UNIVERSITY OF SOUTHAMPTON

Acoustic Phonetics of European Portuguese Fricative Consonants

Luis Miguel Teixeira de Jesus

A Thesis Submitted for the Degree of
Doctor of Philosophy

to the
Faculty of Engineering and Applied Science
Department of Electronics and Computer Science

June 2001

Supervised by Dr. Christine H. Shadle
Examined by Dr. Mark A. Huckvale and Dr. Robert I. Damper

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE

Department of Electronics and Computer Science

Doctor of Philosophy

ACOUSTIC PHONETICS OF EUROPEAN PORTUGUESE FRICATIVE CONSONANTS

by Luis Miguel Teixeira de Jesus

The production of fricatives is not yet fully understood because the mechanism is particularly complex. Studies of Portuguese fricatives have been very limited, so in this thesis a novel methodology of corpus design, and temporal and spectral analysis techniques were developed to enhance our description of the acoustic properties, and to increase our understanding of the production of fricatives. The data presented in this thesis could be used to improve the naturalness of synthetic speech.

Corpora were devised that included the fricatives /f, v, s, z, ʃ, ʒ/ in the following contexts: sustained, repeated nonsense words of the form /pV_1CV_2/, Portuguese words containing fricatives in frame sentences, and the same set of words in sentences. Four subjects (two male, two female) were recorded saying the corpora, using a microphone in the acoustic far-field and a laryngograph. Temporal analysis of the fricatives revealed a large number of devoiced examples. Analysis of variance showed that devoicing was significantly more likely for word-final fricatives and posterior place of articulation.

In addition to the fricatives listed above, we also noticed other fricatives occurring as allophones of /ʀ, ɾ/ in 100 words out of 365. Durations of the fricative segments were comparable to /ʀ, ɾ/ and thus shorter on average than fricatives /f, v, s, z, ʃ, ʒ/. Some of the speech segments were continuous "noisy signals" very similar to those of fricatives. The spectral peak frequencies of the fricatives occurring in place of /ʀ/ were compared to the other fricatives, which indicated a place of articulation further back than /ʃ, ʒ/, and compared to velar and uvular fricative results previously reported for other languages. These comparisons indicated that the uvular fricatives [χ, ʁ] and the voiceless tapped alveolar [ɾ̥] were given the phonological role of /ʀ/ and /ɾ/ respectively, though these fricatives have not previously been reported as phones of standard European Portuguese.

The fricative spectra were parameterised in terms of our knowledge of the underlying aeroacoustics. The parameters spectral slope, frequency of maximum amplitude, and dynamic amplitude were developed to characterise fricative spectra. 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.

A preliminary cross - language study of Portuguese and English fricatives produced by two bilingual siblings is also presented. Although results for Portuguese and English fricatives seem to be very similar this maybe due to the use by bilinguals of different production strategies from monolinguals which attenuate cross - language acoustical contrasts. The English corpus developed for the bilingual subjects could be used to study monolingual English speakers.

Contents

List of Figures

List of Tables

The Kathakali Man is the most beautiful of men. Because his body is his soul. His only instrument.

in “The God of Small Things” by Arundhati Roy. Flamingo, 1997, p. 230.

Fonte 2

No sorriso louco das mães batem as leves
gotas de chuva. Nas amadas
caras loucas batem e batem
os dedos amarelos das candeias.
Que balouçam. Que são puras.
Gotas e candeias puras. E as mães
aproximam-se soprando os dedos frios.
Seu corpo move-se
pelo meio dos ossos filiais, pelos tendões
e órgãos mergulhados,
e as calmas mães intrínsecas sentam-se
nas cabeças filiais.
Sentam-se, e estão ali num silêncio demorado e apressado,
vendo tudo,
e queimando as imagens, alimentando as imagens,
enquanto o amor é cada vez mais forte.
E bate-lhes nas caras, o amor leve.
O amor feroz.
E as mães são cada vez mais belas.
Pensam os filhos que elas levitam.
Flores violentas batem nas suas pálpebras.
Elas respiram ao alto e em baixo. São
silenciosas.
E a sua cara está no meio das gotas particulares
da chuva,
em volta das candeias. No contínuo
escorrer dos filhos.
As mães são as mais altas coisas
que os filhos criam, porque se colocam
na combustão dos filhos, porque
os filhos estão como invasores dentes-de-leão
no terreno das mães.
E as mães são poços de petróleo nas palavras dos filhos,
e atiram-se, através deles, como jactos
para fora da terra.
E os filhos mergulham em escafandros no interior
de muitas águas,
e trazem as mães como polvos embrulhados nas mãos
e na agudeza de toda a sua vida.

E o filho senta-se com a sua mãe à cabeceira da mesa,
e através dele a mãe mexe aqui e ali,
nas chávenas e nos garfos.
E através da mãe of filho pensa
que nenhuma morte é possível e as águas
estão ligadas entre si
por meio da mão dele que toca a cara louca
da mãe que toca a mão pressentida do filho.
E por dentro do amor, até somente ser possível
amar tudo,
e ser possível tudo ser reencontrado por dentro do amor.

in "Poesia Toda" by Herberto Helder. Assírio e Alvim, 1996, p. 43.

Acknowledgements

I would like to thank my parents and brother for their love and tenderness.

I would like to thank my supervisor Dr. Christine H. Shadle for her support and encouragement.

Thank you Carlos, Cristina, Inês, Patrick and Rosemary for your friendship and support.

I would like to thank Doutora Lurdes de Castro Moutinho, at the Universidade de Aveiro in Portugal, for help in designing the corpus.

I also would like to give thanks to Dr. Robert I. Damper for his advice.

I would like to extend my gratitude to the ISIS Research Group for having contributed to a friendly environment where I have spent a challenging time.

This work was supported by Fundação para a Ciência e a Tecnologia, Portugal and the University of Southampton. I would like to thank both for making this work possible.

Para a minha avó Vitória que me ensinou o significado da palavra "Coragem".

Para o Doutor Freitas Gomes que me ensinou o significado da palavra "Amizade".

Chapter 1

Research Overview

1.1 Introduction

Portuguese is an important European language, spoken by over 180 million people worldwide. Studies of Portuguese phonetics and phonology indicate that fricatives are central to some interesting features of the language, yet studies of Portuguese fricatives have been few and limited. In this study, Portuguese fricatives were analysed in ways designed to enhance our description of the language, and to use and increase our understanding of the production of fricatives. The research presented in this thesis aims to investigate the acoustic features which characterise the production of fricative consonants. The production of fricatives is not yet fully understood because the mechanism is particularly complex. We will be focusing on the analysis of frication in the Portuguese language, describing a novel methodology of corpus design, and temporal and spectral analysis techniques. Knowledge accumulated from data could be used for improved speech synthesis.

1.2 Fricative Production Mechanisms

When a vowel is being uttered, the vocal tract is relatively unconstricted ($\sim$1 cm^2 cross-sectional area at the most constricted region) and the vocal folds vibrate periodically, causing the volume of air flowing through the glottis to fluctuate periodically as well. Fricative consonants are produced when the vocal tract is constricted ($\sim$0.1 cm^2 at most constricted region) somewhere along its length, as shown in Figure 1.1, enough to produce turbulence noise when air is forced through the constriction. The place of constriction affects the tract resonances (filter properties), but also affects the shape of the tract downstream of the constriction and thus the source properties: where the turbulent jet will impinge on tract walls, generating more noise, and the particular spectral characteristics of that noise.

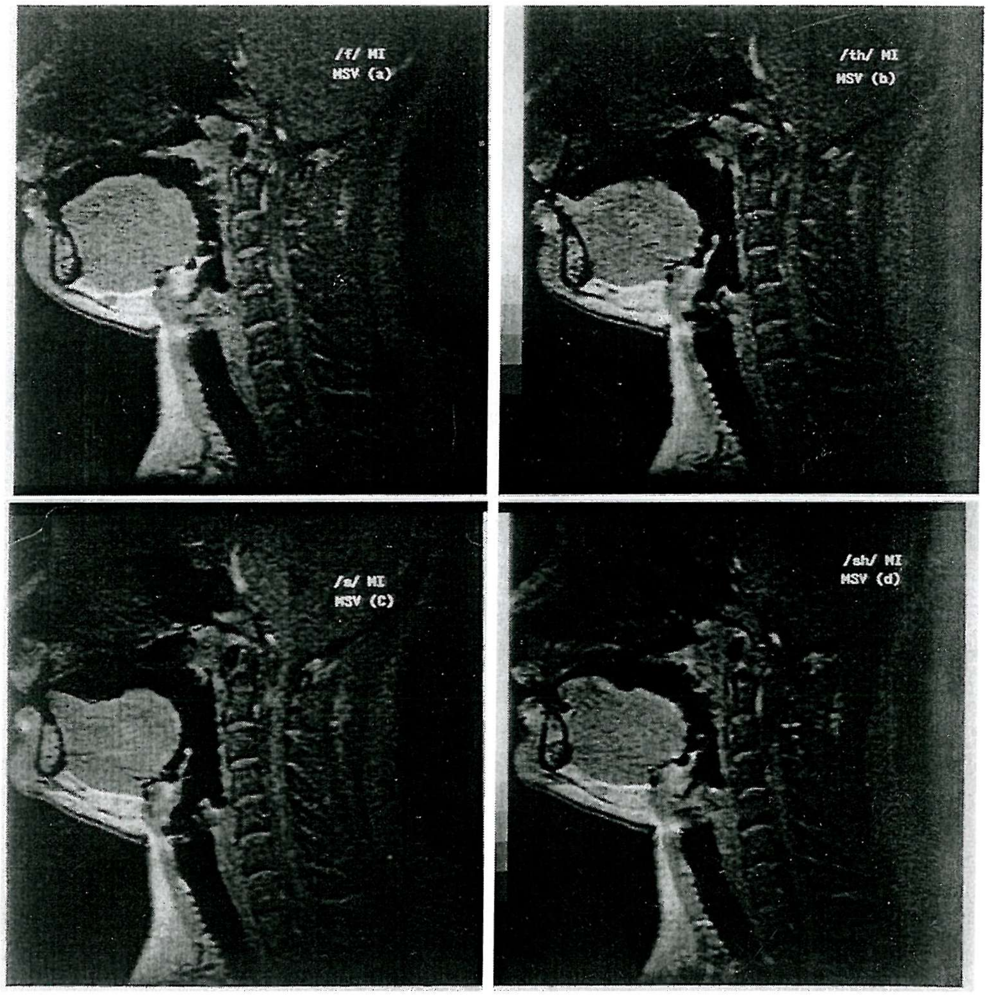

Figure 1.1: Mid-sagittal magnetic resonance imaging (MRI) profiles of the vocal tract during the production of voiceless fricatives: a) /f/, labiodental; b) /θ/ (/th/), dental; c) /s/, alveolar; d) /ʃ/ (/sh/), postalveolar. From Narayanan, Alwan, and Haker (1995).

It is known from studies of jet noise (Goldstein 1976) and mechanical models

(Shadle 1985) that when a particular configuration is held constant, and only the air velocity is increased, the turbulence noise increases (i.e. sound pressure and power), and increases more at higher frequencies. Though it is not easy to control nor measure parameters so precisely in the vocal tract, the same phenomenon appears to occur for fricatives (Hixon 1966; Hixon et al. 1967; Shadle 1985).

The acoustic mechanism for production of fricatives is thus not as well understood as for vowels because:

1. turbulence noise defies an analytic formulation, requiring empirical studies;
2. turbulence noise sources are much more sensitive to changes in the surrounding geometry than are acoustic resonances (Shadle 1991);
3. given the small constriction dimensions and the dependence of all aero-acoustic sources on flow velocities, it is much more difficult and more important to get sufficiently accurate vocal tract *shape* and simultaneous *aerodynamic* and *acoustic* data for fricative configurations.

These difficulties have been reflected in the relatively poor quality of fricative and affricate synthesis. Nevertheless, our understanding of fricative production has been improved by the use of existing expertise in the production of speech corpora, the extraction of magnetic resonance imaging (MRI) data (Narayanan et al. 1995; Shadle et al. 1996; Mohammad 1999; Engwall and Badin 2000), fricative aeroacoustics analysis methods (Shadle and Scully 1995), and the incorporation of three-dimensional vocal tract data in speech synthesis (Davies, McGowan, and Shadle 1993; Motoki, Badin, Pelorson, and Matsuzaki 2000; Motoki, Pelorson, Badin, and Matsuzaki 2000; Niikawa, Matsumura, Tachimura, and Wada 2000).

1.3 Previous Studies of Fricatives

The study of relations between articulatory, acoustic and perceptual cues (Hoole et al. 1989; Hoole et al. 1993; Trong and Hoole 1993; Trong et al. 1994; Stevens 1997) provides crucial information for the articulatory synthesis of fricative consonants (Scully 1979; Scully and Allwood 1985; Scully, Castelli, Brearley, and Shirt 1992). More specific studies of the articulation of fricatives include the palatographic experiments of Fletcher (1989) and Fletcher and Newman (1991), the extensive studies of tongue shapes by Stone et al. (1992) and Stone and Lundberg (1996), and the MRI and electropalatography experiments of Narayanan (1995) and Narayanan et al. (1995). The study of the nature of the interaction between acoustic sources and vocal tract shapes for constricted consonantal configurations (Stevens 1987; Stevens 1991; Badin 1991; Badin et al. 1994; Shadle 1995), and the study of mechanical models by Shadle (1985, 1990, 1991), has supplied important data to drive various parametric multi-tube acoustic models (Zagar 1986; Vescovi and Castelli 1995; Liu and Lacroix 1997; Riegelsberger 1997; Narayanan and Alwan 2000). See also the list of references for additional reading on subjects related to fricative consonants at the end of this thesis.

Further research is needed to determine specific acoustic, aerodynamic and articulatory attributes of fricatives. Analysis methods such as time averaging and ensemble averaging (Shadle, Dobelke, and Scully 1992; Shadle, Moulinier, Dobelke, and Scully 1992; Shadle, Mair, and Carter 1996), and studies that establish cavity affiliation (Shadle et al. 1991) and the effect of vowel context on the acoustic characteristics of fricatives (Shadle et al. 1995), have identified some parameters that might be useful for the analysis of Portuguese fricative consonants. Researchers have used spectral moments and locus equations on fricatives (Forrest et al. 1988; Sussman 1994; Jongman et al. 2000) without much success, although such techniques work well on stops. A different parameter set, described in Chapter 6, based more on the understanding of the acoustic mechanisms of fricative production, was developed and tested against a large corpus of fricatives. An initial study, based on an existing English and French fricative corpus, suggested some fruitful directions to pursue (Shadle and Mair 1996).

1.4 Analysis of Portuguese Fricatives

One of the first specific studies of Portuguese fricatives was that of Lacerda and Rogers (1939), which consisted of the analysis of aerodynamic and acoustic readings using very primitive methods. Johns (1972) observed, from a study of slips of the tongue, that because of the difficulty of motor coordination during the production of unvoiced Portuguese fricatives, place of articulation was often incorrectly executed. A study by Martins (1975) produced a rank order of average duration and "intensity" of Portuguese fricatives. Unfortunately, the methodology used to measure amplitude is rather outdated and averaging of results inappropriate (as shown in Chapters 4 and 5).

Lacerda (1982) 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.

The acoustic and aerodynamic study of Portuguese consonant clusters of Andrade (1982, 1995) 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/, because this vowel is weakened, or even not produced, in final position.

Viana's study of Portuguese plosives (Viana 1984) also includes the analysis of fricative spectra. 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. (1995) observed that when the vowel V_1 in /V_1ʃpV_2/ and /V_1ʃtV_2/ sequences is weakened (or not produced), its formant structure is somewhat "transferred" (as designated by the authors) to the following fricative /ʃ/, allowing the listener to perceive the V_1 vocalic segment.

Andrade et al. (1999) studied the acoustical and perceptual effects of rounding in /s/ produced by a male speaker. They observed, in spectrograms of /sɨ/ and /su/, that the peak frequencies were shifted down and the overall amplitude diminished for the rounded vowel context. Results from perceptual tests of CVs generated by a formant synthesizer (Klatt 1980; Klatt and Klatt 1990) revealed F2 as a relevant cue for the distinction between /sɨ/ and /su/.

Regional variations of Portuguese (Cunha and Cintra 1992) 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 (Mateus 1996).

There have been many studies of the phonetics and phonology of Portuguese, which have shown some interesting features of the language; it is unusually rich in instances of vowel reduction, consonant clusters, and plosives that are realized as fricatives (Viana 1984). The study of Portuguese fricative consonants constitutes a challenging and complex research area, which is as yet incompletely explored. This is in part due to the lack of specific speech corpora that reflect the variety of phonetic contexts in which these speech sounds occur and the large number of variations in fluent Portuguese speech.

1.5 Thesis Overview

This study focusses 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. The corpus is described in Chapter 2. The segmentation and annotation of the recorded material is described in Chapter 3. A detailed description of time and frequency domain analysis methods is also included in Chapter 3. The results of this analysis are presented in Chapter 4 (temporal and devoicing analysis) and Chapter 5 (spectral analysis). Chapter 6 describes the parameterisation of the fricative spectra, done both to aid within- and across-speaker comparisons, and as a first step towards modelling and synthesis of the fricatives. A preliminary cross-language study of Portuguese and English fricatives produced by two bilingual siblings is described in Chapter 7. Conclusions and future work are presented in Chapter 8.

Chapter 2

Design and Recording of a Corpus of Portuguese Fricatives

2.1 Introduction

A speech corpus has been designed to explore the fricatives of standard European Portuguese. The phonetic and phonological evidence underlying the design of the corpus are described in the sections that follow. The complete corpus is described in Appendices A and B. We used methodology of previous fricative studies, begun with the EC SCIENCE "Fricative" Project, conducted by Shadle et al. (Shadle 1992; Shadle and Carter 1993). That study was focused on characterizing fricatives in general. Here, that methodology has been adapted to focus on Portuguese fricatives in particular, and thus uses real words and phonology of Portuguese.

2.2 Design

A rich variety of phonetic contexts using both real Portuguese words and nonsense words was selected to study the most relevant phoneme variants, and describe the spectral and articulatory characteristics of Portuguese fricative consonants. The corpora also included sustained fricatives, which are better controlled (no phenomena such as coarticulation or devoicing occur during the production of sustained fricatives) and easier to analyse than those in words.

Fricatives that were produced more naturally, but still with contextual and stress control, were studied using a corpus of nonsense words. To produce examples that would be phonotactically possible words in Portuguese, the nonsense words all followed these generally accepted (for European Portuguese) language-specific phonological rules (Mateus and Andrade 2000, p. 11):

1. any of the vowels /i, e, ɛ, ɐ, a, ɔ, o, u, ĩ, ẽ, ɐ̃, õ, ũ/ can occur in the tonic syllable [1];
2. any of the vowels /i, ɨ, e, ɛ, ɐ, a, ɔ, o, u, ĩ, ẽ, ɐ̃, õ, ũ/ can occur before the tonic syllable;
3. only vowels /i, ɨ, ɐ, u/ can occur after the tonic syllable;
4. the vowel /i/ does not appear in final position;
5. the fricatives /f, v, s, z, ʃ, ʒ/ can all occur in initial and medial positions;
6. /ʃ/ is the only fricative that can occur in word-final position.

In addition to these constraints and to facilitate comparisons, the corpora were designed to be compatible where possible with the fricative corpora recorded of English, American, French and German subjects (Shadle 1992; Shadle and Carter 1993).

2.2.1 Corpus 1: Sustained Fricatives

Corpus 1a consisted of a set of VCV sequences, where V belongs to the reduced set of Portuguese vowels /ɨ, ɐ, u/, and C is one of the Portuguese fricative consonants /f, v, s, z, ʃ, ʒ/ sustained for 5 s (see Appendix B). As shown by Shadle et al. (1996), the vowel context, even for sustained examples, influences the articulatory and spectral characteristics of fricatives. Since the vocalic contexts of Corpus 1a overlap with those of Corpus 3 (set

[1] Tonic syllable – syllable which carries maximal prominence, usually owing to a major pitch change (Crystal 1997).

of Portuguese words), it is possible to make a comparative study between the fricatives produced within these two experimental conditions.

A separate set of Portuguese fricative consonants, sustained for 3 s, at medium, soft and loud effort levels, was also recorded (and is called Corpus 1b). Ideally we would like the articulation to be held constant, and only the mean flow velocity at the constriction during its production to be varied. We attempt to elicit this by asking for a variation in effort level.

2.2.2 Corpus 2: Nonsense Words

Corpus 2 (see Appendix B) consisted of /pV_1CV_2/ sequences, where V_1, V_2 were one of the vowels /i, ɐ, u/. The set comprised all possible vowel and fricative permutations, each repeated about 12 times in one breath. The phoneme /p/ is an easily identifiable marker for segmentation and spectral analysis, and has been used in Rothenberg mask recordings by Shadle et al. (Shadle 1992; Shadle and Carter 1993) to measure the subglottal pressure and to check where the zero is in the recorded time signal (no flow velocity).

The stress was placed according to language - specific phonological rules, and subjects were instructed to keep it the same through all the repetitions. The subjects were not always able either to produce the indicated stress pattern or to produce a different pattern consistently, so there were some instances with equal stress in both syllables, and with deleted vowels.

2.2.3 Corpus 3: Real Words

Corpus 3 consisted of 154 words, each said within the frame sentence *Diga ..., por favor* /'digɐ ... puɾ 'fɐvoɾ/, which was used to record the words in the corpus in a balanced phonetic context and with a neutral prosody. The words, listed in Appendix A, were presented in a randomised order.

The 154 words consist of 8 words forming nearly minimal pairs with the pattern /FV_1FV_2/; 54 words with the pattern /FV_/ (fricative in initial position); 69 words with the pattern /_V_1FV_2_/ (fricative in medial position); and 23 words with the pattern /_VF/ (fricative in final position).

The vowels in words with sequences /FV_1FV_2/, /FV/, /V_1FV_2/ and /VF/ have been divided into three groups according to their location in the vowel triangle: /i, ɨ, e/ – group 1; /ɛ, ɐ, a/ – group 2; /ɔ, o, u/ – group 3. Appendix A lists examples with nearly all Portuguese non - nasal vowels preceding each of the fricatives, followed by one vowel from each of the vowel groups.

The vowel /ɨ/ is generally deleted in final position, as shown by Andrade (1994), and so the resulting allophone is not expected to influence the preceding fricative. Therefore words such as *chefe* /'ʃɛfɨ/, *ave* /'avɨ/ and *asse* /'asɨ/, were used to 'simulate' final position contexts. As mentioned by Mateus and Andrade (2000), phonologically, only /ʃ/ can occur in final position, but phonetically any fricative can be found in final word position as a consequence of deletion of unstressed vowels. Appendix A lists examples of Portuguese words with fricative consonants in final position.

