ACOUSTIC ANALYSIS OF A SPEECH CORPUS OF EUROPEAN PORTUGUESE FRICATIVE CONSONANTS

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

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, www: http://www.isis.ecs.soton.ac.uk/∼{lmtj97r, chs}

10.21437/Eurospeech.1999-111

ABSTRACT

The study of Portuguese fricatives is a complex problem, which has not been explored fully. As part of a larger study of the acoustic properties of Portuguese fricatives, corpora of Portuguese words containing /f, v, s, z, ʃ, ʒ/, nonsense words with Portuguese phonology, and sustained Portuguese fricatives have been recorded and analysed. Results show that more than half of the voiced fricatives devoice. Averaged power spectra were computed for all fricatives; differences between fricatives, in particular related to syllable stress, are described.

1 INTRODUCTION

Portuguese is an important European language, spoken by over 180 million people worldwide. Portuguese is unusually rich in instances of vowel reduction, consonant clusters, and fricatization of plosives [9]. Further regional variations result in some instances of substitution of fricative /ʃ/ by affricate /ʧ/ in the north of Portugal, and various occurrences of phonemic variation (e.g. [s̺] as produced in Viseu).

Lacerda [3] describes the use of perceptual experiments to identify the acoustic features of /f, s, ʃ/. /f/ has a flat spectrum and low intensity level; /s/ has a broad-band spectral peak between 4.1 kHz and 5.7 kHz, and high intensity level; the energy in the high frequency bands (around 6 kHz) is perceptually important for /s/; /ʃ/ has a high intensity level and an important broad-band spectral peak between 2.7 kHz and 3.5 kHz.

Andrade's [2] study of consonant clusters included an analysis of /asV/ sequences, where V was one of the vowels /ɨ, ɐ, u/. Results showed that the duration of the frication period is longer when the fricative is followed by /ɨ/, due to the fact that this vowel is weakened (or even not produced) in final position.

Viana's [9] results showed that fricative consonants have longer average duration than the neighboring vowels, and that in final position, they have lower average energy than in initial position. /f/ showed the weakest spectra.

Martins, et al. [4] observed that when the vowel /V_1/ in a /V_1ʃV_2/ sequence is weakened (or not produced), its formant structure is somewhat "transferred" to the following fricative /ʃ/, allowing the listener still to perceive the vocalic segment.

In this study we focus on the analysis of frication in Portuguese, by combining analysis of fricative-rich Portuguese words and sentences with techniques developed in previous work using more controlled nonsense utterances.

2 RECORDING METHOD

The subject used in this study was a male adult Portuguese native speaker from the city of Aveiro. A list was created of 154 Portuguese words, each containing one of the fricatives F=/f, v, s, z, ʃ, ʒ/ in combination with the non-nasal vowels V=/i, ɨ, e, ɛ, ɐ, a, ɔ, o, u/. Of the 154 words, 8 follow the same pattern /FV_1FV_2/, 54 present the fricative word-initially, 69 word-medially and 23 word-finally. Strictly, only /ʃ/ can occur in final position [5], but all of the fricatives are elicited word-finally by incorporating words with a reduced /ɨ/ as the final phoneme [1].

Corpus 1 consists of each of these words in the frame sentence "Diga ..., por favor." Corpus 2 consists of 12 sentences containing 61 of the 154 words. Corpus 3 consists of nonsense words, /pV_1FV_2/, where V_i is one of /ɨ, ɐ, u/ and the word follows Portuguese phonological rules [5]. Each word was repeated 12 times on one breath. Corpus 4 consists of sustained fricatives, produced with a short preceding vowel (e.g. /ɨffffff/) at three different effort levels. The nonsense-word and sustained-fricative corpora are similar in design to corpora used previously for American, French and German subjects, and thus facilitate cross-language comparisons.

