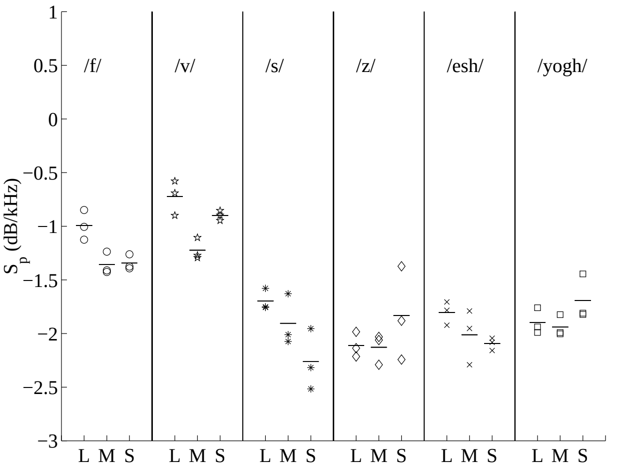

Figure 3: Spectral slope of sustained fricatives from Corpus 1b at Loud (L), Medium (M) and Soft (S) effort levels. The horizontal line is the average value of all the examples. Speaker ACC.

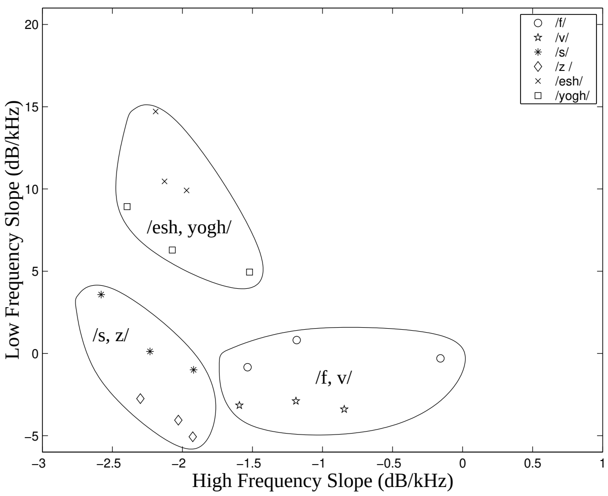

Figure 4: Interactions between the low frequency $S_{p'}$ and high frequency S_p spectral slopes of fricatives from Corpus 1a. Speaker CFGA.

[This work was partially supported by Fundação para a Ciência e a Tecnologia, Portugal.]

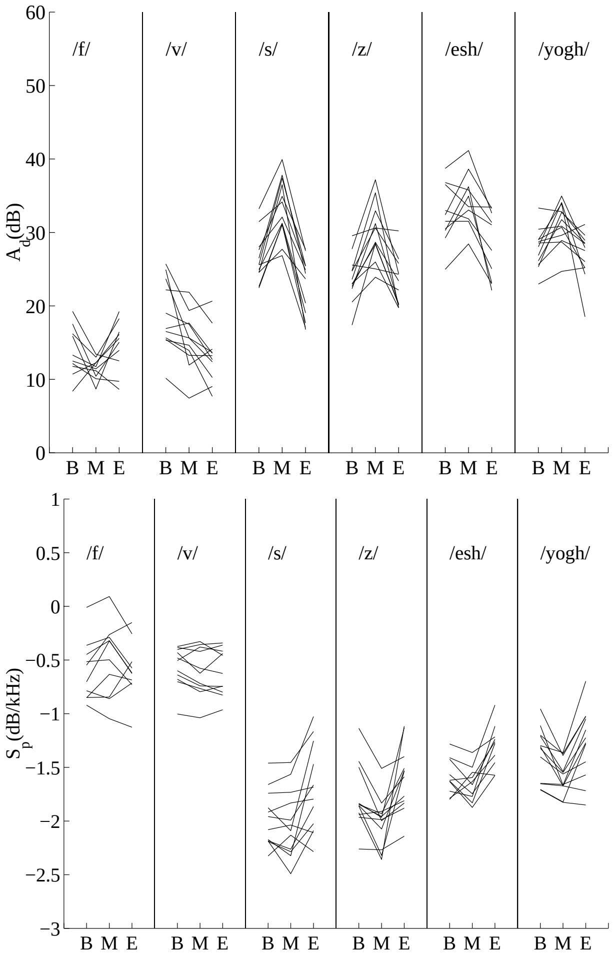

Figure 5: Dynamic amplitude (top) and spectral slope (bottom) of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative. Speaker LMTJ.

REFERENCES

[1] V. Evers, H. Reetz, A. Lahiri (**1998**). "Crosslinguistic Acoustic Categorization of Sibilants Independent of Phonological Status", Journal of Phonetics, **26**, 345-370.

[2] K. Forrest, G. Weismer, P. Milenkovic and R. Dougall (**1988**). "Statistical Analysis of Word-Initial Voiceless Obstruents: Preliminary Data", JASA, **84**, 115-123.

[3] L. Jesus and C. Shadle (**1999**). "Acoustic Analysis of a Speech Corpus of European Portuguese Fricative Consonants", EuroSpeech'99, **1**, Budapest, Hungary, 431-434.

[4] A. Jongman and J. Sereno (**1995**). "Acoustic Properties of Non-Sibilant Fricatives", ICPhS 95, **4**, Stockholm, Sweden, 432-435.

[5] A. Jongman, J. Sereno, R. Wayland and S. Wong (**1998**). "Acoustic Properties of English Fricatives", 16th ICA, **4**, Seattle, USA, 2935-2936.

[6] C. Shadle (**1985**). "The Acoustics of Fricative Consonants", Ph.D. Thesis, MIT, Cambridge, USA.

[7] C. Shadle and S. Mair (**1996**). "Quantifying Spectral Characteristics of Fricatives". ICSLP 96, Philadelphia, USA, 1517-1520.

[8] C. Shadle, S. Mair and J. Carter (**1996**). "Acoustic Characteristics of the Front Fricatives [f, v, θ, ð]". 4th Speech Production Seminar, Autrans, France, 193-196.

[9] H. Sussman (**1994**). "The Phonological Reality of Locus Equations Across Manner Class Distinctions: Preliminary Observations", Phonetica, **51**, 119-131.

[10] L. Wilde (**1995**). "Analysis and Synthesis of Fricative Consonants", Ph.D. Thesis, MIT, Cambridge, USA.