2.2.4 Corpus 4: Real Words in Connected Speech

Corpus 4 consisted of a set of sentences (see Appendix A) including 60 words from Corpus 3. Ten of the sentences are meaningful; two include word boundaries within some of the phonetic sequences in Corpus 3, but are semantically nonsense.

2.3 Recording Method

The subjects used in this study were two male (LMTJ and CFGA) and two female (ACC and ISSS) adult Portuguese native speakers, with no reported history of hearing or speech disorders. Subject LMTJ, age 26, is from the city of Aveiro (at the centre of Portugal), and CFGA, age 26, is from Braga (in north Portugal). Speaker ACC, age 33, is from Sintra (a city very close to Lisbon), and ISSS, age 21, is from Lisbon. At the time of the recordings all subjects had been studying in England for a period of two to three years.

Recordings were made in a sound treated room using a Bruel & Kjaer 4165 $\frac{1}{2}$ inch microphone located 1 m in front of the subject's mouth, connected to a Bruel & Kjaer 2639 pre - amplifier. The signal was amplified and filtered by a Bruel & Kjaer 2636 measurement amplifier, with high - pass cut - on frequency of 22 Hz and low-pass cut-off frequency of 22 kHz. A laryngograph signal (Lx) was also collected using a laryngograph processor [2]. The acoustic speech signal and Lx were recorded with a Sony TCD-D7 DAT system at 16 bits, with a sampling frequency of 48 kHz, and digitally transferred to a computer for post-processing. A 94 dB, 1000 Hz calibration tone produced by a Bruel & Kjaer 4620 calibrator was also recorded on the same tape on which speech was recorded (see Appendix C).

2.4 Summary

In this chapter, the design of the speech corpus and the underlying phonological rules that determined the selection of Portuguese words and nonsense words have been described. Details of a corpus of sustained fricatives and a corpus of Portuguese sentences have also been presented. This chapter has also provided a description of the recording apparatus and techniques. Chapter 3 describes the segmentation and annotation, and presents the various methods used to analyse the speech corpora.

[2]Model LxProc, type PCLX produced by Laryngograph Ltd (UK).

Chapter 3

Methods for Segmentation, Annotation, and Analysis

3.1 Introduction

This chapter describes the method used to segment and annotate the various corpora presented previously. This is followed by an extensive definition of both a manual method and an automatic method to determine if a fricative is devoiced. Finally we describe the three techniques used to average the power spectra of the fricatives: time-averaging, frequency-averaging and ensemble-averaging.

3.2 Segmentation and Annotation

The data on the DAT tape were digitally transferred to `.wav` computer files, which contain the acoustic speech signal in the right channel and the laryngograph signal on the left channel, recorded at 16 bit, with a sampling frequency of 48 kHz.

The time waveforms of all the corpus words were manually analysed to detect the start of the vowel-fricative transition, the start of the fricative, the end of the fricative, and the start of the fricative-vowel transition. During the vowel - fricative transition, there is a decrease in amplitude, voicing ceases (for unvoiced fricatives) and frication noise starts, as shown in Figure 3.1. During the fricative - vowel transition, there is an increase in amplitude, voicing starts (for unvoiced fricatives) and frication noise ceases (Docherty 1992, pp. 118 - 119). These events do not occur simultaneously or always in the same order, making the segmentation a somewhat subjective process. However, it is important to segment consistently, because the results of the analysis methods depend on where the boundaries are placed (Docherty 1992, pp. 103 - 110). The amplitude and voicing changes appear in both acoustic and Lx signals, which aids the segmentation process. For example, as can be seen in Figure 3.1, the FV transition also includes some frication noise because we've established that an unvoiced fricative would only correspond to a steady - state noise segment.

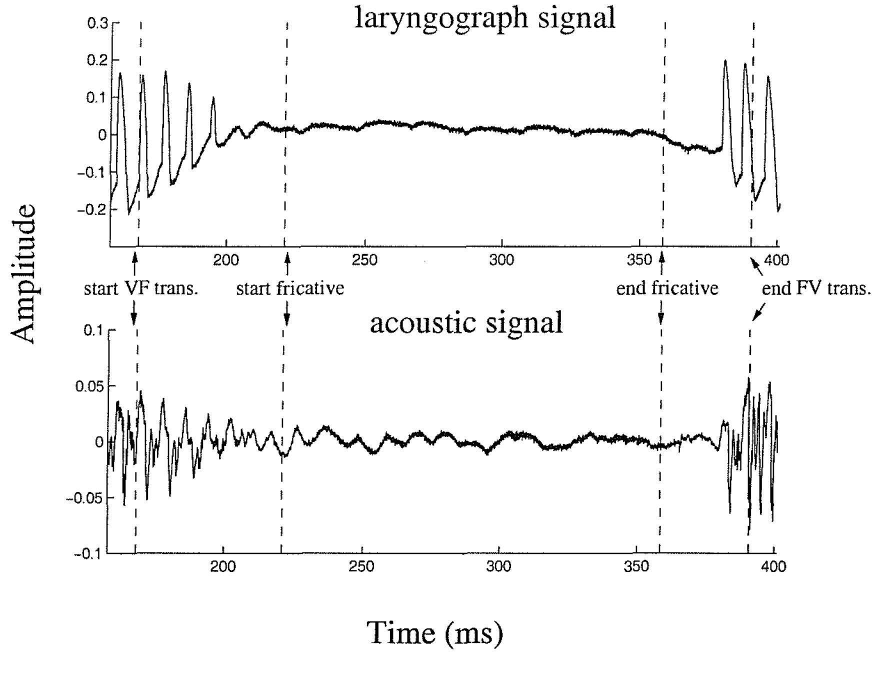

Figure 3.1: Laryngograph signal and acoustic signal of fricative /f/ in *café* /kɐˈfɛ/, showing the start of the vowel - fricative transition, the start of the fricative, the end of the fricative, and the end of the fricative - vowel transition. Corpus 3 (Speaker LMTJ).

The laryngograph signal was also used in the decision process to determine the VF and FV boundaries (see Figure 3.1). For unreduced vowels there was always significant voicing, and for the duration of most fricatives the laryngograph signal changed drastically. Therefore, the amplitude of the laryngograph signal was an important cue in determining the boundaries between the different phones. When it was not clear from the acoustic signal where the fricative started and ended (especially for voiced fricatives), the laryngograph signal was used as an additional cue, because its amplitude diminishes during the VF transition and increases during the FV transition.

The annotation files generated for Corpus 3, which have been used by various analysis programs, consist of eight sample numbers referring to the following locations within the corpus word:

1. start of first vowel - fricative transition;
2. start of first fricative;
3. end of first fricative (or start of first fricative - vowel transition);
4. end of first fricative - vowel transition;
5. start of second vowel - fricative transition;
6. start of second fricative;
7. end of second fricative (or start of second fricative - vowel transition);
8. end of second fricative - vowel transition.

For corpus words with only one fricative (e.g. *fala* /ˈfalɐ/), values 5 through 8 are set to zero.

In examples such as *este* /ˈeʃtɨ/, where we have a vowel - fricative - plosive segment, the fourth annotated value corresponds to the end of first fricative - plosive transition. When the words contain a final fricative, the fourth annotated value has the same sample value as the third, or the fourth annotated value corresponds to a marker in the "silence" that follows the fricative.

The four speakers produced, on average, more than 12 repetitions of each nonsense word in Corpus 2. However, the first and last, and any atypical tokens were eliminated, thus resulting in ensembles of nine tokens each (see Section 3.5.2).

We have also created a set of files containing a phonetic transcription, according to the International Phonetic Alphabet (IPA 1999), of all recorded speech material.

3.3 Devoicing

In the word corpora, there were large amounts of devoicing. When the vocal tract is constricted for a voiced fricative, voicing is often maintained only over part of the fricative, because if strong voicing was to be maintained then the amplitude of the noise source would become small, whereas a strong noise source can only be achieved at the expense of weakened voicing or cessation of voicing (Stevens 1987, p. 388). When voiced fricatives devoice, it is with a whisper phonation (Abercrombie 1967, p. 137), distinguishing them from their voiceless counterparts which are realised with a glottal abduction gesture. Smith (1997) also suggested that the glottis is in a state intermediate between voicing and voicelessness, like the state of the glottis that is used in whisper, with the glottis open but the folds very close together. The signal shown in Figure 3.2 is one such case. It corresponds to a segment that starts at the onset of the VF transition, and ends at offset of the FV transition.

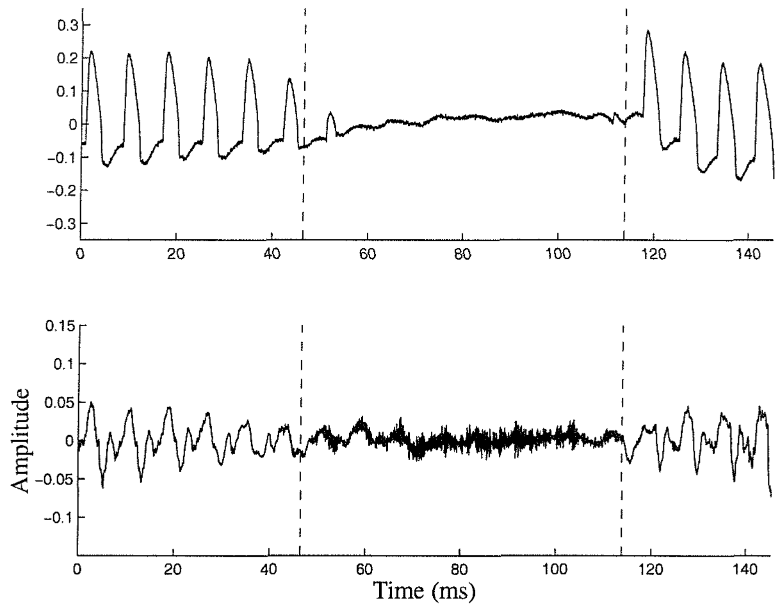

Figure 3.2: Laryngograph signal and acoustic signal of fricative /z/ in *diga zarpar* /ˈdigɐ zɐrˈpar/. The dashed lines mark the start and end of the fricative. Corpus 3 (Speaker LMTJ).

Haggard (1978) defined devoicing as presence of measurable friction in the absence of continued glottal vibration, i.e., the periodic component of the voiced fricative ceased before the friction component.

Smith (1997, p. 478) 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.

She described three processes which may result in devoicing: an assimilatory process, in which the position of the vocal folds is more open (a change in articulation); a lenition process, in which the transglottal pressure drop or the volume velocity of the airflow across the glottis is insufficient to maintain vocal fold vibration (a change in aerodynamics); and a change in the aerodynamic conditions caused by the articulatory movement of the supralaryngeal constriction.

A study by Pirello et al. (1997, p. 3756) 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.

Both the acoustic signal and the laryngograph signal were used to determine if a fricative was devoiced. A fricative was called *devoiced* when less than one-third of the frication interval showed periodic structure in the acoustic or laryngograph signals. The term *partially devoiced* was used when more than one-third but less than half of the frication interval contained steady acoustic and laryngograph signal cycles. A fricative was called *voiced* when more than half of the frication interval showed steady acoustic and laryngograph signal cycles, even if the amplitude was much lower than in the vowel (Docherty 1992, p. 13). If the laryngograph signal was clearly periodic, the interval was classified as voiced; if the laryngograph signal was zero or distorted, the signal was classified as voiced only if the acoustic signal was unambiguously periodic.

In Chapter 4 we present an inventory of all cases of devoicing, a detailed analysis of the devoicing patterns found, and try to identify the factors (such as vowel context, word position, etc.) influencing this phenomenon.

3.3.1 Automatic Criterion for Devoicing

As pointed out by Docherty (1992, p. 102), a number of techniques for automatically detecting whether a portion of a signal is voiced or not has been used in the past, but proved to be unsuitable for fricatives because in this class of speech sounds voicing has low energy. Therefore a new criterion, based on the laryngograph signal, was tested for the corpora used in the present study. The sample mean

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.1}$$

and sample variance

$$\sigma^2(x) = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{3.2}$$

of the laryngograph signal x_N were calculated during the VF transition and during the fricative. The ratio of variances of the two intervals,

$$r_{\sigma^2(x)} = \frac{\sigma_t^2(x)}{\sigma_f^2(x)}, \tag{3.3}$$

where $\sigma_t^2(x)$ is the variance of the signal during the VF transition and $\sigma_f^2(x)$ is the variance of the signal during the fricative, was used as an automatic criterion for devoicing. Obviously the ratio gets bigger if the laryngograph signal during the fricative gets really small relative to the transition. A heuristic threshold of 15 was used: for $r_{\sigma^2(x)} \geq 15$, the fricative is labelled *devoiced*; if $r_{\sigma^2(x)} < 15$, *voiced*. Fricatives manually classified as *partially voiced* were considered to be in the *devoiced* category when comparing the manual and automatic criteria.

The laryngograph signal presents, in some voiced fricative examples and in most unvoiced fricative examples, a slowly increasing or decreasing amplitude over the frication interval, which results in a large variance, and therefore a misclassification as voiced. This problem has been solved using an averaged $r_{\sigma^2(x)}$. We have computed the mean $\overline{x}$ and the variance $\sigma^2(x)$ for three consecutive equal length sections of the frication interval, calculated the average frication interval variance, and used it to compute a new ratio of variances. We have tried to use a larger number of sections over which we calculate the averaged $r_{\sigma^2(x)}$ but this does not improve significantly the efficiency of this measure of devoicing.

3.3.2 Vowel Reduction

Another striking feature of the corpus is the large number of highly reduced vowels, which are often also devoiced. Figure 3.3 shows an example of a reduced /u/, and Figure 3.4 shows a reduced /i/. There are different patterns of reduction (reduced throughout, partially reduced, ...) depending on the phonetic context. Vowel reduction results in a larger number of word-final fricatives and consonant clusters with a fricative, both factors contributing to devoicing of the fricative, especially if the phoneme preceding or following the fricative is devoiced (Hogan and Rozsypal 1980; Stevens et al. 1992; Pirello et al. 1997; Smith 1997).

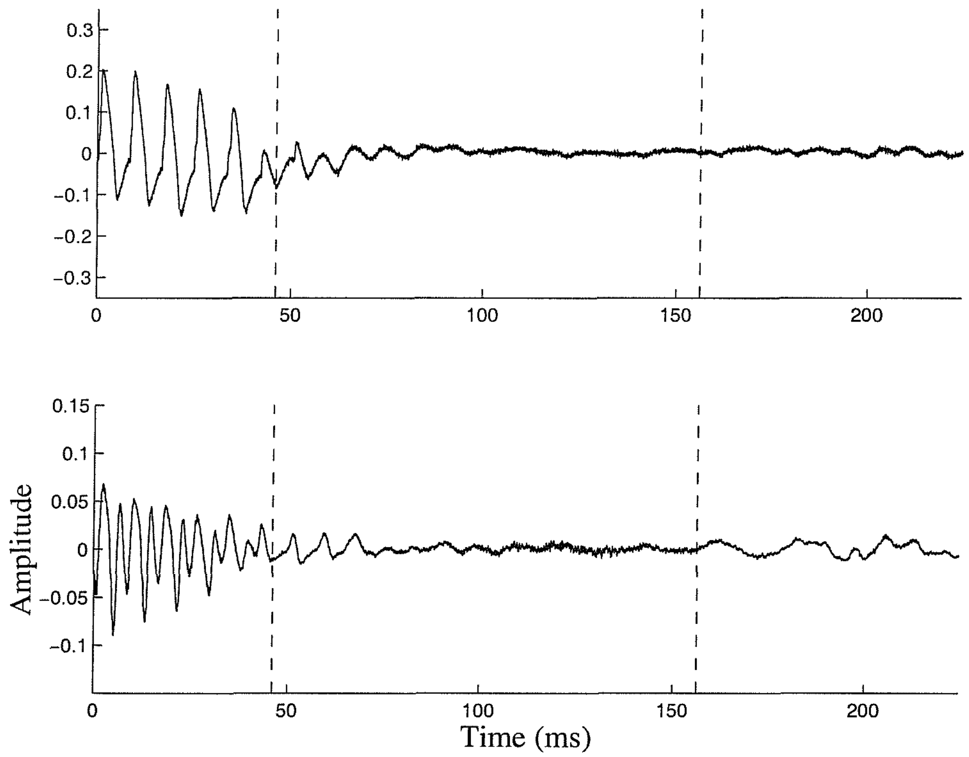

Figure 3.3: Laryngograph signal and acoustic signal of /iv/ transition, fricative /v/ and /vu/ transition where /u/ is reduced (from the word *altivo* [al'tiv]). The dashed lines mark the start and end of the fricative. Corpus 3 (Speaker LMTJ).

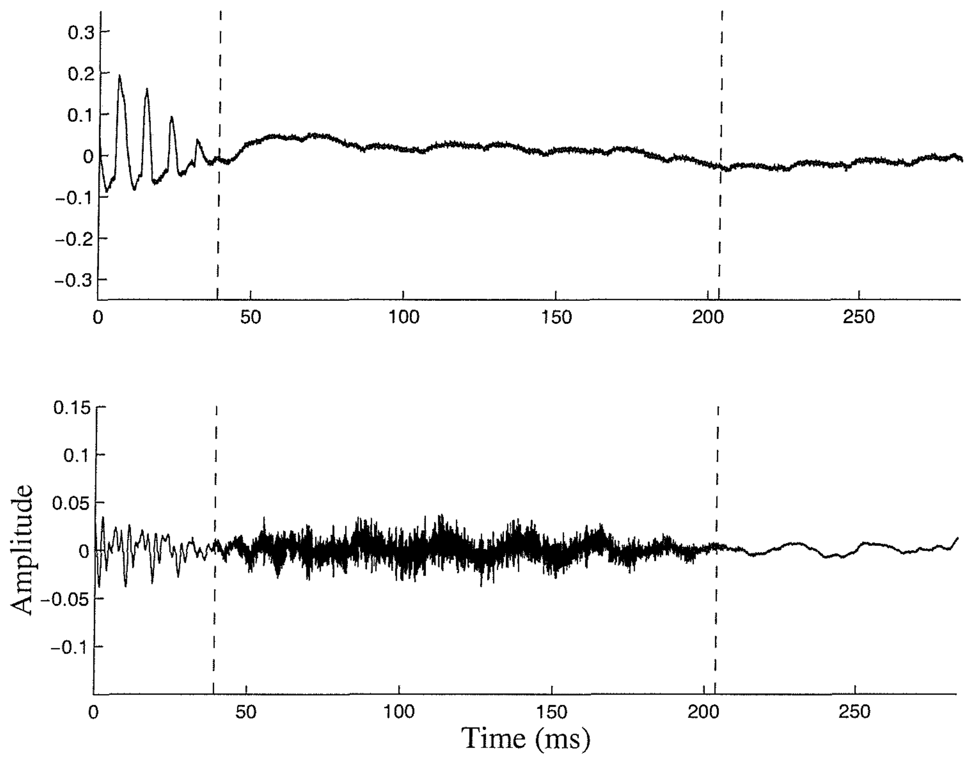

Figure 3.4: Laryngograph signal and acoustic signal of /es/ transition, fricative /s/ and /sɨ/ transition where /ɨ/ is reduced (from the word *pêssego* [ˈpesgu]). The dashed lines mark the start and end of the fricative. Corpus 3 (Speaker LMTJ).

3.4 Uvular Fricatives and Voiceless Tapped Alveolar Fricatives

The voiced uvular fricative [ʁ] has been referred to in a book by Mateus and Andrade (2000) as forming part of the phonetic repertoire of European Portuguese, manifested in a rhotic phonological role (of the uvular trill /ʀ/). The voiceless uvular fricative [χ] is mentioned in the same book as a phone in Brazilian Portuguese. Ladefoged and Maddieson (1996, pp. 166-167) reported uvular fricatives in a number of different languages, and assumed these had similar "vocal tract shape" as uvular stops. For an extensive articulatory study of pharyngeal consonants see the work of Esling (1996).

To our knowledge, there is no reference to the voiceless tapped alveolar fricative [ɾ̥] in any study of European Portuguese. Voiceless tapped alveolar fricatives are described by Laver (1994, p. 263) as follows:

> A tapped fricative is made by a swift movement of the active articulator towards the passive articulator, but where the maximum degree of stricture reached is that of close approximation rather than complete closure.

This constriction of the vocal tract lasts for a very short time and results in a low amplitude burst of frication noise. Laver (1994) mentions a Nigerian language, Etsako, that uses a tense voiceless tapped alveolar fricative. Ladefoged and Maddieson (1996, pp. 232 - 242) also reported several examples of "fricative rhotics" in various world languages (e.g. English, French, Czech, Edo). Solé et al. (1998) analysed the variation, impairment and extinction of voiced and voiceless trills as a function of intraoral pressure and airflow. Results revealed that the intraoral pressure and airflow conditions for voiced trills and fricatives are very similar. As the intraoral pressure dropped below a certain threshold trilling ceased, resulting in a fricative.

During the annotation phase, we noticed that 100 words contained a second fricative, besides the one initially selected for analysis. These included 21 examples of [χ], similar to that shown in Figure 3.5, and two of its voiced counterpart [ʁ].

Some of the speech segments ("noisy signals" very similar to those of fricatives) were classified as voiceless tapped alveolar fricatives [ɾ̥], because for most of the examples in the corpus they were given the rhotic /ɾ/ phonological role. However, they also took the phonological role of an uvular trill /ʀ/, as can be seen in Tables 3.1 and 3.2.

Not all of the examples from Corpus 3 and 4 were reported in Tables E.1 to E.3 and used for analysis because some of the fricatives were part of a consonant cluster where it was impossible to determine precise boundaries for each phoneme. Also from the phonetic transcriptions of Tables E.3 to E.6 we can see that the speakers sometimes produced the close vowel [ɨ] following or preceding [ɾ̥], and that there was one tap [ɾ] followed by [ɾ̥] and tapped fricative [ɾ̥] - tap [ɾ] clusters.