Recordings were made in a sound-treated booth using a Bruel & Kjaer 4165 1/2 inch microphone located 1 m in front of the subject's mouth. The signal was amplified and filtered by a B & K 2636 measurement amplifier, with high-pass cut-on frequency of 22 Hz and low-pass cut-off frequency of 22 kHz. A laryngograph (Lx) signal was also recorded using a Laryngograph Processor. The acoustic and Lx signals were recorded with a Sony TCD-D7 DAT recorder (16 bits, sampling frequency 48 kHz), and digitally transfered to a computer for analysis.

3 ANALYSIS METHOD

3.1 Temporal Analysis

The time waveforms of all the corpus words were manually analysed to detect the start of the VF transition, the start of the fricative, the end of the fricative, and the start of the FV transition. During the VF transition, amplitude decreases, voicing ends (for unvoiced fricatives) and frication noise starts. During the FV transition, amplitude increases, voicing starts (for unvoiced fricatives) and frication noise ends. These events overlap in time, making the segmentation a somewhat subjective process. However, it is important to segment consistently, since the analysis methods depend on the boundaries so identified.

The Lx signal was also used in segmentation. For unreduced vowels there was always significant voicing, and for the duration of most fricatives Lx changed drastically. While segmenting, we noticed a large number of devoiced examples. To study this phenomenon further we devised a new criterion for devoicing based on both the acoustic and Lx signals.

Smith [8] used a criterion for devoicing in American English based on the amplitude of the electroglottograph (EGG) cycles: "The fricative was considered to be voiced during the portion of its duration that the amplitude of the EGG cycles exceeded one-tenth of the EGG cycle amplitude at the time of maximum energy in the preceding vowel." A study by Pirello et al. [6] presents an alternative measure of voicing based on the acoustic signal: "An amplitude difference greater than 10 dB between the amplitude of the vowel and frication noise was classified as voiceless. A difference of less than or equal to 10 dB sustained over 30 ms was classified as voiced."

We used both the relative value of the amplitude and duration of the acoustic signal, and the amplitude of the Lx signal to determine if a fricative is devoiced. Devoiced is defined as no periodic structure of the Lx signal on the frication interval. Partially devoiced is a few steady Lx cycles during frication; voiced is steady Lx cycles throughout the whole frication, even if the amplitude is much lower than in the vowel. Section 4 gives a detailed analysis of the devoicing patterns found.

The mean and variance of the laryngograph signal were calculated during the VF transition and during the fricative. The relative variance of the two intervals, $\sigma_r(x) = \sigma_t(x)/\sigma_f(x)$, where $\sigma_t(x)$ is the variance of the signal during the VF transition and $\sigma_f(x)$ is the variance of the signal during the fricative, was used as a quantitative criterion for devoicing.

3.2 Spectral Analysis

The first phase of spectral analysis consisted of a study of the averaged power spectra, used to see the broad characteristics of the fricatives. We used time averaging on Corpus 1 with nine 9 ms Hamming windows, one left-aligned to the start of the fricative, one right-aligned to the end of the fricative, and the remaining six windows evenly spaced in between. For the shorter fricatives the windows overlap as much as 98.6%. The same time-averaging technique was used to calculate the spectra of sustained fricatives on Corpus 4 (100 non-overlapping, 9 ms windows).

The time-averaged power spectrum for each fricative is given by

$$\langle X(f) \rangle = \frac{1}{N} \sum_{i=1}^{N} |X_i(f)|^2 \tag{1}$$

where X_i is the DFT of a portion of the fricative signal, x_i, corresponding to the ith windowed segment, and $N = 9$ (Corpus 1) or $N = 100$ (Corpus 4).

Ensemble averaging was used for Corpus 3. The ensemble-averaged power spectrum of each fricative is given by

$$\overline{X(f)} = \frac{1}{9} \sum_{k=1}^{9} |X_k(f)|^2 \tag{2}$$

where X_k is the DFT of a portion of the fricative signal, x_k, corresponding to the windowed segment (18 ms Hamming window placed at the beginning, middle or end of the fricative) of the kth token. The first and last, and any atypical tokens were eliminated to obtain the ensemble of nine tokens.