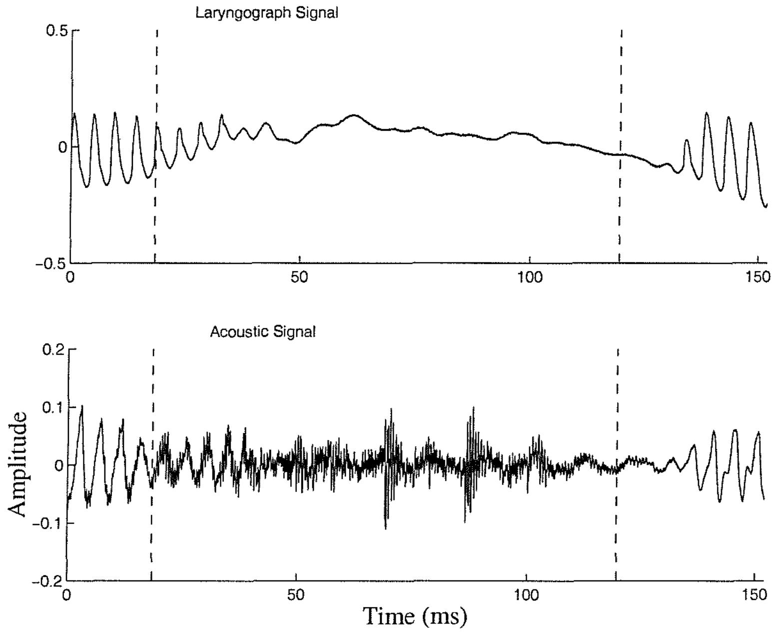

Figure 3.5. Laryngograph signal and acoustic signal of fricative [χ] in *relevo* [ɐ χiˈlev] (preceded by the word *diga* [ˈdigɐ]). The dashed lines mark the start and end of the fricative. Corpus 3 (Speaker ISSS).

Table 3.1: Number of occurrences of phones [ʀ], [χ] and [ʁ], and their particular phonological role.

Corpus	Speaker	[ʀ]	[χ]	[ʁ]
3	LMTJ	4	-	-
3	CFGA	1	5 – /ʀ/ phon. role	1 – /ʀ/ phon. role 1 – /ɾ/ phon. role
3	ACC	1	1 – /ʀ/ phon. role	1 – /ʀ/ phon. role
3	ISSS	-	6 – /ʀ/ phon. role	-
4	LMTJ	-	-	-
4	CFGA	-	5 – /ʀ/ phon. role	1 – /ʀ/ phon. role
4	ACC	-	-	-
4	ISSS	-	4 – /ʀ/ phon. role	-

Table 3.2: Number of occurrences of phones [ɾ] and [ɾ̥], and their particular phonological role.

Corpus	Speaker	[ɾ]	[ɾ̥]
3	LMTJ	25	26 – /ɾ/ phon. role 2 – /ʀ/ phon. role
3	CFGA	50	-
3	ACC	30	22 – /ɾ/ phon. role 2 – /ʀ/ phon. role
3	ISSS	42	8 – /ɾ/ phon. role
4	LMTJ	28	9 – /ɾ/ phon. role 4 – /ʀ/ phon. role
4	CFGA	61	2 – /ɾ/ phon. role
4	ACC	51 – /ɾ/ phon. role 3 – /ʀ/ phon. role	9 – /ɾ/ phon. role 3 – /ʀ/ phon. role
4	ISSS	37	6 – /ɾ/ phon. role

3.5 Spectral Analysis of Labiodental, Alveolar and Postalveolar Fricatives

Stochastic signals require some form of averaging for their spectra to be both consistent and low - error estimates of the underlying distribution (Bendat and Piersol 2000, pp. 423 - 442). For a stationary signal, time - averaging can be used; for nonstationary signals where an ensemble exists, ensemble - averaging can be used; or a single spectrum can be smoothed, and the averaging achieved at the expense of frequency resolution. We varied the method according to corpus. In Corpus 2, where multiple tokens of the nonsense words existed, ensemble - averaging was used; time - averaging was used elsewhere (Corpus 1a, 1b, 3 and 4), as shown in Figure 3.6.

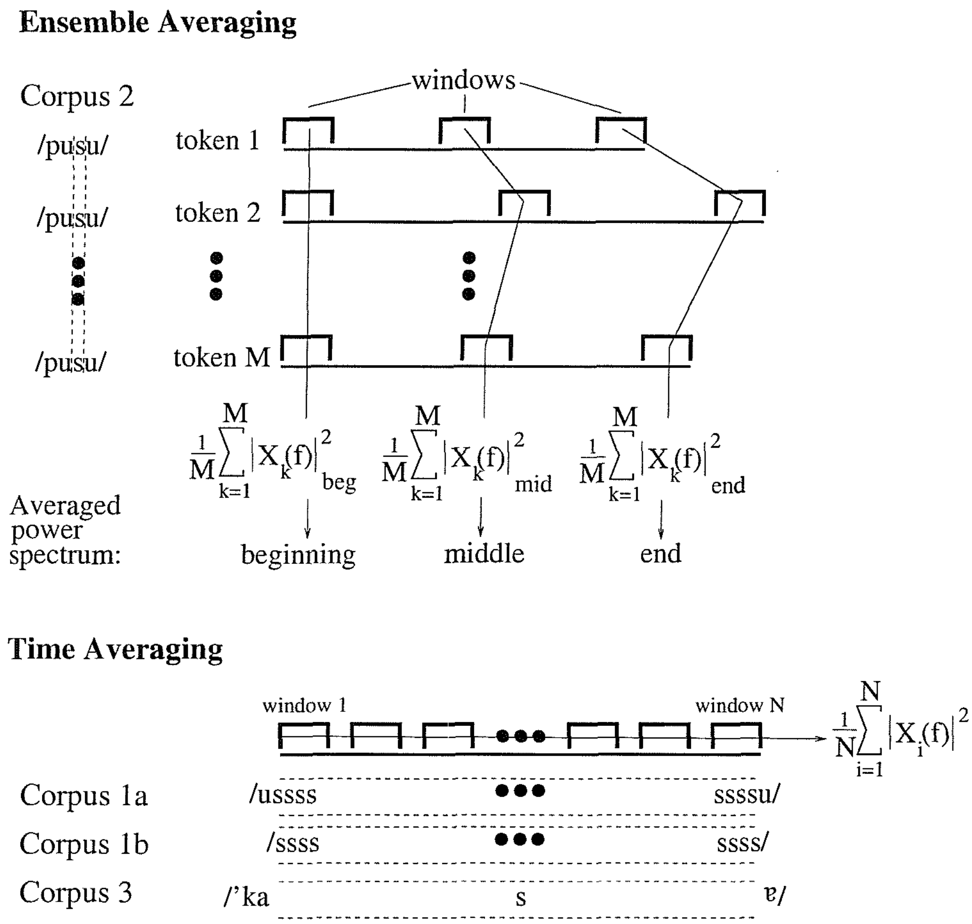

Figure 3.6: Diagram showing how windows are placed within fricative tokens, and how the corresponding power spectra $X_k(f)$ are combined to compute ensemble and time averaged spectra.

3.5.1 Time - Averaged Spectra

The first phase of spectral analysis consisted of a study of the averaged Discrete Fourier Transform (DFT) spectra, used to see the broad characteristics of the fricatives. The duration of the windows (10 ms both for time and ensemble - averaging) was chosen so that a reasonable number of windows could be used to cover adequately the wide range of fricative durations in the corpora (from 40 ms to 200 ms). For the shorter fricatives the windows overlapped. We used nine windows to calculate the averaged spectra, because observation of Corpus 2 indicated that there were always at least nine valid repetitions of the nonsense words. This allowed us to compare the spectra of fricatives calculated from real Portuguese words (Corpus 3 and 4) with the ensemble-averaged spectra of nonsense words.

A window's placement was related to the proportion of the distance through the fricative interval. Thus, regardless of the fricative length, the segments used to calculate the averaged spectra were always placed in a time position that corresponded to the same speech event. This allowed us to compare the spectra of short and long fricatives.

We used time-averaging with nine 10 ms Hamming windows, one left-aligned to the start of the fricative, one right-aligned to the end of the fricative, one centered at the middle of the fricative, and the rest evenly distributed in between (centered at $\frac{1}{8}$, $\frac{1}{4}$, $\frac{3}{8}$, $\frac{5}{8}$, $\frac{3}{4}$ and $\frac{7}{8}$ times the total length of the fricative). This meant that the amount of overlap varied according to fricative duration. The longest fricatives ($\sim$ 200 ms) had zero overlap, as shown in Figure 3.7; the shortest ($\sim$ 40 ms) had approximately 60% overlap, as shown in Figure 3.8, between successive windows. The time-averaged power spectrum for each fricative is given by

$$P_T(f) = \frac{1}{N}\sum_{i=1}^{N}|X_i(f)|^2 \tag{3.4}$$

where X_i is the DFT of a portion of the fricative signal, x_i, corresponding to the ith windowed segment.

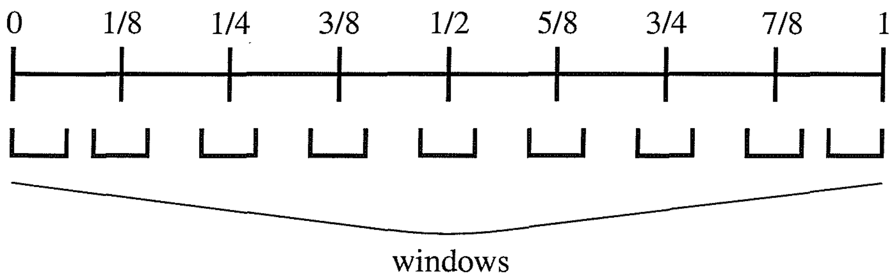

Figure 3.7: Placement of windows for the calculation of the time-averaged spectrum of a fricative (more than 120 ms).

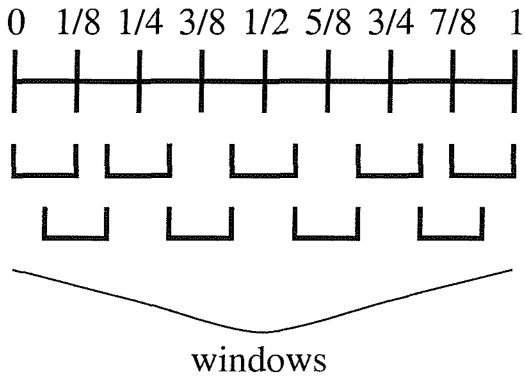

Figure 3.8: Placement of windows for the calculation of the time-averaged spectrum of a fricative (less than 120 ms).

The same time-averaging technique was used to calculate the spectra of sustained fricatives in Corpus 1a and Corpus 1b, but since the sustained fricatives were so long, we used $N = 100$ windows (each 10 ms long). A comparative study of all fricatives in different phonetic contexts was therefore possible. For example, the main spectral peak of fricative /s/ in Figure 3.9 is shifted down when compared with the spectra of the same fricative shown in Figure 3.10 (for medium effort level it is actually fairly similar), because of the rounded vowel /u/ context.

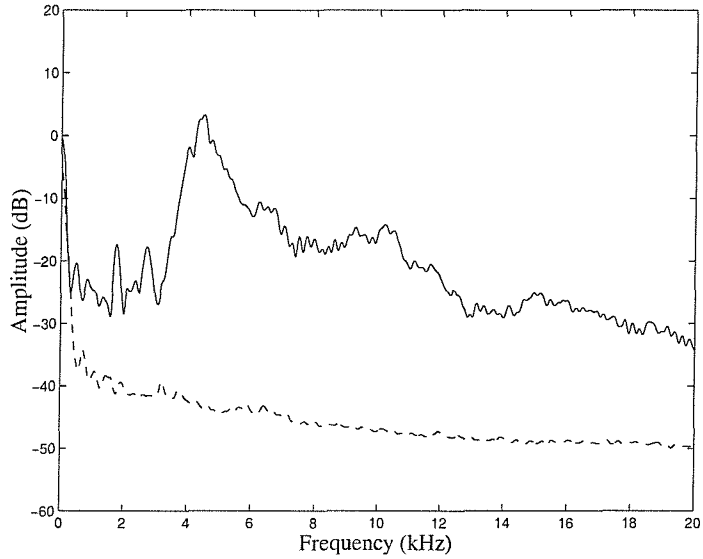

Figure 3.9: Time-averaged spectrum of sustained fricative /ussss ... u/. The dashed curve is the averaged spectrum of the room noise. Corpus 1a (Speaker LMTJ).

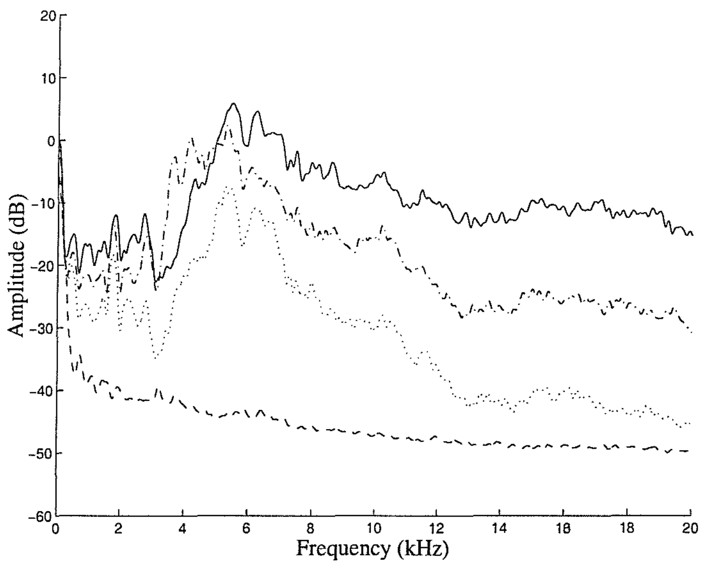

Figure 3.10: Time-averaged spectra of fricative /s/ sustained at three different effort levels: soft (dotted line), medium (dash-dotted line) and loud (solid line). The dashed curve is the averaged spectrum of the room noise. Corpus 1b (Speaker LMTJ).

An example of the time-averaged spectrum of a fricative in a Portuguese word is presented in Figure 3.11. The spectrum is not shown above 20 kHz because human hearing does not usually go beyond that limit (and the low-pass frequency is 22 kHz), and the spectrum below 22 Hz has also been filtered out (0 Hz in the figure actually corresponds to 21 Hz of the whole frequency range) because it clearly corresponds to room noise and other external artifacts. The dashed curve in the same figure corresponds to the time average of the room noise ($N = 500$; 10 ms windows). The figure has a significant low frequency peak that corresponds to room noise. The speech signal amplitude is considerably higher than the noise amplitude for most of the examples.

In the future we should also consider having a neutral vowel /ə/ context for fricatives in Corpus 1b, because the speaker naturally uses some vowel before starting the production of the actual fricative.

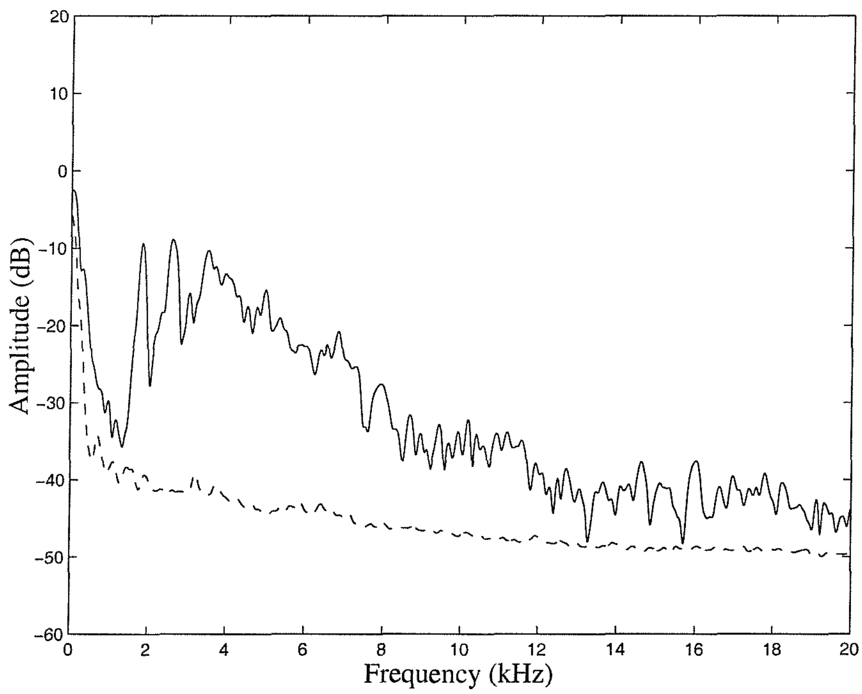

Figure 3.11: Time-averaged spectrum of fricative /ʒ/ in *jaqueta* /ʒɐˈketɐ/. The dashed curve is the averaged spectrum of the room noise. Corpus 3 (Speaker LMTJ).

3.5.2 Ensemble - Averaged Spectra

Ensemble averaging, based on one DFT computed at the same event in each of nine tokens, was used for Corpus 2. Three 10 ms windows located relative to events within one fricative were used: one left - aligned to the start of the fricative, one centred at the centre of the fricative, and one right - aligned to the end of the fricative, as shown in Figure 3.12. The ensemble - averaged power spectrum of each fricative is given by

$$P_E(f) = \frac{1}{9} \sum_{k=1}^{9} |X_k(f)|^2 \tag{3.5}$$

where X_k is the DFT of a portion of the fricative signal, x_k, corresponding to the windowed segment (at the beginning, middle or end of the fricative) of the kth token. Figure 3.13 shows examples of the ensemble - averaged spectra of fricative /f/ in Corpus 2.

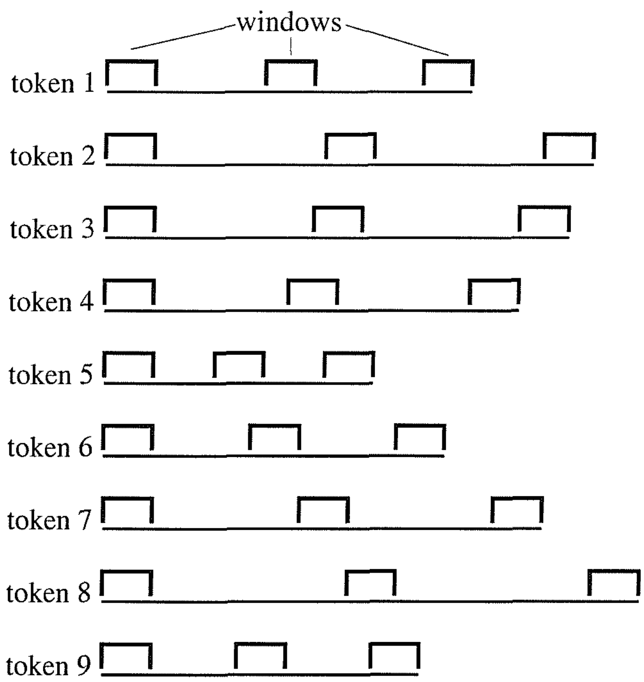

Figure 3.12: Schematic drawing of the placement of windows for the calculation of the ensemble - averaged spectra of fricative consonants, allowing for differing lengths of tokens.

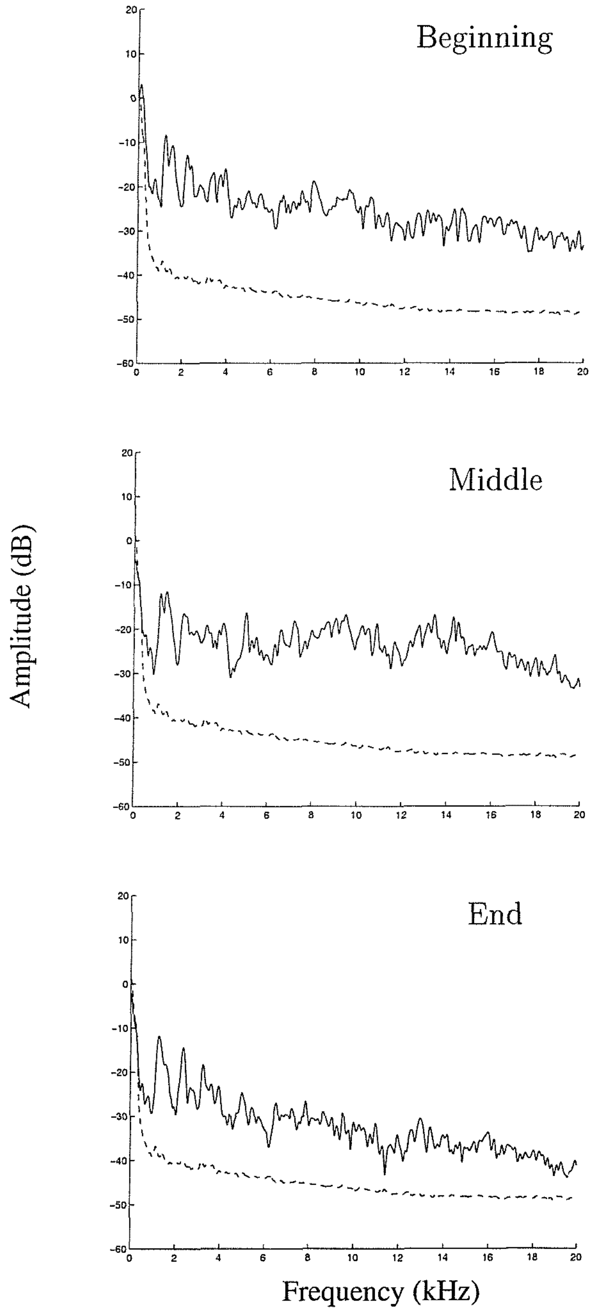

Figure 3.13: Ensemble-averaged spectra of fricative in /pɐfɐ/; Top: beginning of the fricative; Centre: middle of the fricative; Bottom: end of the fricative. The dashed curve is the time-averaged spectrum of the room noise. Corpus 2 (Speaker CFGA).

3.6 Spectral Analysis of Uvular Fricatives and Voiceless Tapped Alveolar Fricatives

The spectral smoothing procedure was different from the one used to analyse fricatives /f, v, s, z, ʃ, ʒ/. Since some of the fricatives were very short (11 to 117 ms), the necessary spectral averaging was achieved by first time-averaging and then spectral smoothing. Time-averaged spectra such as the one shown in Figure 3.14 (top), computed using three 10 ms windows (one left-aligned to the start of the fricative, one right-aligned to the end of the fricative and one centred at the middle of the fricative), were smoothed in the frequency domain (according to Bendat and Piersol 2000, pp. 432-434) by averaging together the results of 10 contiguous spectral components (see Figure 3.14, bottom).