4 RESULTS

4.1 Temporal Analysis

Observation of laryngograph and acoustic signals of the Portuguese words in Corpus 1 revealed that 50.0% of the examples of fricative /v/, 66.7% of /z/, and 80.0% of /ʒ/ were totally devoiced (see Table 1).

If classifying is instead done using $\sigma_r \geq 15$ as the criterion for devoicing, results are as shown in Table 2. There are a few examples which are classified differently from the initial empirical observation of the laryngograph signal (see Table 1). Still, the percentage of examples from Corpus 1 which were classified in the same category using the two methods is quite high: /v/: 91.7%, /z/: 96.7%, and /ʒ/: 80%. Most of the discrepancies result from cases on the partially devoiced / completely devoiced borderline, giving promise that this automatic measure can be used for future subjects.

Table 1: Inventory of all cases of complete devoicing from Corpus 1 (using an empirical criterion). Values given are number of devoiced examples divided by the total number of examples.

	Word- Initial	Word- Medial	Word- Final	All Pos.
/v/	6/14 (42.9%)	7/16 (43.8%)	5/6 (83.3%)	18/36 (50.0%)
/z/	5/10 (50.0%)	12/18 (66.7%)	2/2 (100%)	20/30 (66.7%)
/ʒ/	7/10 (70.0%)	14/16 (87.5%)	3/4 (75.0%)	24/30 (80.0%)
All Fric.	18/34 (52.9%)	34/50 (68.0%)	10/12 (83.3%)	62/96 (64.6%)

Table 2: Inventory of all cases of complete devoicing from Corpus 1 (using the relative variance criterion). Values given are number of devoiced examples divided by the total number of examples.

	Word- Initial	Word- Medial	Word- Final	All Pos.
/v/	9/14 (64.3%)	10/16 (62.5%)	5/6 (83.3%)	24/36 (66.7%)
/z/	6/10 (60.0%)	15/18 (83.3%)	2/2 (100%)	23/30 (76.7%)
/ʒ/	4/10 (40.0%)	13/16 (81.3%)	3/4 (75.0%)	20/30 (66.7%)
All Fric.	19/34 (55.9%)	38/50 (76.0%)	10/12 (83.3%)	67/96 (69.8%)

The nonsense words from Corpus 3 were analysed with the empirical criteria for devoicing. Results show that 63.4% (64 out of 101) of the tokens of fricative /v/, 54.3% (63 out of 116) of /z/, and 46.8% (58 out of 124) of /ʒ/ were totally devoiced.

Both measures of devoicing occur more often in word-final than word-initial position. Devoicing rate by fricative differs between the two measures, and between Corpus 1 and 3.

The minimum and maximum durations of the fricatives from Corpus 1 are: 86 ms $\leq$ /f/ $\leq$ 182 ms (mean = 129 ms), 37 ms $\leq$ /v/ $\leq$ 135 ms (75 ms), 109 ms $\leq$ /s/ $\leq$ 220 ms (151 ms), 46 ms $\leq$ /z/ $\leq$ 117 ms (81 ms), 76 ms $\leq$ /ʃ/ $\leq$ 194 ms (132 ms), and 60 ms $\leq$ /ʒ/ $\leq$ 139 ms (93 ms). For fricative /s/ in initial position, as the following vowel's place of articulation moves further back, the duration of the fricative diminishes.

4.2 Spectral Analysis

Substantial differences are found between spectra of voiced and unvoiced, same-place fricatives: not only are the voiced spectra lower in amplitude, as expected, but differences in spectral shape occur, particularly for /ʃ–ʒ/. Examples of averaged power spectra of the sustained fricatives (Corpus 4) are shown in Figure 1. In general, amplitude differences between the three effort levels are smallest at low frequencies. The amount of amplitude difference at high frequencies varies with the fricatives, from 10 dB for /f/ to 15 dB for /ʃ/; it tends to be smaller for the voiced fricatives, from 5-10 dB for /v/ to 10 dB for /ʒ/. These differences are associated with source type and strength, and are similar to results for American English and French subjects.

We expected that effort level of sustained fricatives would correspond with stress of the syllable containing the fricative in Corpora 1 and 3, and possibly also position within the word in Corpus 1. Figure 2 contrasts spectra for /ʃ/ with similar vowel context from Corpora 1 and 3: [kɐˈpuʃu] (final /u/ reduced) and [puʃu] (both syllables equal stress). Four of the seven such pairs-by-vowel-context showed this pattern, with the Corpus 3 spectral amplitude consistently higher (by 5-10 dB) than that of Corpus 1 for frequencies above 2 kHz. This amplitude difference across the frequency range corresponds strongly with a difference in stress between the two fricatives. The spectral shapes and amplitudes are similar to the soft effort level for all destressed /ʃ/ fricatives, and to the medium effort level for stressed /ʃ/ fricatives. No fricatives from Corpus 1 or 3 resemble their high-effort-level Corpus 4 counterparts. Some vowel context effects were noted: the main peak in Figure 2, for /uʃu/ context, was at a significantly lower frequency than in, e.g., /ɨʃi/ context; this is as expected from previous work [7].

These points taken together give us information needed to model the fricative. They also indicate that the nonsense word corpus follows Portuguese phonological rules. Corpus 3 is better controlled and easier to analyse than Corpus 1 or 2; validating its use would give an important advantage.

5 CONCLUSIONS

In this paper, preliminary work on the acoustics of Portuguese fricatives has been described. Corpora were designed including Portuguese words, nonsense words following Portuguese phonology, and sustained fricatives at three different effort levels; these were recorded for one speaker. Voiced fricatives devoice in over one half the cases in both nonsense and real words; two measures of devoicing were developed and compared. Spectral analysis revealed a correspondence between the effect of effort level and of syllable stress, and showed some effect of vowel context. Future work will include analysis of other Portuguese speakers.