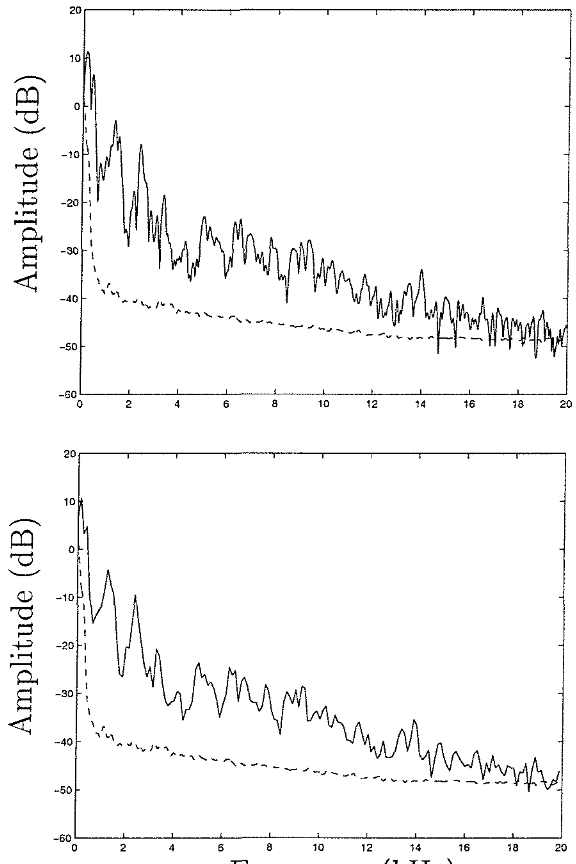

Figure 3.14: Time - averaged (top) and time and frequency - averaged (bottom) spectra of fricative [χ] in "ressaca" [ɐ χ'sakɐ] (preceded by the word "diga" ['digɐ]). The dashed curve is the time - averaged spectrum of the room noise. Corpus 4 (Speaker CFGA).

3.7 Summary

In this chapter the method used to segment and annotate the corpora has been described. This was followed by the definitions of manual and automatic criteria for voicing classification. Techniques used to average the power spectra of the fricatives (time - averaging, frequency - averaging and ensemble - averaging) were also presented.

Chapter 4

Results of Temporal and Devoicing Analysis

4.1 Introduction

This chapter presents a detailed discussion of the results from the temporal analysis of both the acoustic and laryngograph signals. This includes detailing the durations of the fricatives, and of the VF and FV transitions (Corpora 3 and 4) and a study of devoicing in Corpora 2, 3 and 4, together with discussion of some possible causes of this phenomenon. The correlation between devoicing and duration was also investigated.

Two devoicing criteria (a manual criterion and a criterion based on the ratio of variances of the laryngograph signal during the VF transition and during the fricative) were used to classify the examples into two/three categories. The results of the automatic measure of devoicing are compared with the manual ones, and an explanation for observed misclassifications is presented.

4.2 Duration of Fricatives, and of the VF and FV Transitions

There have been various studies in the past reporting the duration of fricatives, and of the VF and FV transitions, which identified several temporal cues aiding the perception of these sounds. The most relevant to the study presented in this thesis are revised in the section that follows.

4.2.1 Literature Review

Research results characterising the duration of American English fricatives include those of Behrens and Blumstein (1988) for voiceless fricatives in nonsense words: /f/ – 149 ms; /θ/ – 134 ms; /s/ – 174 ms; /ʃ/ – 175 ms. Stevens et al. (1992) also used nonsense words, but obtained significantly different results: /f/ – 94 ms; /v/ – 64 ms; /s/ – 108 ms; /z/ – 78 ms. They observed that the preceding vowel was longer when followed by a voiced fricative. Pirello et al. (1997) showed that the alveolar fricatives /s, z/ tend to have longer durations than the labiodental fricatives /f, v/ (mean durations in VC syllables said in isolation: /f/ – 214 ms; /v/ – 128 ms; /s/ – 236 ms; /z/ – 167 ms). Although vowel length played a role in the perception of voicing in word-final fricatives, it did not serve such a role in word-initial fricatives. Jongman et al. (2000) showed that the mean duration of fricatives, produced by 20 American English Speakers in nonsense words, differed significantly for sibilants [1] versus nonsibilants: /f/ – 166 ms; /v/ – 80 ms; /θ/ – 163 ms; /ð/ – 88 ms; /s/ – 178 ms; /z/ – 118 ms; /ʃ/ – 178 ms; /ʒ/ – 123 ms.

Docherty (1992), in a study of British English obstruents in real words said in isolation and in a carrier sentence, reported mean durations of 110 ms for /f/, of 108 ms for /θ/ and of 137 ms for /s/. The mean durations according to word-position are shown in Table 4.1. Results for Hebrew reported by Berkovits (1993) showed that the position in the sentence played an important role in the duration of word-final fricatives: sentence-medial /f/ – 84 ms; sentence-final /f/ – 235 ms; sentence-medial /v/ – 63 ms; sentence-final /v/ – 184 ms; sentence-medial /x/ – 80 ms; sentence-final /x/ – 257 ms.

[1] Sibilant – a fricative sound made by producing a narrow, groove-like stricture between the blade of the tongue and the back part of the alveolar ridge (Crystal 1997); /s, z/ and /ʃ, ʒ/ are examples. This is a phonetic classification based on the manner of articulation (see also definition of *strident* on page 89).

Table 4.1: Mean durations of British English fricatives in ms. After Docherty (1992).

	/f/	/v/	/θ/	/ð/	/s/	/z/	/ʃ/
Word - Initial	158	91	143	62	157	102	156
Word - Final	132	103	125	85	167	99	168

The difference in duration between unvoiced and voiced fricatives has been reported in many studies. One of the earliest ones was that of Slis and Cohen (1969), who reported unvoiced fricatives were on average 50 ms longer than voiced for Dutch. O'Shaughnessy (1974) proposed a durational model for the synthesis of American English consonants, observing that unvoiced fricatives were on average 30 - 40 ms longer than voiced fricatives, and that word - initial consonants were longer than word - final consonants. The results of Docherty (1992) also clearly showed that the duration of British English unvoiced fricatives was longer than their voiced counterparts, as shown in Table 4.1. Scully, in an acoustic and aerodynamic study of British English real words, reported mean durations of 90 ms for /s/ and of 45 ms for /z/, but also concluded that the "relative durations of vowels and fricatives are crucial for the perception of the [Vs] versus [Vz] contrast in English ..." (Scully 1979, p. 46), a fact which had been previously reported in a perceptual study of American English by Cole and Cooper (1975, pp. 1286 - 1287):

> In general, the primary cue for voiced - voiceless distinctions in syllable - final fricatives appears to be the ratio of durations of the frication and the preceding vowel, whereas the primary cue in syllable - initial fricatives appears to be duration of the following vowel.

These are interesting conclusions which will be referred to briefly when the results for Portuguese are discussed in Section 4.2.2. It is worth investigating further if Portuguese has the same duration cues.

Hogan and Rozsypal (1980) reported that the durations of Canadian English word - final /f/ and /s/ were longer on average than their voiced counterparts /v/ and /z/. Manrique and Massone's (1981) study of Argentine Spanish fricatives showed that the average durations of voiceless fricatives were 850 ms in isolation and 550 ms in CV context, and the durations of voiced fricatives were 400 ms in isolation and 250 ms in CV context. Soli (1982, pp. 376 - 377), in a study of American English, showed that the proportional durations of the transition and steady-state portions of the vowel in nonsense words (utterances of /jus/ and /juz/), were the main source of fricative information in the time domain:

> ... as the duration of a vowel is lengthened or shortened due to the voicing of a postvocalic fricative, the relative timing of the transition and steady-state portions of the vowel is also modified. These differences in the temporal structure of the vowel were shown to provide information for the voicing contrast, indicating that the vowel structure can combine with durational cues to specify linguistic information... Moreover, it appears that in natural speech the preceding vowel contains the major voicing cues for a final fricative, while the acoustic characteristics of the friction noise itself provide secondary voicing cues.

Scully et al. (1992, p. 40) stated that the underlying perceptual mechanisms of voiced-voiceless contrast should also be taken into account:

> The cross-over point for the perception of a voiced vs. voiceless fricative seemed to be associated with continued presence or absence of voicing at the time of rapid formant transitions associated with rapid changes of the vocal tract alveolar constriction cross-section areas.

Baum and Blumstein's (1987) acoustic and perceptual analysis of durational characteristics distinguishing American English fricatives in nonsense words showed that while the overall mean value of the voiceless fricatives (/f/ – 149 ms, /θ/ – 134 ms and /s/ – 174 ms) was longer than the voiced fricatives (/v/ – 116 ms, /ð/ – 107 ms and /z/ – 152 ms), there was considerable overlap in the duration distribution of voiced and voiceless fricative tokens. This was later confirmed by Crystal and House (1988), who also found that voiceless fricatives were longer than voiced fricatives (47 ms difference), and that the probability density distribution curves of the durations of voiced and voiceless fricatives, overlapped significantly, as shown in Figure 4.1. Stevens et al. (1992) reported a 30 ms voiced-voiceless difference for American English. Smith (1995, 1997) reported that the duration of frication of American English /z/ was shorter than for /s/, and that vowels were significantly longer before /z/ than before /s/. Mair and Shadle (1996) studied voiced-voiceless distinction using EPG, acoustic and aerodynamic data for a male

French speaker, by comparing the voiceless-voiced pairs /s, z/ and /ʃ, ʒ/ in nonsense words. A detailed statistical study revealed that the duration of voiceless fricatives was longer, and that durations of the preceding and following vowels were longer for voiced fricative consonants. The total VCV duration was longer for voiced fricatives, and voicing amplitude diminished earlier and resumed later in the voiceless compared with the voiced fricatives.

A variety of work has also concentrated on investigating the influence of stress on fricative durations. Klatt (1971, 1974) studied real word-level phenomena (stress, word position, number of syllables, etc.) that influenced the duration of /s/ produced by three male American English speakers (RK, KNS and DHK). Analysis of broadband spectrograms showed that the mean fricative durations, in words where /s/ was in various sV, VsV and sV vowel contexts (the corpus also included words where /s/ was part of a consonant cluster), were 127 ms for RK, 125 ms for KNS and 100 ms for DHK. Stressed /s/ was approximately 15% longer than /s/ in an unstressed syllable (see Table 4.2). The duration of /s/ in a word with many syllables was shorter than in single syllable words, but /s/ was more resistant to shortening than vowels.

Table 4.2: Duration of fricative /s/ in ms for American English speakers RK, KNS and DHK. After Klatt (1971, 1974).

	RK	KNS	DHK
Word-initial and stressed (primary)	142	140	106
Word-initial and stressed (secondary)	122	125	93
Word-initial and unstressed	118	119	96
Word-medial and stressed (primary)	136	131	108
Word-medial and stressed (secondary)	134	116	95
Word-medial and unstressed	115	117	96
Word-final	130	121	98

However, in a more complete study of connected discourse read by an American English speaker, Klatt (1975) reported little or no effect of stress on the mean duration of fricatives, as shown in Table 4.3. The results of Crystal and House (1988) are also inconclusive regarding the effect of stress. The probability densities of the durations of fricatives occurring in stressed and unstressed syllables overlapped significantly, as shown in Figure 4.1. However, the results of Manrique and Massone's (1981) study of Argentine Spanish fricatives showed that fricatives were longer in unstressed syllables than in stressed ones, as shown in Table 4.3.

Table 4.3: Duration of fricatives in American English (Klatt 1975) and Argentine Spanish (Manrique and Massone 1981).

	Duration (ms)	
	American English	Argentine Spanish
[f] stressed	110	147
[f] unstressed	105	192
[v] stressed	75	-
[v] unstressed	65	-
[θ] stressed	100	-
[θ] unstressed	95	-
[ð] stressed	60	98
[ð] unstressed	60	104
[s] stressed	120	148
[s] unstressed	120	187
[z] stressed	60	-
[z] unstressed	60	-
[ʃ] stressed	110	170
[ʃ] unstressed	110	210
[ʒ] stressed	-	98
[ʒ] unstressed	-	149

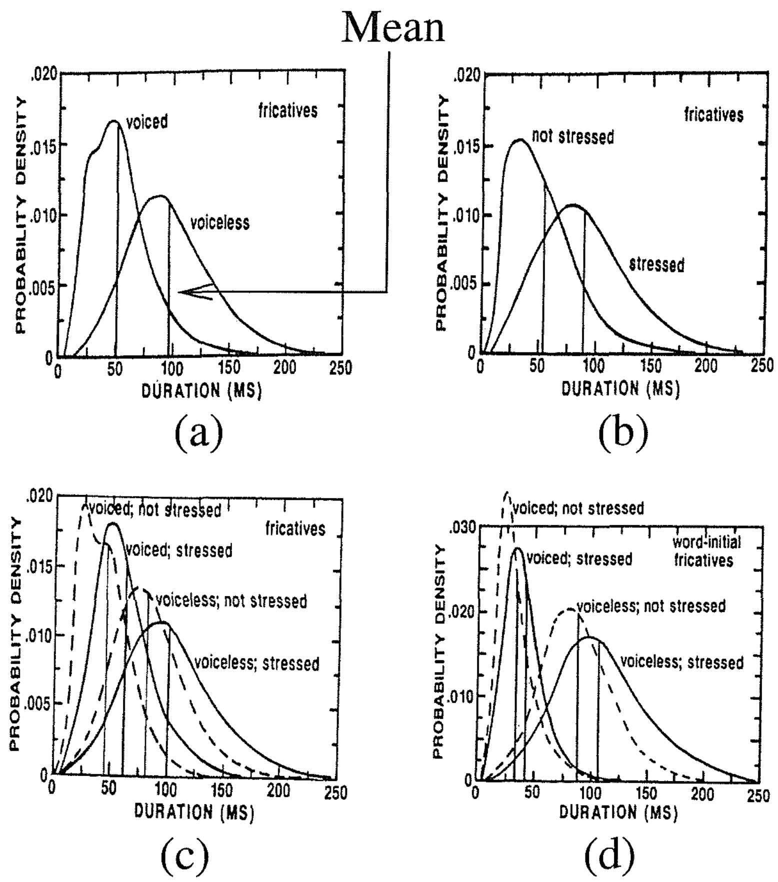

Figure 4.1: Probability density distribution curves of the durations of: (a) voiced /v, ð, z/ and voiceless /f, θ, s/ fricatives; (b) fricatives occurring in stressed and unstressed syllables; (c) voiced and voiceless fricatives occurring in stressed and unstressed syllables; (d) word-initial voiced and voiceless fricatives occurring in stressed and unstressed syllables. From Crystal and House (1988).

The influence of vowel context on the duration of fricatives has also been widely studied. Schwartz (1969) reported that, for American English fricatives in nonsense words, the duration of /s/ and /ʃ/ in /i-i/ vowel context was significantly longer than in /a-a/ vowel context. Schwartz also concluded that both the /s/ and /ʃ/ duration differences were the result of an influence of the final vowel and not the initial. This interpretation was drawn from the finding that the fricatives in /a-i/ and /i-i/ vowel contexts were longer than those in /i-a/ and /a-a/, whereas no significant duration differences existed between the /i-a/ and /a-a/ and between the /a-i/ and /i-i/ vowel contexts. LaRiviere et al. (1975) also reported significant effects of vowel context on the duration of American English unvoiced fricatives:

- /fi/ – 126 ms; /fa/ – 130 ms; /fu/ – 175 ms;
- /θi/ – 122 ms; /θa/ – 142 ms; /θu/ – 128 ms;
- /si/ – 160 ms; /sa/ – 148 ms; /su/ – 148 ms;
- /ʃi/ – 160 ms; /ʃa/ – 190 ms; /ʃu/ – 175 ms.

However, the study of Behrens and Blumstein (1988) showed minimal effects of vowel context on duration of American English fricatives. Mair and Shadle (1996) reported that, for a male French speaker, the total VCV duration was affected by vowel context, and that duration was longer for /a - a/ than /i - i/ and longer for /i - i/ than /u - u/ in both the voiced and voiceless fricatives, but durations were inconsistent with regard to fricative place of articulation.

4.2.2 Results of Temporal Analysis of Labiodental, Alveolar and Postalveolar Fricatives

A complete analysis of duration in Corpus 3 is shown in Table 4.4 and illustrated in Figures 4.5 to 4.8. The minimum, maximum and mean durations (averaged over the four speakers) of fricatives from Corpus 3 were: $77 \leq$ /f/ ≤ 180 ms (mean = 120 ms); $39 \leq$ /v/ ≤ 128 ms (mean = 72 ms); $88 \leq$ /s/ ≤ 180 ms (mean = 127 ms); $48 \leq$ /z/ ≤ 120 ms (mean = 79 ms); $62 \leq$ /ʃ/ ≤ 191 ms (mean = 114 ms); $55 \leq$ /ʒ/ ≤ 128 ms (mean = 85 ms).

The mean duration of the unvoiced fricatives is always greater than the mean duration of the voiced fricatives, as shown in Figure 4.2. There is no significant difference by place of articulation. The mean duration of the fricative is greater than the mean duration of the VF and FV transitions, and the mean duration of the VF transition is greater than the mean duration of the FV transition for speakers LMTJ, ACC and ISSS, as shown in Figures 4.3 and 4.4. For Speaker LMTJ's /s/ in word - initial position, as the following vowel's place of articulation moves further back, the duration of the fricative diminishes. This was only observed for this fricative produced by Speaker LMTJ.

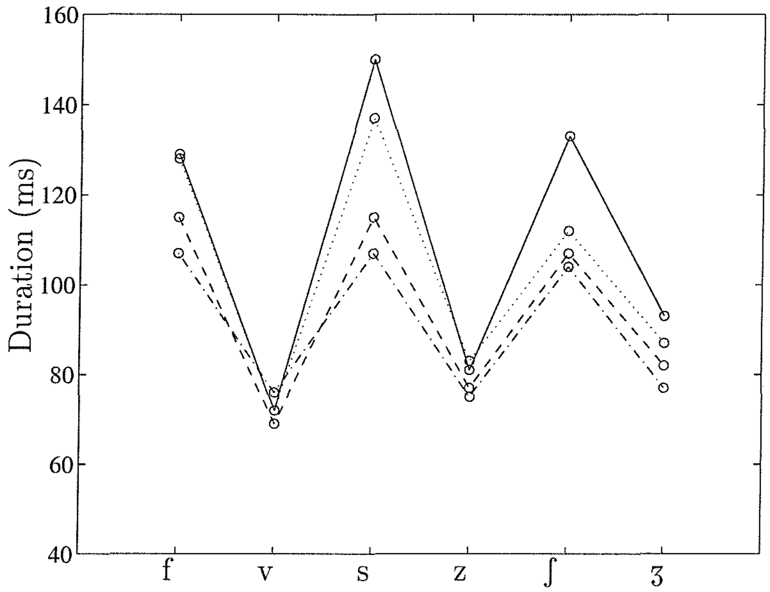

Figure 4.2: Mean duration of fricatives /f, v, s, z, ʃ, ʒ/ in Corpus 3. Speaker LMTJ – solid line; Speaker CFGA – dash-dotted line; Speaker ACC – dashed line; Speaker ISSS – dotted line.

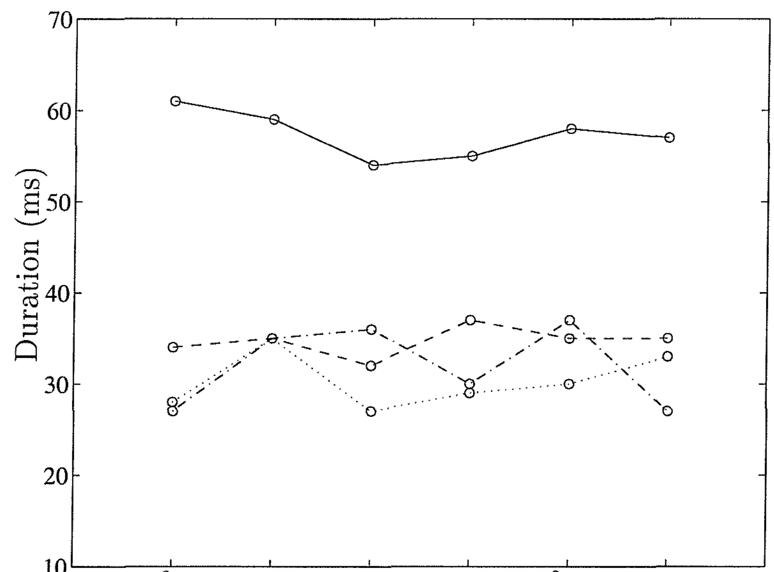

Figure 4.3: Mean duration of VF transitions in Corpus 3. Speaker LMTJ – solid line; Speaker CFGA – dash-dotted line; Speaker ACC – dashed line; Speaker ISSS – dotted line.

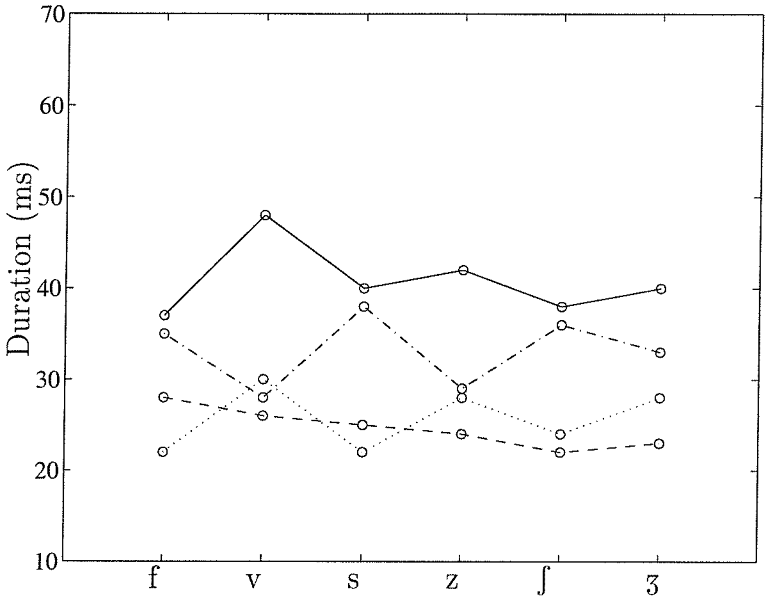

Figure 4.4: Mean duration of FV transitions in Corpus 3. Speaker LMTJ – solid line; Speaker CFGA – dash-dotted line; Speaker ACC – dashed line; Speaker ISSS – dotted line.

Table 4.4: Duration of fricatives and of VF and FV transitions in Corpus 3.