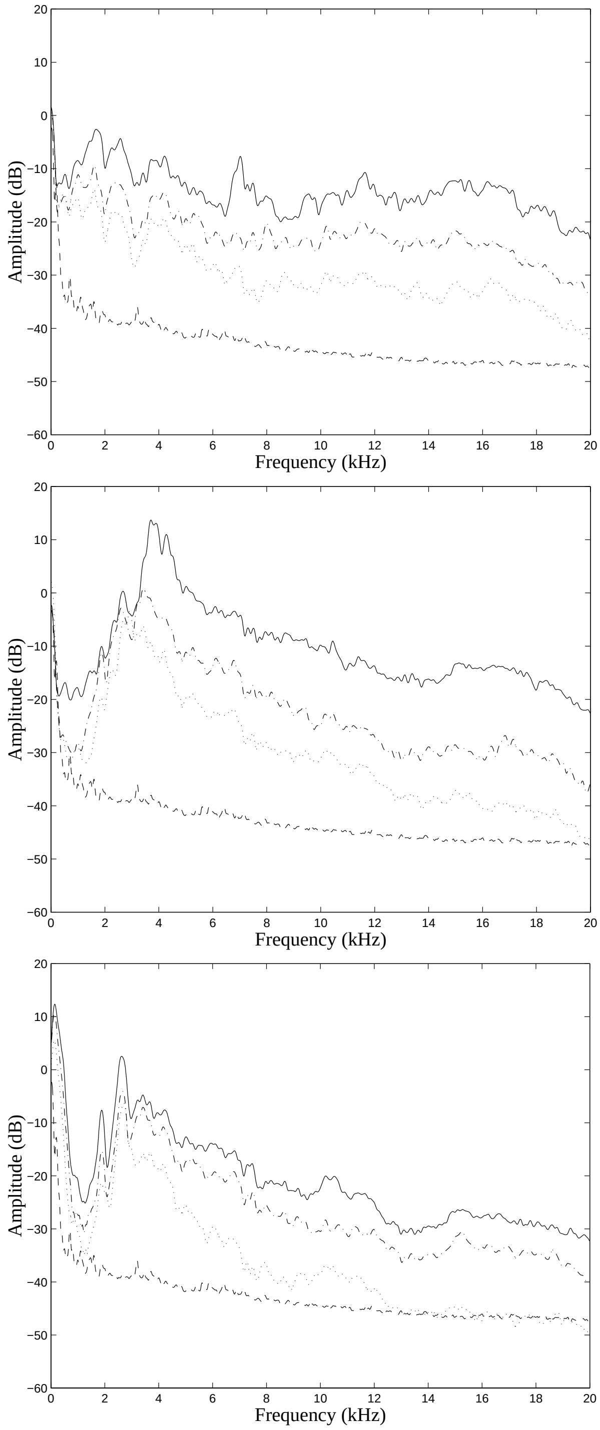

Figure 1: Averaged power spectra of sustained fricatives at three effort levels (Corpus 4). Top: /f/; centre: /ʃ/; bottom: /ʒ/. Solid line: loud; dash-dotted: medium; dotted: soft levels. The dashed curve is the averaged spectrum of the room noise.

ACKNOWLEDGEMENTS

The authors would like to thank Dra. Lurdes de Castro Moutinho at the Universidade de Aveiro in Portugal, for the invaluable help during the design phase of the corpus used in this research. This work was partially supported by Fundação para a Ciência e a Tecnologia, Portugal.

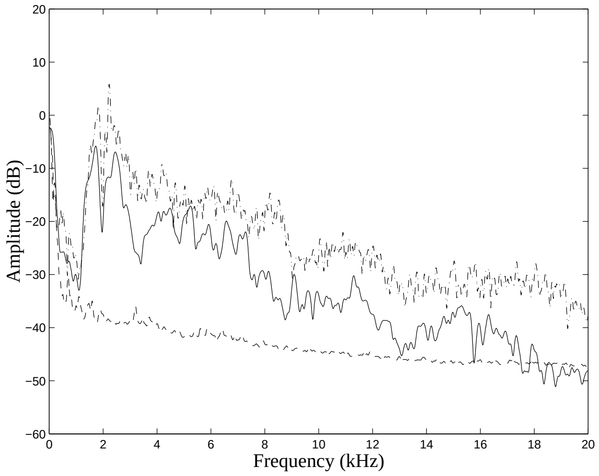

Figure 2: Averaged power spectra of fricative /ʃ/ from Corpora 1 (solid line) and 3 (dash-dotted line). The dashed curve is the averaged spectrum of the room noise.

REFERENCES

[1] Andrade, A. (**1994**). "Reflexões Sobre o 'E Mudo' em Português Europeu," Cong. Int. Sobre o Português, Lisboa, Portugal, **2**, 303-344.

[2] Andrade, A. (**1995**). "Percepção de C ou CC Oclusivas por Ouvintes Nativos de Português Europeu," XI Enc. Nac. da APL, Lisboa, **3**, 153-186.

[3] Lacerda, F. (**1982**). "Acoustic Perceptual Study of the Portuguese Voiceless Fricatives," J. Phonetics, **10**, 11-22.

[4] Martins, M. and Harmegnies, B. and Poch, D. (**1995**). "Changement Phonétique en Cours du Portugais Européen," XI Enc. Nac. da APL, Lisboa, **3**, 249-259.

[5] Mateus, M. (**1975**). "Aspectos da Fonologia Portuguesa," Centro de Estudos Filológicos, Lisboa.

[6] Pirello, K. and Blumstein, S. and Kurowski, K. (**1997**). "The Characteristics of Voicing in Syllable-Initial Fricatives in American English," JASA, **101**, 3754-3765.

[7] Shadle, C. and Scully, C. (**1995**). "An Articulatory-Acoustic-Aerodynamic Analysis of [s] in VCV Sequences," J. Phonetics, **23**, 53-66.

[8] Smith, C. (**1997**). "The Devoicing of /z/ in American English: Effects of Local and Prosodic Context," J. Phonetics, **25**, 471-500.

[9] Viana, M. (**1984**). "Etude de Deux Aspects du Consonantisme du Portugais: Fricatisation et Dévoisement," Doct., U. Sciences Humaines de Strasbourg, France.