Speaker LMTJ									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	37	61	108	86	129	182	13	37	80
/v/	24	59	100	22	72	135	26	48	85
/s/	25	54	90	106	150	220	17	40	65
/z/	24	55	101	46	81	117	21	42	79
/ʃ/	29	58	123	76	133	194	21	38	53
/ʒ/	24	57	129	60	93	139	15	40	115

Speaker CFGA									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	15	27	41	70	107	170	22	35	48
/v/	17	35	116	46	76	133	17	28	39
/s/	15	36	59	63	107	142	23	38	50
/z/	18	30	44	41	75	133	23	29	48
/ʃ/	20	37	67	58	104	237	19	36	49
/ʒ/	18	27	44	54	77	122	23	33	53

Speaker ACC									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	27	34	49	77	115	161	17	28	41
/v/	13	35	56	48	69	117	14	26	48
/s/	24	32	58	83	115	170	13	25	39
/z/	20	37	60	55	77	100	13	24	36
/ʃ/	21	35	50	56	107	176	13	22	34
/ʒ/	17	35	48	46	82	131	13	23	39

Speaker ISSS									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	18	28	42	76	128	207	17	22	31
/v/	25	35	56	38	72	126	16	30	51
/s/	21	27	37	100	137	188	13	22	37
/z/	20	29	40	50	83	129	16	28	39
/ʃ/	15	30	48	58	112	158	15	24	38
/ʒ/	22	33	51	61	87	119	13	28	48

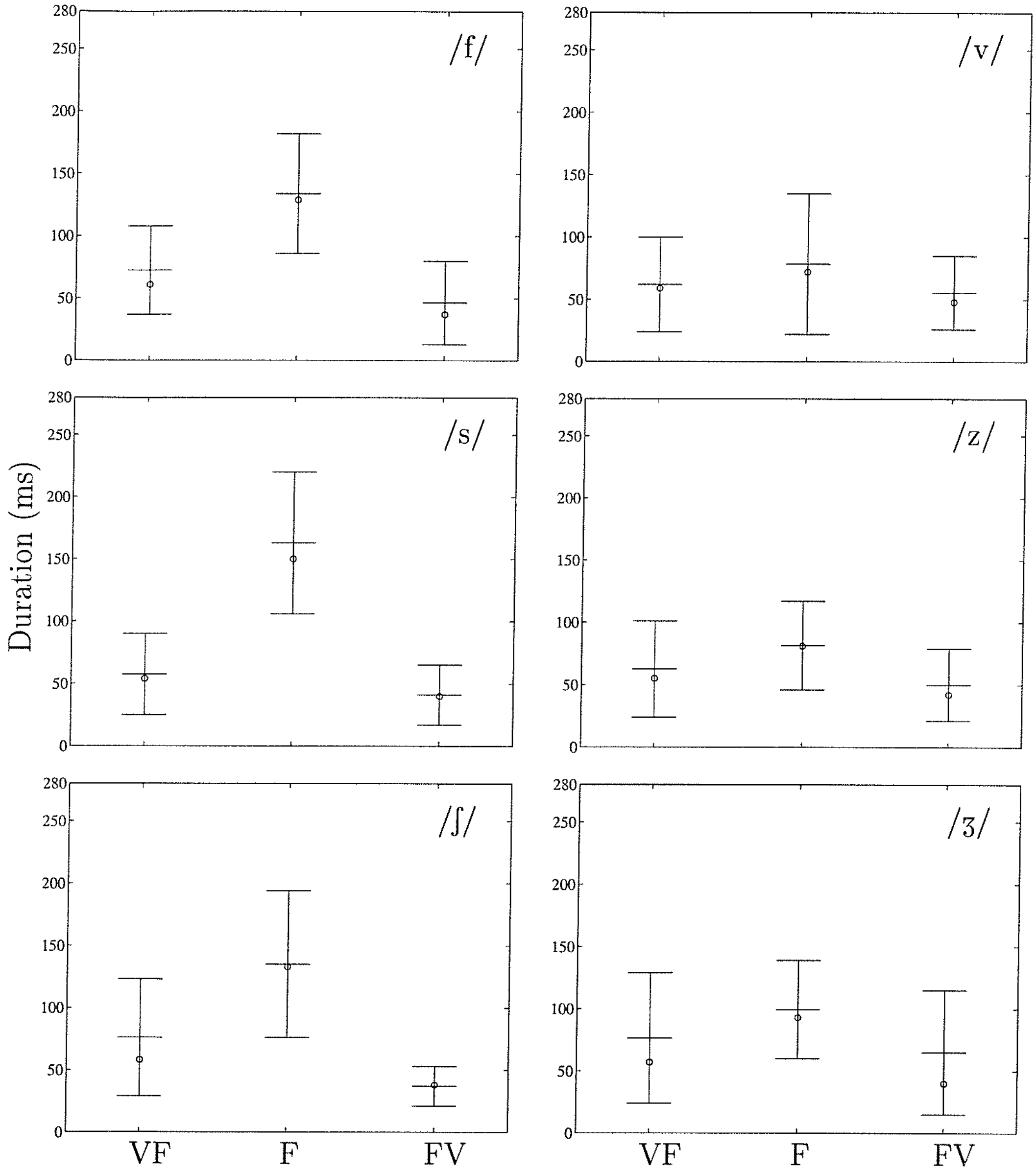

Figure 4.5: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 3 (Speaker LMTJ).

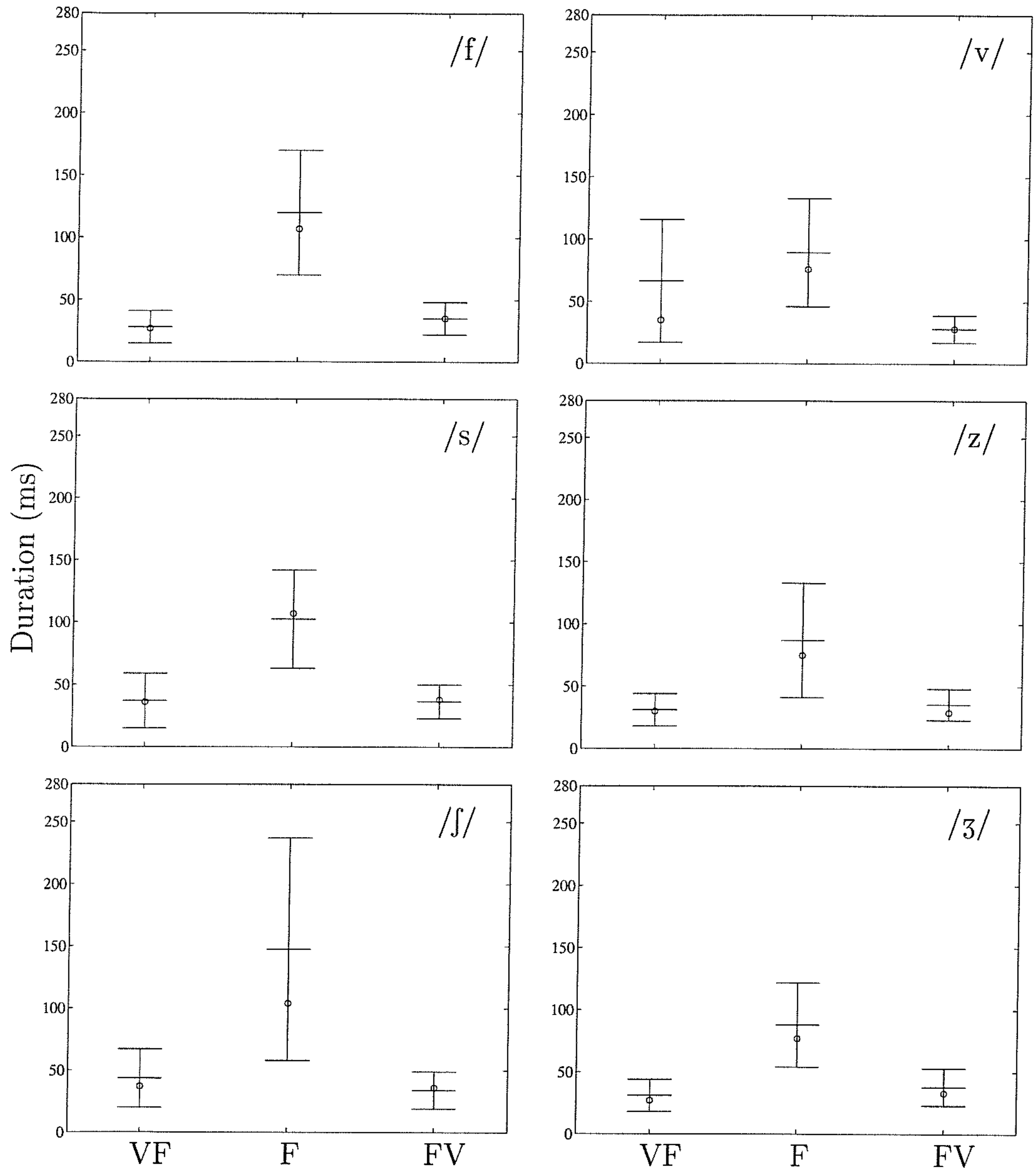

Figure 4.6: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 3 (Speaker CFGA).

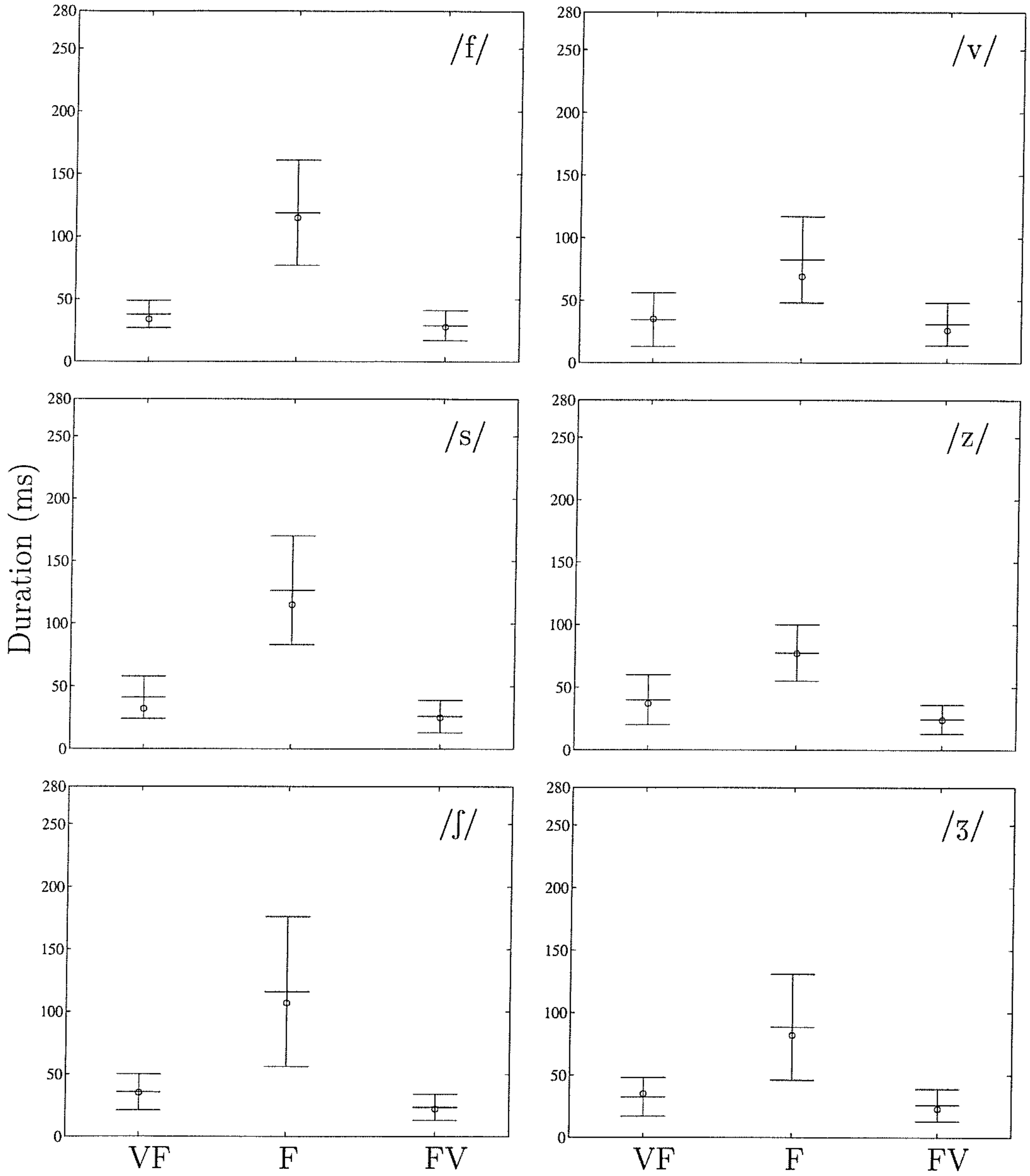

Figure 4.7: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 3 (Speaker ACC).

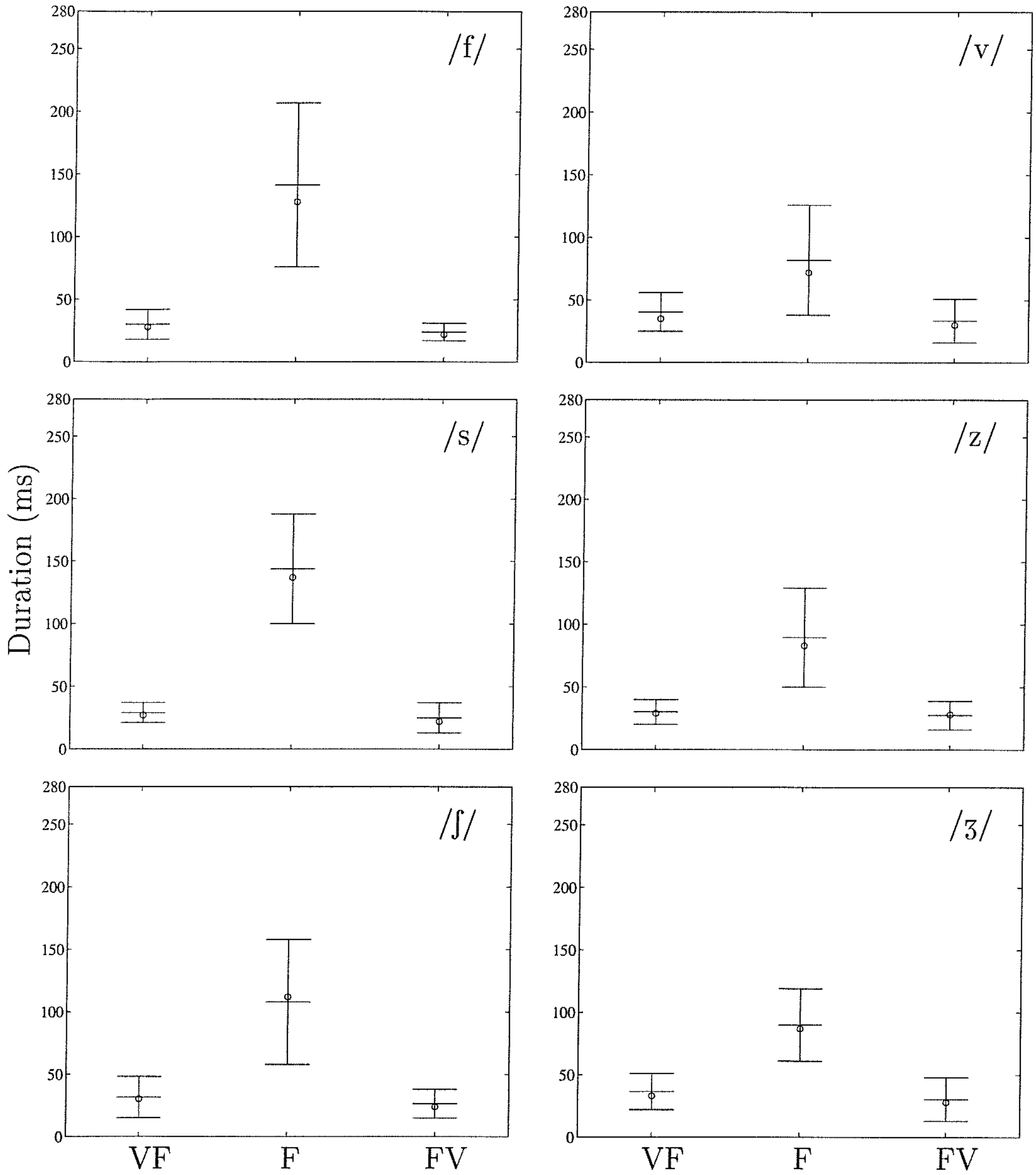

Figure 4.8: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 3 (Speaker ISSS).

A detailed time analysis of fricatives, and VF and FV transitions in Corpus 4 is presented in Table 4.5, illustrated in Figures 4.12 to 4.15, and can be summarized as follows: $72 \leq$ /f/ ≤ 148 ms (mean = 110 ms); $36 \leq$ /v/ ≤ 104 ms (mean = 60 ms); $78 \leq$ /s/ ≤ 230 ms (mean = 130 ms); $44 \leq$ /z/ ≤ 178 ms (mean = 82 ms); $75 \leq$ /ʃ/ ≤ 159 ms (mean = 120 ms); $36 \leq$ /ʒ/ ≤ 135 ms (mean = 74 ms). Minimum, maximum and mean durations of fricatives were averaged over the four speakers to allow comparisons to most other studies. The mean duration of the fricatives in the word corpus (Corpus 3) is quite similar to the mean duration of fricatives from the sentence corpus (Corpus 4).

As previously observed for Corpus 3 fricatives, the duration of Corpus 4 unvoiced fricatives is always greater than the duration of their voiced counterparts, as shown in Figure 4.9, which agrees with results for the English language (Hogan and Rozsypal 1980; Crystal and House 1988; Stevens et al. 1992; Pirello et al. 1997). The mean duration of the fricatives is greater than the duration of the VF and FV transitions, and comparing the mean duration of the VF and FV transitions, no consistent pattern can be found for any of the speakers (compare with results for Corpus 3), as shown in Figures 4.10 and 4.11.

Corpus 4 alveolar fricatives /s, z/ are on average longer than labiodentals /f, v/, a fact that had been previously reported by Pirello et al. (1997) for English, but which was not observed in Corpus 3 fricatives. Word-final fricatives at the end of the sentences read by Speaker LMTJ have much longer duration than other examples in Corpus 4.

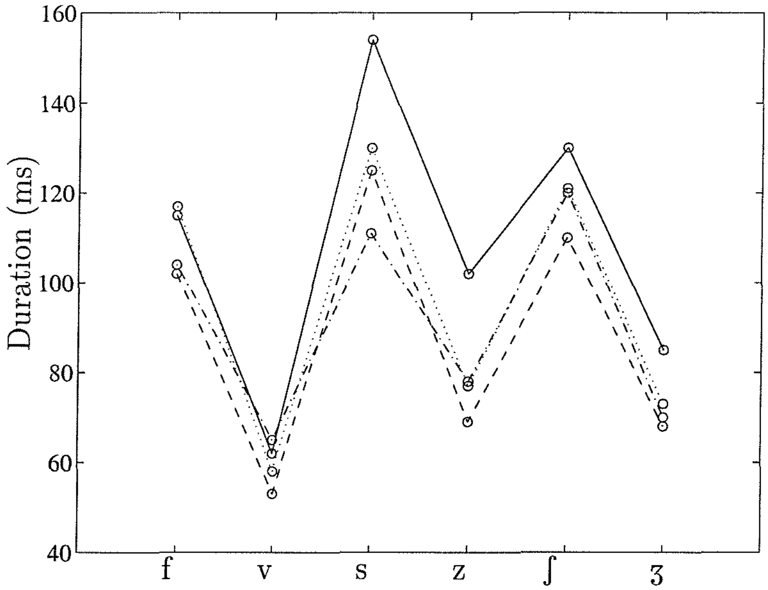

Figure 4.9: Mean duration of fricatives /f, v, s, z, ʃ, ʒ/ in Corpus 4. Speaker LMTJ – solid line; Speaker CFGA – dash-dotted line; Speaker ACC – dashed line; Speaker ISSS – dotted line.

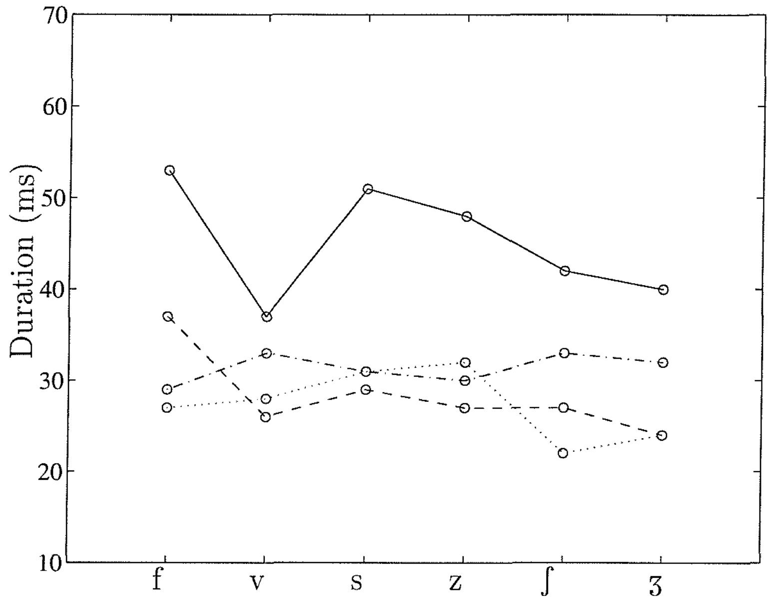

Figure 4.10: Mean duration of VF transitions in Corpus 4. Speaker LMTJ – solid line; Speaker CFGA – dash-dotted line; Speaker ACC – dashed line; Speaker ISSS – dotted line.

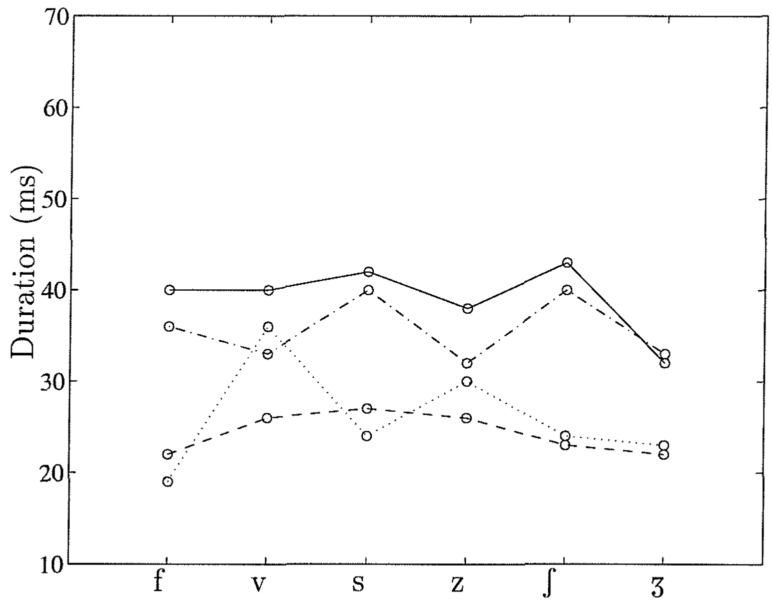

Figure 4.11: Mean duration of FV transitions in Corpus 4. Speaker LMTJ – solid line; Speaker CFGA – dash-dotted line; Speaker ACC – dashed line; Speaker ISSS – dotted line.

Table 4.5: Duration of fricatives and of VF and FV transitions in Corpus 4.

Speaker LMTJ									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	37	53	68	85	115	149	29	40	62
/v/	10	37	64	41	62	101	22	40	56
/s/	24	51	104	95	154	272	30	42	67
/z/	23	48	104	49	102	268	21	38	55
/ʃ/	31	42	58	40	130	186	23	43	74
/ʒ/	17	40	79	48	85	156	21	32	40
Speaker CFGA									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	16	29	39	61	104	152	19	36	55
/v/	17	33	73	40	65	100	17	33	68
/s/	20	31	53	67	111	199	23	40	63
/z/	17	30	53	49	78	168	16	32	50
/ʃ/	15	33	54	99	120	145	19	40	72
/ʒ/	16	32	65	36	70	123	23	33	47
Speaker ACC									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	25	37	142	70	102	135	11	22	40
/v/	11	26	65	30	53	109	12	26	70
/s/	16	29	48	85	125	229	14	27	57
/z/	12	27	49	41	69	122	12	26	44
/ʃ/	17	27	39	79	110	146	13	23	45
/ʒ/	12	24	47	40	68	121	12	22	41
Speaker ISSS									
	VF Transition			Fricative			FV Transition		
	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)	Min. (ms)	Mean (ms)	Max. (ms)
/f/	16	27	38	73	117	157	14	19	27
/v/	20	28	47	31	58	106	15	36	252
/s/	20	31	50	64	130	221	14	24	41
/z/	21	32	55	35	77	152	16	30	61
/ʃ/	16	22	28	80	121	157	13	24	41
/ʒ/	19	24	40	21	73	140	13	23	40

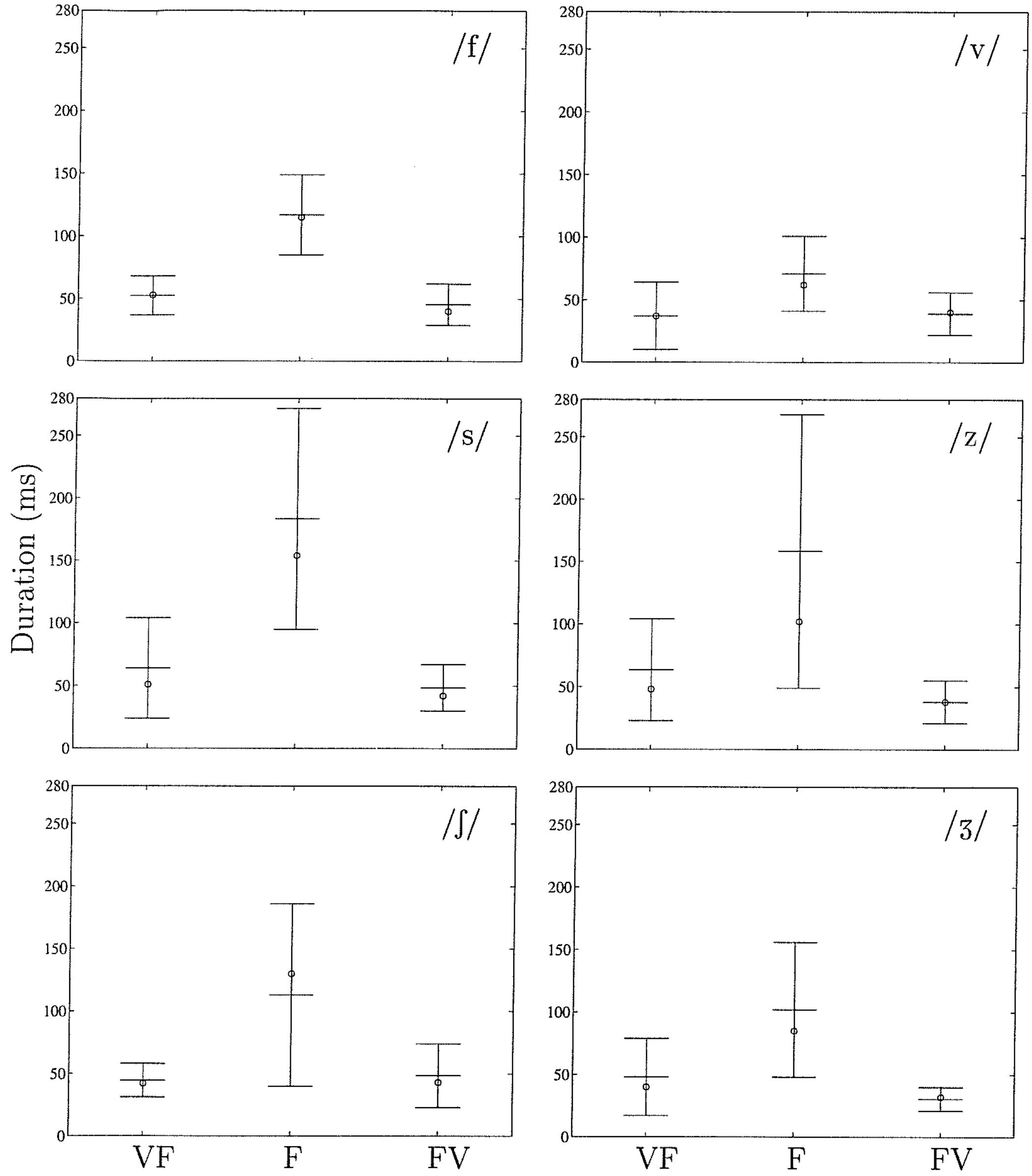

Figure 4.12: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 4 (Speaker LMTJ).

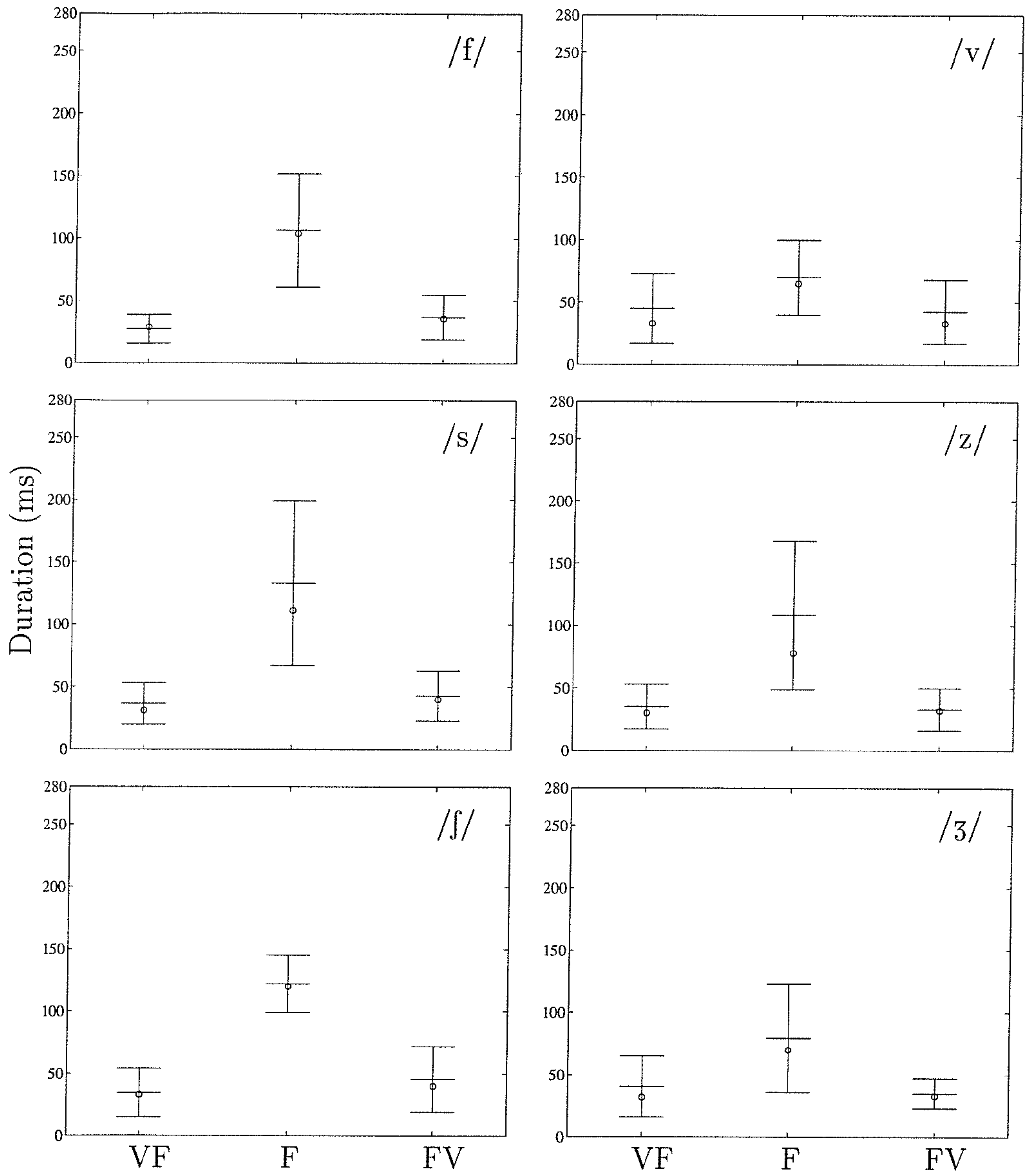

Figure 4.13: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 4 (Speaker CFGA).

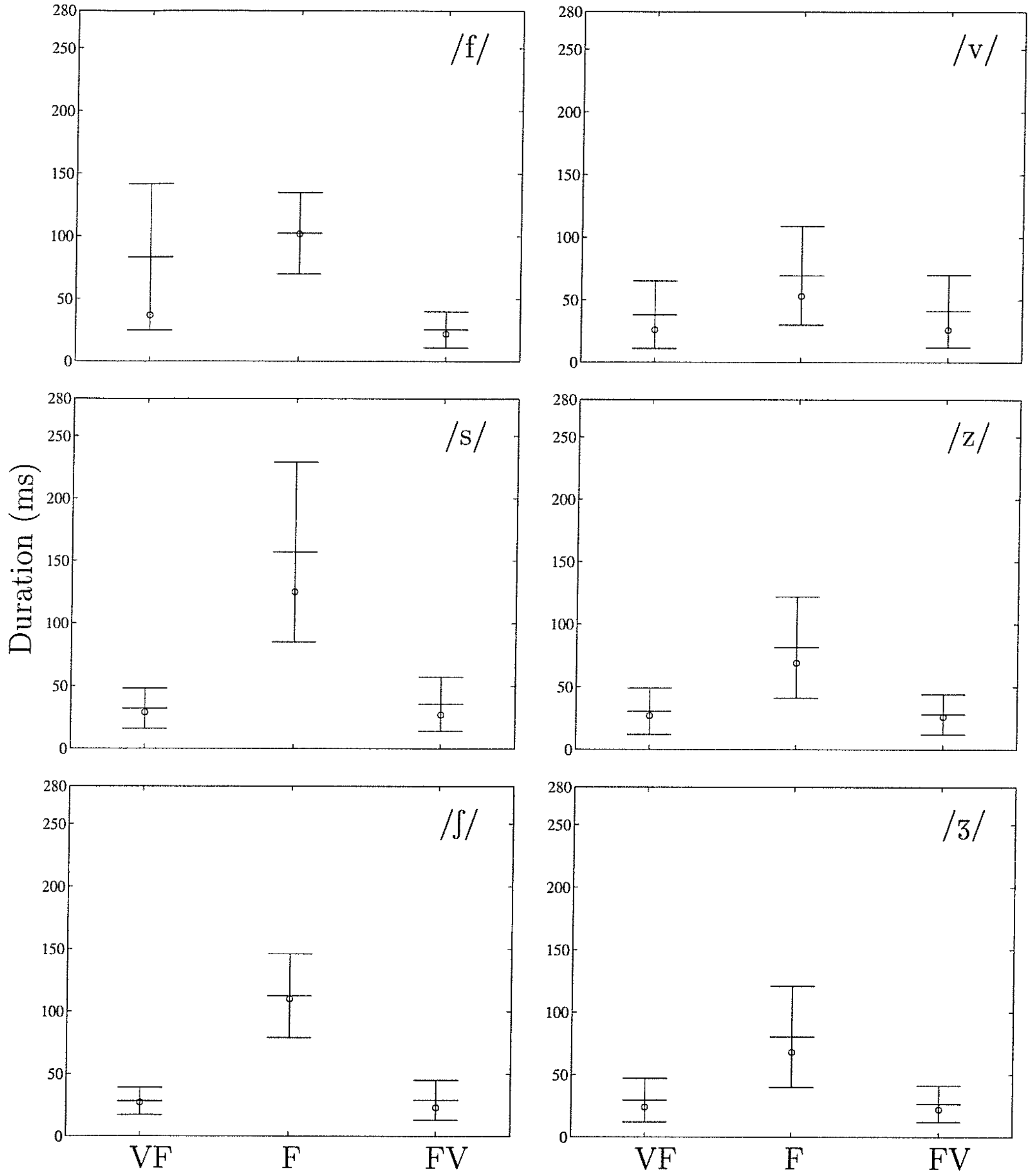

Figure 4.14: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 4 (Speaker ACC).

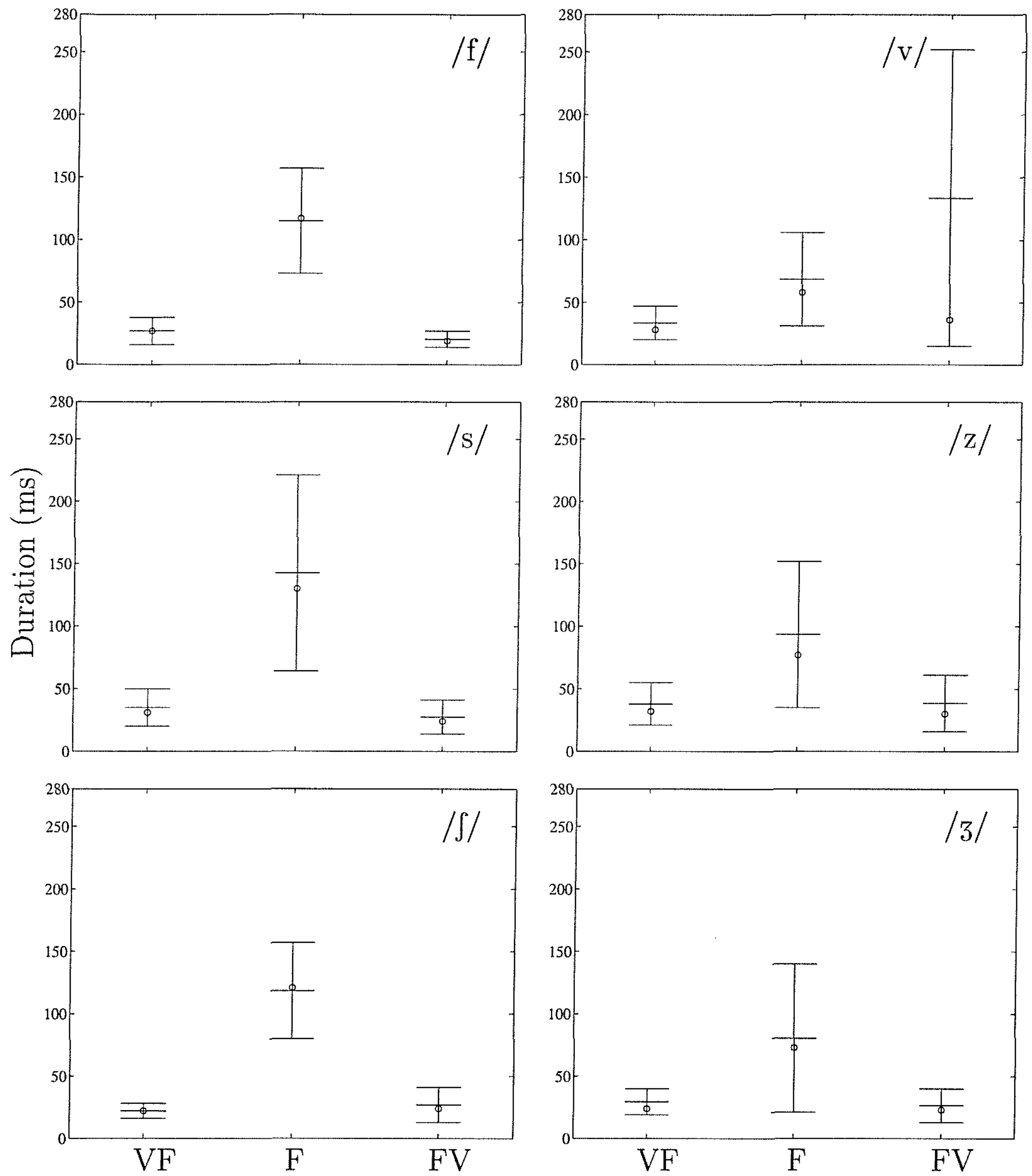

Figure 4.15: Range of values of duration of fricatives /f, v, s, z, ʃ, ʒ/, and of VF and FV transitions. o is the mean. Corpus 4 (Speaker ISSS).

4.2.3 Analysis of Variance of Duration of Labiodental, Alveolar and Postalveolar Fricatives

One-way analysis of variance (ANOVA) was used to study the effects of the independent variables (factors) speaker (LMTJ, CFGA, ACC and ISSS), place of articulation (labiodental, alveolar and postalveolar) and position in word (word - initial, word - medial and word - final) on the dependent variable duration of fricatives in Corpus 3 and in Corpus 4. Two separate ANOVAs were run to study the correlation between place of articulation and duration, because from the analysis presented in previous subsections, it was clear that the duration of unvoiced fricatives was always significantly greater than their voiced counterparts. Since place was the independent variable and not voicing, the two subgroups /f, s, ʃ/ and /v, z, ʒ/ had to be analysed separately. The software package SPSS 10.0 for Windows (SPSS 1999a; SPSS 1999b) was used to run all statistical tests and the results were as follows. There was a significant effect of the factor `speaker` on the duration of fricatives /f, s, ʃ/ both in Corpus 3 ($F(3, 344) = 21.667$, $p < 0.001$) [2] and Corpus 4 ($F(3, 318) = 11.041$, $p < 0.001$). Both values of F are inaccurate because the Levene test [3]was significant. There was a significant linear trend in Corpus 3 ($F(1, 344) = 5.775$, $p = 0.017$) and Corpus 4 ($F(1, 318) = 4.853$, $p = 0.028$).

There was no significant effect of the factor `speaker` on the duration of fricatives /v, z, ʒ/ in Corpus 3 ($F(3, 337) = 1.876$, $p = 0.133$), but in Corpus 4 there was a significant effect ($F(3, 409) = 10.064$, $p < 0.001$). Both values of F are inaccurate because the Levene test was significant. There was a significant linear trend in Corpus 4: $F(1, 409) = 16.567$, $p < 0.001$.

There was a significant effect of the factor `place` on the duration of fricatives /f, s, ʃ/ both in Corpus 3 ($F(2, 345) = 5.358$, $p = 0.005$) and Corpus 4 ($F(2, 319) = 10.680$, $p < 0.001$). The value of F in Corpus 4 is inaccurate because the Levene test was significant. There was a significant linear trend in Corpus 4: $F(1, 319) = 6.391$, $p = 0.012$.

There was a significant effect of the factor `place` on the duration of fricatives

[2]The F statistic is the ratio of the mean squares for each source of variability (model mean squares and residual mean squares). The degrees of freedom used to assess F are the degrees of freedom of the model, 3, and the degrees of freedom for the residuals of the model, 344. The p-value is derived from F, and as F increases the p-value decreases. For $p < 0.05$ results are significant. For this example, the p-value of less than 0.001 means that there is less than 0.1% chance that the F-ratio of 21.667 would happen by chance alone (Field 2000, p. 112).

[3]"... if Levene's test is significant at $p \leq 0.05$ then we can conclude that the null hypothesis is incorrect and that the variances are significantly different – therefore, the assumption of homogeneity of variances has been violated. If, however, Levene's test is non-significant (i.e. $p > 0.05$) then we must accept the null hypothesis that the difference between the variances is zero – the variances are roughly equal and the assumption tenable" (Field 2000, p. 238).

/v, z, ʒ/ both in Corpus 3 ($F(2, 338) = 10.943$, $p < 0.001$) and Corpus 4 ($F(2, 410) = 30.693$, $p < 0.001$). Both values of F are inaccurate because the Levene test was significant. There was a significant linear trend (Corpus 3 – $F(1, 338) = 21.787$, $p < 0.001$; Corpus 4 – $F(1, 410) = 26.091$, $p < 0.001$). In Corpus 3, as the place of articulation moved further back, duration increased proportionately.

There was a significant effect of the factor `position in word` on the duration of fricatives /f, s, ʃ/ both in Corpus 3 ($F(2, 345) = 11.547$, $p < 0.001$) and Corpus 4 ($F(2, 319) = 10.309$, $p < 0.001$). Both values of F are inaccurate because the Levene test was significant. There was a significant linear trend in Corpus 3: $F(1, 345) = 11.128$, $p = 0.001$.

There was a significant effect of the factor `position in word` on the duration of fricatives /v, z, ʒ/ both in Corpus 3 ($F(2, 338) = 49.834$, $p < 0.001$) and Corpus 4 ($F(2, 410) = 5.425$, $p = 0.005$). Both values of F are inaccurate because the Levene test was significant. There was a significant linear trend (Corpus 3 – $F(1, 338) = 91.513$, $p < 0.001$; Corpus 4 – $F(1, 410) = 10.415$, $p = 0.001$) indicating that as the position of the fricative moves from initial, through medial, to final word position, duration increased proportionately.

4.2.4 Temporal Analysis of Uvular Fricatives and Voiceless Tapped Alveolar Fricatives

Duration of the steady state portion of the fricatives and of their VF and FV transitions were measured to describe the characteristics of European Portuguese and to compare them with results reported for various other languages. The duration of fricative [χ] varied from 23 ms to 117 ms (median duration = 69 ms), the VF transition from 15 ms to 41 ms (median duration = 25 ms), and the FV transition from 22 ms to 43 ms (median duration = 32 ms), as shown in Table E.1. The duration of fricative [ɾ̥] varies from 11 ms to 85 ms (median duration = 22 ms), the VF transition from 12 ms to 103 ms (median duration = 30 ms) and the FV transition from 13 ms to 58 ms (median duration = 21 ms), as shown in Tables E.3 to E.6. Twelve out of the 86 examples of this fricative occurred word-initially (14 %), 20 word-medially (23 %), and 54 word-finally (63 %).

These durational results for Portuguese uvular fricatives are different from those of Manrique and Massone's (1981) study of Argentine Spanish, who showed that for velar fricatives the average durations for four speakers were: /x/ (unstressed) – 196 ms; /x/ (stressed) – 147 ms; /ɣ/ (unstressed) – 92 ms; /ɣ/ (stressed) – 58 ms. Results of a study by Alwan (1986), of the production and perception of Arabic pharyngeal and uvular consonants, revealed fricative duration values similar to those of Manrique and Massone. The voiceless consonants were longer (/χ/ – 169 ms averaged across all vowel contexts and four speakers) than the voiced consonants (/ʁ/ – 113 ms), but vowel context did not significantly affect the duration of the consonants. European Portuguese fricatives are shorter probably as a result of the naturalness of the corpora studied here (only real words), contrary to the focus on nonsense words in the studies of Manrique and Massone (1981) and Alwan (1986).

4.3 Devoicing

In this section a study of devoicing in Corpora 2, 3 and 4 is presented. We start by reviewing early works reporting frequent devoicing of voiced fricatives, including those of Raphael (1972), Klatt (1976), Haggard (1978), Hogan and Rozsypal (1980), Scully (1971), Soli (1982), Veatch (1989), and Scully (1992).

4.3.1 Literature Review

Stevens et al. (1992) reported a significant number of devoiced examples of American English /v/ and /z/, e.g., 22% of the tokens of the fricative in [azə] were devoiced. /v/ tended to exhibit glottal vibration continuing throughout the entire fricative more often than /z/, and fricatives in word-final position were virtually always devoiced. An early study of the effect of devoicing on the duration of fricatives and of the preceding vowels led Klatt (1976, p. 1219) to conclude that:

> Many English speakers devoice postvocalic voiced fricatives, suggesting that English may be changing in the direction of using vowel duration or the ratio of fricative duration to preceding vowel duration as a primary cue for the voicing contrast.

Hogan and Rozsypal (1980, p. 1770), quoting the same (Klatt 1976) study and also an earlier one by Raphael (1972), supported the same theory:

> Vowel to fricative duration ratio appears to be pertinent for recognition. Klatt (1976) has noted that many English speakers devoice postvocalic voiced fricatives indicating that vowel duration or vowel to fricative duration ratio may be assuming the functional load for the voicing contrast. It was noted by Raphael (1972) that when voicing was added during the frication portion, change in the preceding vowel duration became less effective in influencing the recognition scores.

Haggard (1978) studied devoicing of British English voiced fricatives in vowel context, and the effects of a stop before the fricative and a voiceless phoneme after it. The speech material was originally collected in 1969. Results showed that devoicing depends upon place of articulation, a preceding stop or a succeeding voiceless phoneme, and that there is higher incidence of devoicing for /z, ʒ/ than for /v, ð/. The mean percentage of devoicing in various contextual conditions in real words is shown in Table 4.6.

Table 4.6: Mean percentage of devoicing. After Haggard (1978).

	/v/	/z/	/ʒ/
Word - initial before stressed vowel	21	34	-
Word - initial before unstressed vowel	24	-	-
Word - medial before stressed vowel	23	30	-
Word - medial before unstressed vowel	8	39	29
Word - medial between stressed vowels	37	90	-
Word - final after stressed vowel	95	99	100
Word - final after unstressed vowel	92	99	-

In a second experiment, the degree of glottal opening was measured for nonsense words, revealing a slight overall tendency for devoicing to be higher when the fricative follows a stressed vowel than when it precedes a stressed vowel, or in word - initial position. This effect was due to the stress placement resulting from an emphatic stress [4] rather than a lexical stress [5] as in the experiment with real words.

[4]Emphatic stress – used to provide a means of distinguishing degrees of emphasis or contrast in sentences.

[5]Lexical stress – stress pattern as marked in language specific phonetic dictionaries.

Scully's (1979) acoustic and aerodynamic study of British English /s, z/, produced in real words by one female speaker, showed that there was continuation of voicing through the whole frication segment for only 7 out of 18 (39%) tokens of /z/. However, Docherty's (1992) study of British English obstruents in real words, said in isolation and in a carrier sentence by five male speakers, revealed that most of the /v, ð, z/ tokens were fully voiced, as shown in Table 4.7, and that /v/ had longer intervals of voicing than /z/. Fricative /ð/ had longer intervals of voicing than /v/ and /z/, probably because of its short duration.

Table 4.7: Inventory of all cases of devoicing. Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. After Docherty (1992).

	Word - Initial	Word - Final	All Pos.	
/v/	14/68 (21%)	12/65 (19%)	26/133 (20%)	Devoiced
	16/68 (24%)	20/65 (31%)	36/133 (27%)	Partially Devoiced
	38/68 (56%)	33/65 (51%)	71/133 (53%)	Voiced
/ð/	4/21 (19%)	5/18 (28%)	9/39 (23%)	Devoiced
	-	7/18 (39%)	7/39 (18%)	Partially Devoiced
	17/21 (81%)	6/18 (33%)	23/39 (59%)	Voiced
/z/	3/41 (7%)	17/113 (15%)	20/154 (13%)	Devoiced
	8/41 (20%)	68/113 (60%)	76/154 (49%)	Partially Devoiced
	30/41 (73%)	28/113 (25%)	58/154 (38%)	Voiced
All Fric.	21/130 (16%)	34/196 (17%)	55/326 (17%)	Devoiced
	24/130 (19%)	95/196 (49%)	119/326 (37%)	Partially Devoiced
	85/130 (65%)	67/196 (34%)	152/326 (47%)	Voiced

Soli (1982) reported that most of the tokens of the fricative in the nonsense word /juz/, produced by an American English speaker, were devoiced and exhibited a brief interval of voicing in the VF and FV transitions.

Smith (1995, 1997) also observed high percentages of devoicing for American English /z/: 47% devoiced; 36% partially devoiced; 17% voiced. She also noted a large variability amongst speakers. Devoicing was least likely when /z/ was followed by a vowel and most likely at the end of a sentence or in a syllable coda. All sentence - final /z/s were completely devoiced, more than at the end of syllables. There was more devoicing in syllable - initial, word - medial position than in word - initial position. /z/ was devoiced more often word - finally than at the middle of a word. /z/ was distinct from /s/ occurring in the same context, regardless of whether the /z/ was voiced or devoiced. Vowels were significantly longer before /z/ than before /s/, a durational difference which might aid the perceptual distinction between /z/ and /s/ in the absence of vocal fold vibration. /z/ was influenced by the preceding as well as following context, and the likelihood of devoicing increased under voiceless context. There was no consistent pattern across speakers and different /z/ tokens as to the effect of stress and context on devoicing.

The study of Pirello et al. (1997) investigated whether or not systematic patterns of voicing could be identified as a function of phonetic context for American English fricatives and whether or not an acoustic property for voicing in fricatives, that remained stable despite various types of variability, could be identified. There was a greater preponderance of voicing throughout for the labiodental fricatives: /v/ – 5% devoiced, 35% partially devoiced and 60% voiced; /z/ – 20% devoiced, 40% partially devoiced and 40% voiced. Although contextual influences emerged, they did not necessarily occur in the majority of the utterances, and there was variation amongst different speakers.

4.3.2 Results of Devoicing Analysis Using the Manual Criterion

The fricatives in nonsense words from Corpus 2 were analysed using the manual criterion for devoicing. Results show a very high percentage of devoiced examples for speakers LMTJ, CFGA and ISSS, as can be seen in the inventory presented in Table 4.8. Speaker ACC voiced most of her tokens, possibly as a result of her careful, even somewhat unnatural articulation of the fricative in the nonsense words. Speaker ACC reported having used a "dictation style" when reading the nonsense words, because she did not have any reference in the Portuguese language as to how to pronounce them in a "conversational style". She also had a different background from the other three speakers, which might account for a more precise control of voicing: she had singing training and was a secondary school teacher.

Other particular characteristics of Speaker ACC include: most examples of voiced fricatives present a very low laryngograph signal amplitude during the frication interval. For some of the nonsense words with voiceless fricatives /f, s, ʃ/, the laryngograph signal amplitude is zero during the VF transition, but the vowel is still voiced, as can be seen from the acoustic signal. Some vowels devoice, as judged from the laryngograph signal, before the start of the vowel-to-unvoiced fricative transition, which could mean that the larynx moves vertically. The amplitude of the laryngograph signal for voiced fricatives increases to "vowel level" before the fricative ends. Although the other speakers were faced with the same problems when reading the nonsense words, they seem to have dealt with it in a more natural way.

Table 4.8: Inventory of all cases of devoicing in Corpus 2 (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. The partially devoiced category has not been used to classify the examples of Speaker LMTJ (only the devoiced and voiced categories have been used).

Speaker LMTJ			
	Devoiced	Partially Devoiced	Voiced
/v/	64/101 (63.4%)	-	37/101 (36.6%)
/z/	63/116 (54.3%)	-	53/116 (45.7%)
/ʒ/	58/124 (46.8%)	-	66/124 (53.2%)
All Fric.	185/341 (54.3%)	-	156/341 (45.8%)
Speaker CFGA			
	Devoiced	Partially Devoiced	Voiced
/v/	95/145 (65.5%)	30/145 (20.7%)	20/145 (13.8%)
/z/	79/174 (45.4%)	30/174 (17.2%)	65/174 (37.4%)
/ʒ/	124/173 (71.7%)	30/173 (17.3%)	19/173 (11%)
All Fric.	298/492 (60.6%)	90/492 (18.3%)	104/492 (21.1%)
Speaker ACC			
	Devoiced	Partially Devoiced	Voiced
/v/	32/132 (24.2%)	8/132 (6.1%)	92/132 (69.7%)
/z/	50/130 (38.5%)	21/130 (16.2%)	59/130 (45.4%)
/ʒ/	16/120 (13.3%)	26/120 (21.7%)	78/120 (65%)
All Fric.	98/382 (25.7%)	55/382 (14.4%)	229/382 (60%)
Speaker ISSS			
	Devoiced	Partially Devoiced	Voiced
/v/	60/108 (55.6%)	7/108 (6.5%)	41/108 (38%)
/z/	64/118 (54.2%)	22/118 (18.6%)	32/118 (27.1%)
/ʒ/	75/108 (69.4%)	11/108 (10.2%)	22/108 (20.4%)
All Fric.	199/334 (59.6%)	40/334 (12%)	95/334 (28.4%)

In a preliminary study of the influence of vowel context in the devoicing of Corpus 2 fricatives, speakers LMTJ and ACC have been analysed, with the results presented in Figures 4.16 and 4.17. No consistent pattern can be observed in the two graphs, so this study was not pursued for speakers CFGA and ISSS. Smith (1997) and Pirello et al. (1997) had also previously observed that there is no clear effect of vowel context on devoicing.

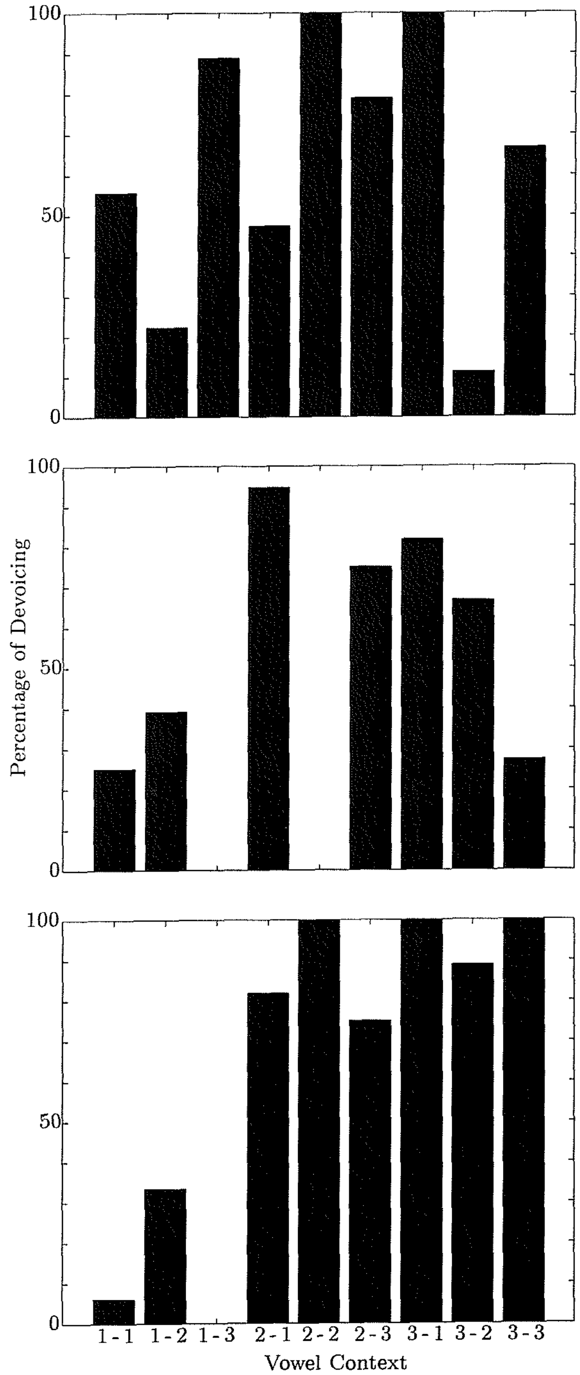

Figure 4.16: Percentage of totally devoiced fricatives in different /pV$_1$FV$_2$/ vowel contexts: 1 – /ɨ/; 2 – /ɐ/; 3 – /u/ (e.g. 1-1 – /pɨvɨ/; 1-2 – /pɨvɐ/). Top: fricative /v/; Middle: fricative /z/; Bottom: fricative /ʒ/. There were no examples of /pɐzɐ/ (group 2-2). Corpus 2 (Speaker LMTJ).

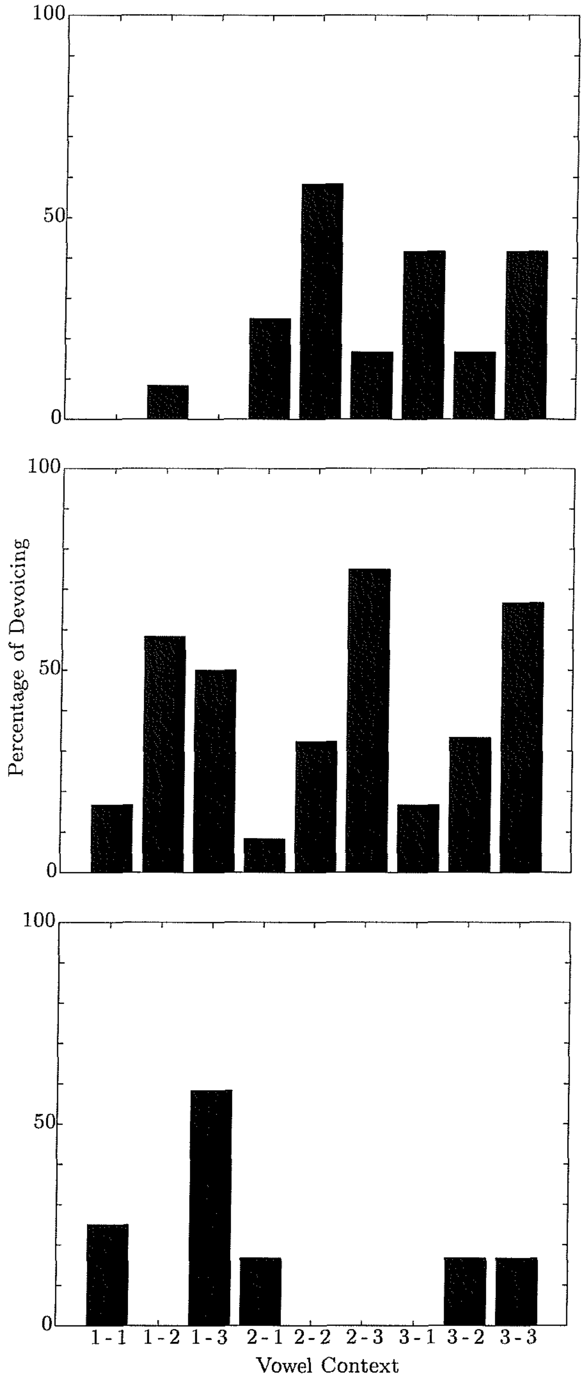

Figure 4.17: Percentage of totally devoiced fricatives in different /pV_1FV_2/ vowel contexts: 1 – /ɨ/; 2 – /ɐ/; 3 – /u/ (e.g. 1-1 – /pɨvɨ/; 1-2 – /pɨvɐ/). Top: fricative /v/; Middle: fricative /z/; Bottom: fricative /ʒ/. Corpus 2 (Speaker ACC).

A complete inventory of devoiced, partially devoiced and voiced examples in Corpus 3 is presented for all four speakers in Tables D.1 to D.4 of Appendix D. Results for all four subjects showed that 55% (70 out of 127) of the examples of fricative /v/ were totally devoiced (see Section 3.3 for devoicing criteria). 74% (79 out of 107) of the examples of fricative /z/ were totally devoiced. 86% (92 out of 107) of the examples of fricative /ʒ/ were totally devoiced. Most word-final fricative examples (93% – 55 out of 59) were totally devoiced, and the percentage of devoicing increased as the place of articulation moved posteriorly.

Veatch's (1989) study of American English fricatives showed that word-final fricatives devoiced 25-100% of the time, depending on context. Although devoicing in our Corpus 3 ranges from 9 to 100%, word-final fricatives devoice 93% of the time.

A different carrier sentence (*Diga ..., bem dito* /'digɐ ... bɐ̃ 'ditu/) was used in the second recording session of Speaker LMTJ, to test the influence of the phoneme that follows the word where the fricative is contained. Although it was expected that the voiced plosive /b/ might result in less devoicing of the target word's final fricative, this was not borne out. Essentially the same amount of devoicing occurred.

In a preliminary study, similar to that reported previously for Corpus 2, the influence of vowel context on devoicing was analysed for a limited number of words from Corpus 3 (Speakers LMTJ and ACC) that follow the pattern /V_1FV_2/. Results are shown in Figures 4.18 and 4.19, where V_1 and V_2 are vowels which belong to one of the groups: group 1 – /i, ɨ, e/; group 2 – /ɛ, ɐ, a/; group 3 – /ɔ, o, u/. Again, there does not seem to be any particular vowel context influencing the voicing of /v, z, ʒ/. Therefore, this study has not been extended to speakers CFGA and ISSS. Both the results of Corpus 2 and Corpus 3 show that there is no effect of vowel context on devoicing of Portuguese fricatives, just as previously observed by Smith (1997) and Pirello et al. (1997) for the English language.

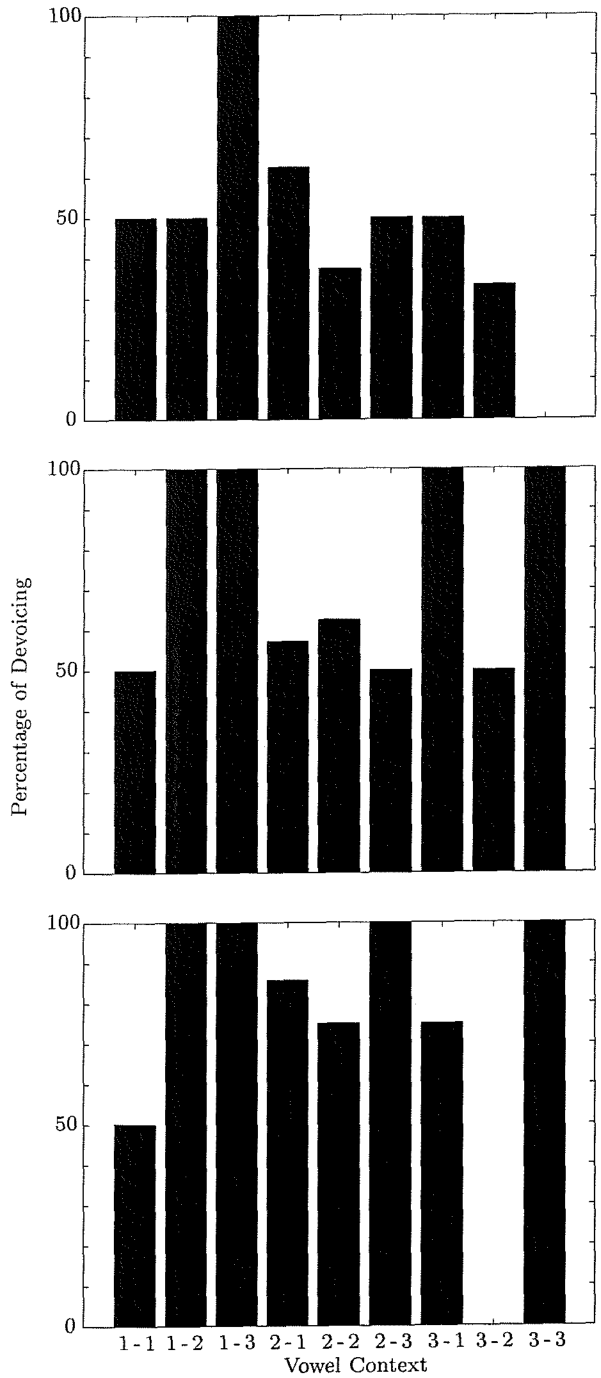

Figure 4.18: Percentage of totally devoiced fricatives in different /V_1FV_2/ vowel contexts: 1 – /i, ɨ, e/; 2 – /ɛ, ɐ, a/; 3 – /ɔ, o, u/ (e.g. 1-1 – /ivɨ/; 1-2 – /ivɐ/). Top: fricative /v/; Middle: fricative /z/; Bottom: fricative /ʒ/. There were no examples of fricative /v/ in 3-3 vowel context. Corpus 3 (Speaker LMTJ).

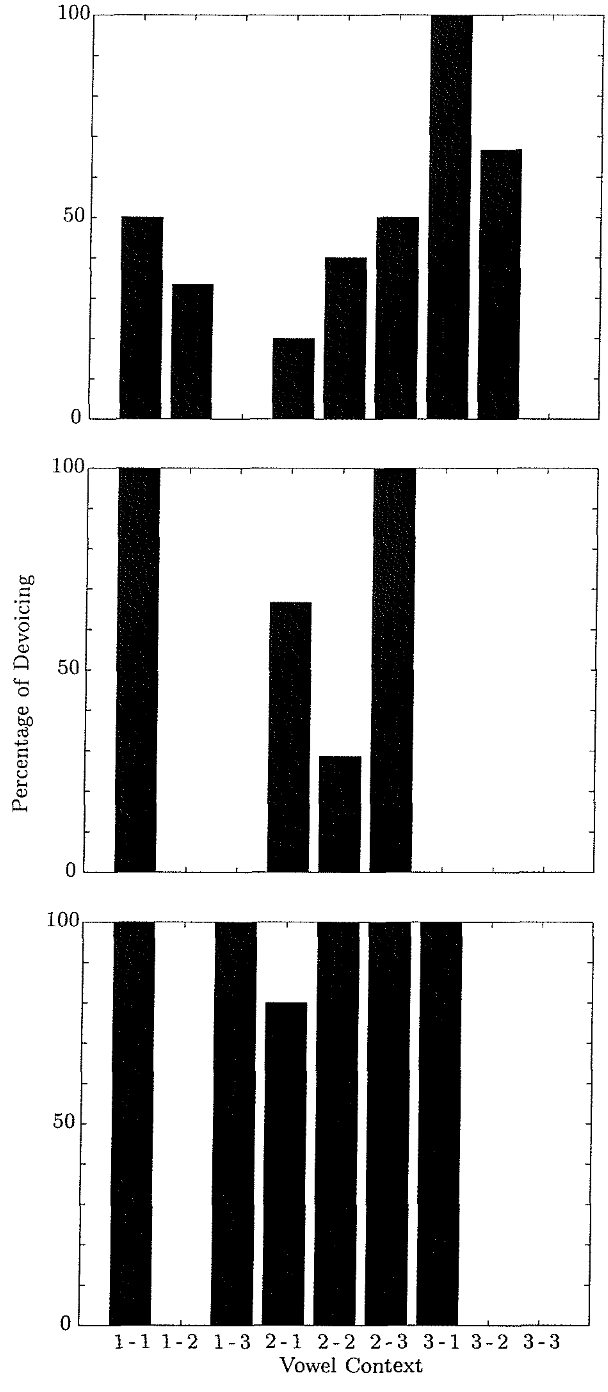

Figure 4.19: Percentage of totally devoiced fricatives in different /V_1FV_2/ vowel contexts: 1 – /i, ɨ, e/; 2 – /ɛ, ɐ, a/; 3 – /ɔ, o, u/ (e.g. 1-1 – /ɨvɨ/; 1-2 – /ɨvɐ/). Top: fricative /v/; Middle: fricative /z/; Bottom: fricative /ʒ/. There were no examples of fricative /v/ in 1-3 and 3-3 vowel contexts, of fricative /z/ in 1-3, 3-1 and 3-3 vowel contexts, and of fricative /ʒ/ in 1-2, 3-2 and 3-3 vowel contexts. Corpus 3 (Speaker ACC).

The results shown in Tables D.5 to D.8 indicate that in Corpus 4: 44% (77 out of 177) of the examples of fricative /v/ were totally devoiced; 78% (86 out of 110) of the examples of fricative /z/ were totally devoiced; 71% (89 out of 126) of the examples of fricative /ʒ/ were totally devoiced. The Corpus 4 fricatives devoiced mostly word-finally, but less often than in Corpus 3: word-initial – 97/157 = 62% (in Corpus 3, $x_1/y_1 = z_1\%$); word-medial – 111/195 = 57% (in Corpus 3, $x_2/y_2 = z_2\%$); word-final – 44/61 = 72% (in Corpus 3, $x_3/y_3 = z_3\%$). As can be seen from the listings of Corpus 4 sentences in Appendix A, some of the fricatives that have been classified as word-final are followed by voiced phonemes. Some of the words that follow these fricatives even start with a vowel. This might account for the lower word-final average percentage of devoicing in Corpus 4 when compared with Corpus 3. Indeed, some voiceless fricatives become voiced in Corpus 4, likely as a result of cross-word coarticulation: eleven tokens of word-final /ʃ/ were produced as [ʒ] by speakers LMTJ and ACC when followed by a word starting with a voiced phoneme ([d] or [m]).

There is a slightly lower number of devoiced examples of /ʒ/ than /z/, which contradicts the very clear results of Corpus 3 fricatives (in which the percentage of devoiced examples decreases as the place of articulation moves anteriorly). One possible explanation could be that /ʒ/ is produced in a more anterior place in continuous speech than in isolated word production. This hypothesis can only be confirmed with additional articulatory data, which is planned as future work.

In summary, overall results from the analysis of devoicing in Corpora 2, 3 and 4, using the manual criteria, show that more than 50% of the fricatives devoice (see Figure 4.20) for all speakers except for Speaker ACC, who has a very low percentage of devoiced tokens in Corpus 2 (see Figure 4.21). There is more final devoicing in Corpus 3 than in Corpus 4 (see Figure 4.22), but the real words of Corpus 4 do exhibit more than 50% total devoicing for all subjects. Devoicing rate differs among the three fricatives, and among Corpora 2, 3 and 4. There is no apparent pattern.

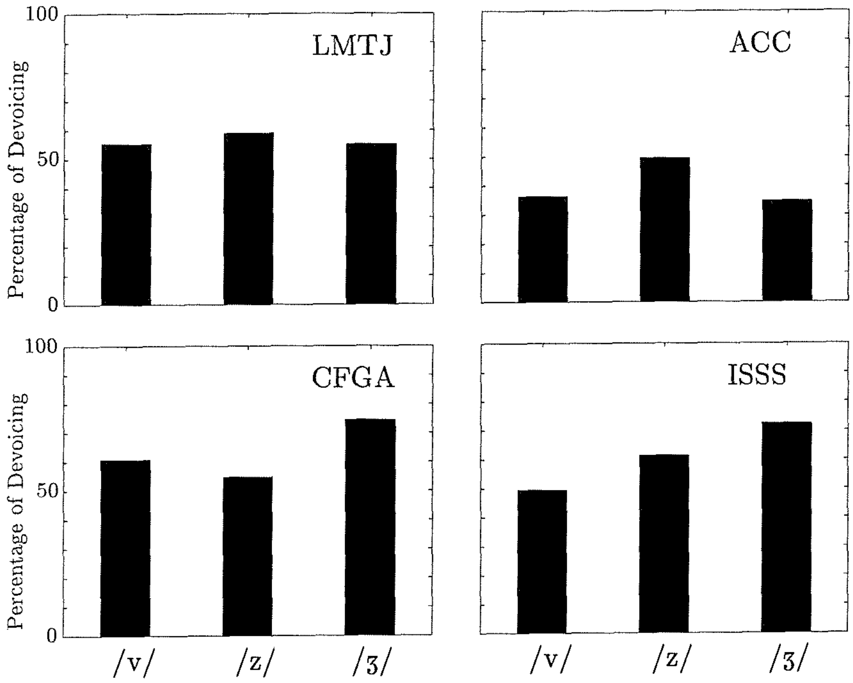

Figure 4.20: Percentage of total devoicing by fricative for each subject; Corpus 2, 3 and 4 combined.

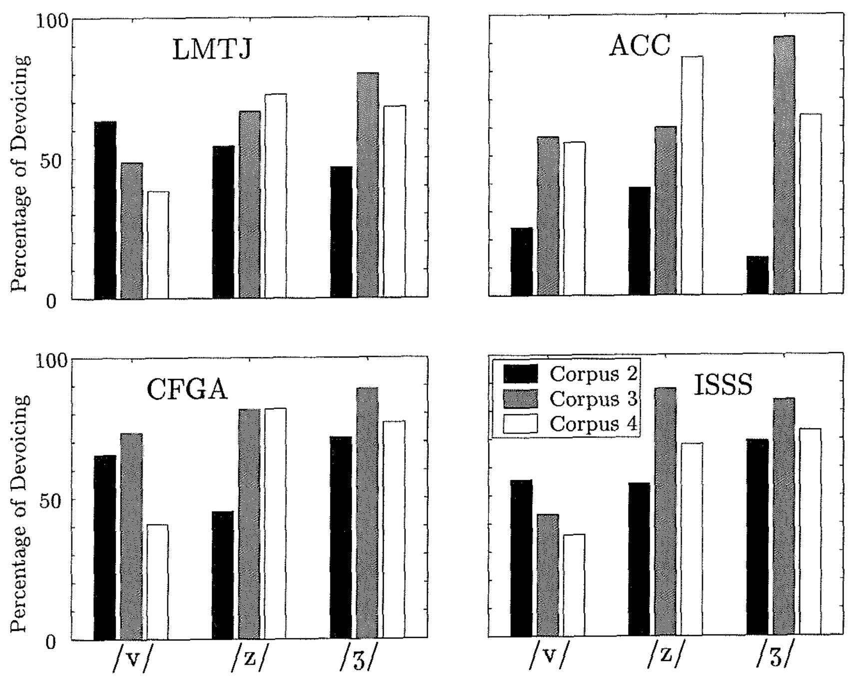

Figure 4.21: Percentage of total devoicing by corpus for each subject.

Figure 4.22 shows the percentage of devoicing as a function of position within a word, and relates it to syllable stress. There is a significant increase in devoiced examples from word-initial, through word-medial to word-final positions, for Corpus 3 fricatives: word-initial – 75/127 = 59%; word-medial – 111/155 = 72%; word-final – 55/59 = 93%, across all subjects. Although we might expect that voicing contrast weakens in unstressed syllables and, thus, unstressed fricatives would devoice disproportionately, the opposite appears to be the case. Our totals for all fricatives of Corpora 3 and 4, all subjects, are: /v/ – 48% (147 out of 304) devoiced and 12% (37 out of 304) partially devoiced; /z/ – 77% (166 out of 217) devoiced and 12% (26 out of 217) partially devoiced; /ʒ/ – 78% (181 out of 233) devoiced and 13% (30 out of 233) partially devoiced.

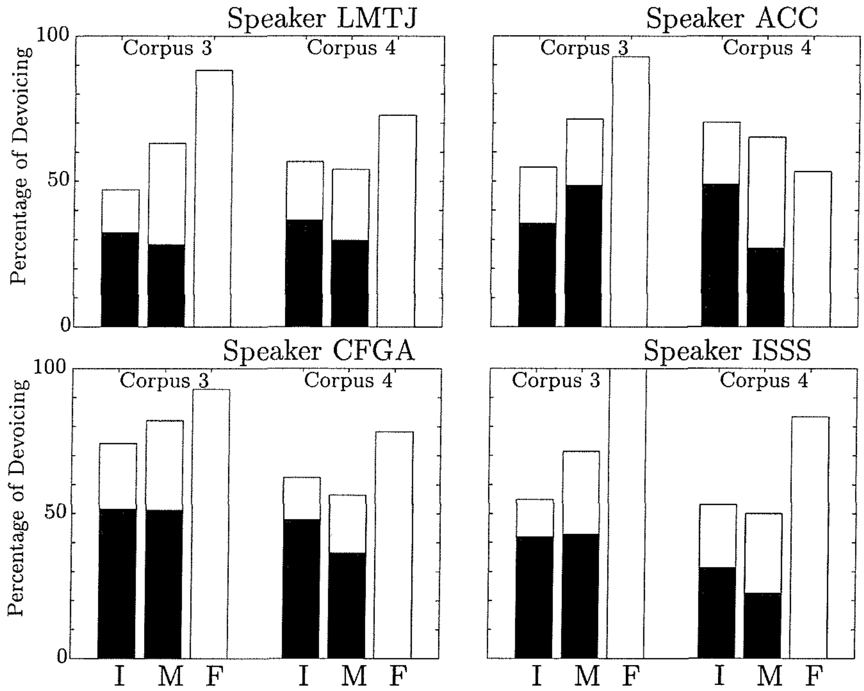

Figure 4.22: Percentage of total devoicing by position in word, word-initial (I), word-medial (M) and word-final (F), fricatives /v, z, ʒ/ combined. The black portion of each bar in the graph corresponds to fricatives in a stressed syllable and the white portion to fricatives in an unstressed syllable. There are no Portuguese fricatives in final stressed position.

These results can be compared to those of Smith (1997) and Pirello et al. (1997) for American English. Smith (1997) studied only /z/ in a range of contexts and measured 47% devoiced and 36% partially devoiced. Pirello et al.

(1997) studied /v, z/ in nonsense words (fricatives in initial and stressed position) and measured, respectively, 5%, 20% devoiced, and 35%, 40% partially devoiced. The comparable figures for initial stressed /v, z/ in Corpora 3 and 4 are 32%, 55% devoiced and 16%, 23% partially devoiced. It thus appears that there is more devoicing in Portuguese than in American English.

4.3.3 Evaluation of the Automatic Devoicing Criterion

The ratio of variances ($r_{\sigma^2} \geq 15$), described in Section 3.3.1, was used as the criterion for devoicing for the Corpus 3 and 4 fricatives of Speaker LMTJ. Results are as shown in Tables 4.9 and 4.10. There are some examples which are classified differently from the manual criterion (see Section 3.3.1) shown in Tables D.1 and D.5. Still, the percentage of examples from Corpus 3 which were classified in the same category using the two methods is quite high: /v/ – 86.1% (31 out of 36), /z/ – 93.3% (28 out of 30), and /ʒ/ – 83.3% (25 out of 30). The percentage of "correctly classified" examples from Corpus 4 was: 79% (27 out of 34) of /v/; 77% (17 out of 22) of /z/; and 64% (14 out of 22) of /ʒ/.

Table 4.9: Inventory of all cases of complete devoicing (using the automatic criterion). Values given are in the form x/y, where x = number of devoiced examples, and y = total number of examples. Corpus 3 (Speaker LMTJ).

	Word - Initial	Word - Medial	Word - Final	All Pos.
/v/	9/14 (64.3%)	8/13 (61.5%)	8/9 (88.9%)	25/36 (69.4%)
/z/	6/10 (60%)	15/17 (88.2%)	3/3 (100%)	24/30 (80%)
/ʒ/	5/10 (50%)	13/15 (86.7%)	4/5 (80%)	22/30 (73.3%)
All Fric.	20/34 (58.8%)	36/45 (80%)	15/17 (88.2%)	71/96 (74%)

Table 4.10: Inventory of all cases of complete devoicing (using the automatic criterion). Values given are in the form x/y, where x = number of devoiced examples, and y = total number of examples. Corpus 4 (Speaker LMTJ).

	Word - Initial	Word - Medial	Word - Final	All Pos.
/v/	3/14 (21.4%)	7/18 (38.9%)	0	10/34 (29.4%)
/z/	3/8 (37.5%)	8/10 (80%)	3/4 (75%)	14/22 (63.6%)
/ʒ/	4/8 (50%)	4/9 (44.4%)	3/5 (60%)	11/22 (50%)
All Fric.	10/30 (33.3%)	19/37 (51.4%)	6/11 (54.6%)	35/78 (44.9%)

For Corpus 3 most of the discrepancies result from cases on the partially devoiced / completely devoiced borderline, giving promise that this automatic measure can be reliably used in the future. Some examples present a few peaks in the laryngograph waveform, which contribute to a larger variance than initially expected. Whether these peaks are included in the fricative or in the adjacent vowels depends on the criteria used for segmentation. There are also some examples in Corpus 3 manually classified as voiced but with a ratio of variances greater than 15. Although there is voicing throughout the whole frication interval, the amplitude of the laryngograph signal during the fricative is much lower than during the VF transition. Most examples present a significant amplitude reduction of the laryngograph signal for the duration of the voiced fricative.

A total of 39 examples from Corpus 4 was misclassified as voiced because there was no VF transition or there was devoicing during the VF transition (51% – 20 out of 39), and because there were a few cycles of the laryngograph during the production of the fricative (41% – 16 out of 39). The remainder of misclassified examples (8% – 3 out of 39) resulted from a dc drift in the laryngograph signal for the duration of the fricative.

The r_{σ^2} metric was also successful when used for the unvoiced fricatives /f, s, ʃ/ of Corpus 3 and 4. The percentage of "correctly classified" examples of Corpus 3 was: /f/ – 96.2% (25 out of 26); /s/ – 85.2% (23 out of 27); and /ʃ/ – 93.8% (30 out of 32). For Corpus 4 the results were: 81% (13 out of 16) of /f/; 55% (16 out of 29) of /s/; 85% (17 out of 20) of /ʃ/.

4.3.4 Analysis of Variance of Devoicing

One-way analysis of variance (ANOVA) was used to study the effects of the independent variables (factors) speaker (LMTJ, CFGA, ACC and ISSS), place of articulation (labiodental, alveolar and postalveolar) and position in word (word-initial, word-medial and word-final) on the dependent variable amount of devoicing of fricatives (using the manual criterion) in Corpus 3 and in Corpus 4. There was a significant effect of the factor `speaker` on the voicing of fricatives /v, z, ʒ/ in Corpus 3 ($F(3, 337) = 3.146$, $p = 0.025$), but no significant effect in Corpus 4 ($F(3, 409) = 0.890$, $p = 0.446$). The value of F in Corpus 3 is inaccurate because the Levene test was significant.

There was a significant effect of the factor `place` on the voicing of fricatives both in Corpus 3 ($F(2, 338) = 19.652$, $p < 0.001$) and Corpus 4 ($F(2, 410) = 32.393$, $p < 0.001$). Both values of F are inaccurate because the Levene test was significant. There was a significant linear trend (Corpus 3 – $F(1, 338) = 21.182$, $p < 0.001$; Corpus 4 – $F(1, 410) = 42.167$, $p < 0.001$). In Corpus 3, as the place of articulation moved further back the amount of devoicing increased proportionately.

There was a significant effect of the factor `position in word` on the voicing of fricatives in Corpus 3 ($F(2, 338) = 10.983$, $p < 0.001$), but no significant effect in Corpus 4 ($F(2, 410) = 2.164$, $p < 0.116$). Both values of F are inaccurate because the Levene test was significant. There was a significant linear trend in Corpus 3 ($F(1, 338) = 13.285$, $p < 0.001$) indicating that as the position of the fricative moves from initial, through medial, to final word position, the amount of devoicing increased proportionately.

4.4 Duration and Devoicing Correlations

Evidence of correlation between duration and devoicing has been reported by Smith (1997) in a study of four American English speakers. The mean duration of /s/ was 101 ms and for /z/ the mean durations grouped into voicing categories were: 81 ms (devoiced), 61 ms (partially devoiced) and 64 ms (voiced). However, in a previous study by Crystal and House (1988), results show no clear correlation between devoicing and duration, as can be seen from the overlap of the probability density distribution curves of duration shown in Figure 4.1. In the present study there is also no consistent pattern between the percentage of devoiced tokens and the average duration of the nine tokens used to ensemble average the spectra of fricatives from Corpus 2, as shown in Figure 4.23.

Figures 4.24 to 4.27 present the average durations of the fricatives /v, z, ʒ/ from Corpus 3 and relate them to devoicing. The manual criterion has been used to classify the fricatives, considering only totally devoiced examples. When a fricative devoices its FV transition, duration diminishes and the duration of fricatives increases (for a few examples the duration remains the same). The VF transition duration is fairly stable.

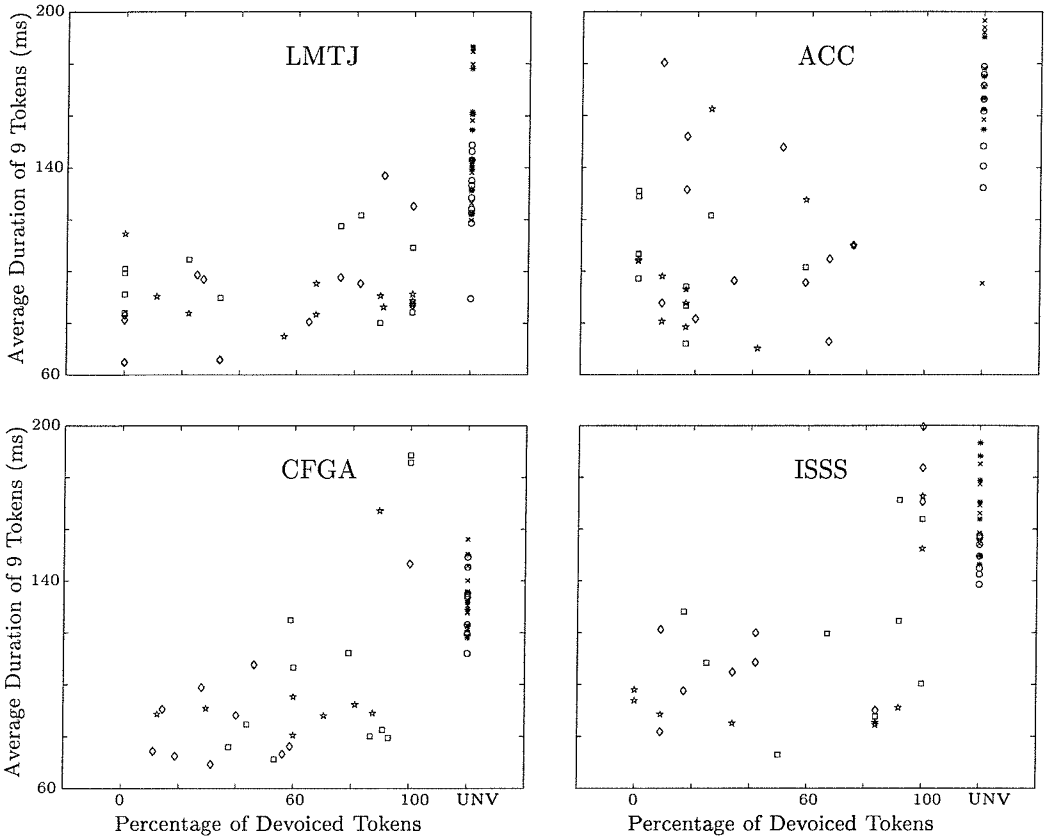

Figure 4.23: Relationship between the percentage of devoiced tokens and duration in Corpus 2. The graphs also include the duration of unvoiced (UNV) fricatives. ○ – /f/, ⋆ – /v/, ∗ – /s/, ⋄ – /z/, × – /ʃ/ and □ – /ʒ/.

For speakers LMTJ and ACC, when Corpus 4 fricatives devoice their duration increases, and the VF and FV transitions are longer (though for a few examples the duration remains the same), as shown in Figures 4.28 and 4.30. This contradicts the result obtained for Corpus 3 where the FV transition duration diminishes when the fricative devoices. There are some examples (e.g., /ʒ/ produced by Speaker CFGA) where the duration of devoiced fricatives is smaller, but this is because almost all fricatives are devoiced, and so the duration of the few voiced examples is clearly atypical.

However, results of linear regression analysis of the amount of devoicing (dependent variable) vs. the duration of fricatives (independent variable), for all speakers, fricatives /v, z, ʒ/ and all positions in word, showed a significant linear trend both in Corpus 3 ($F(1,339) = 49.153$, $p < 0.001$) and in Corpus 4 ($F(1,411) = 27.455$, $p < 0.001$), i.e., as duration of fricatives increased the amount of devoicing increased proportionately.

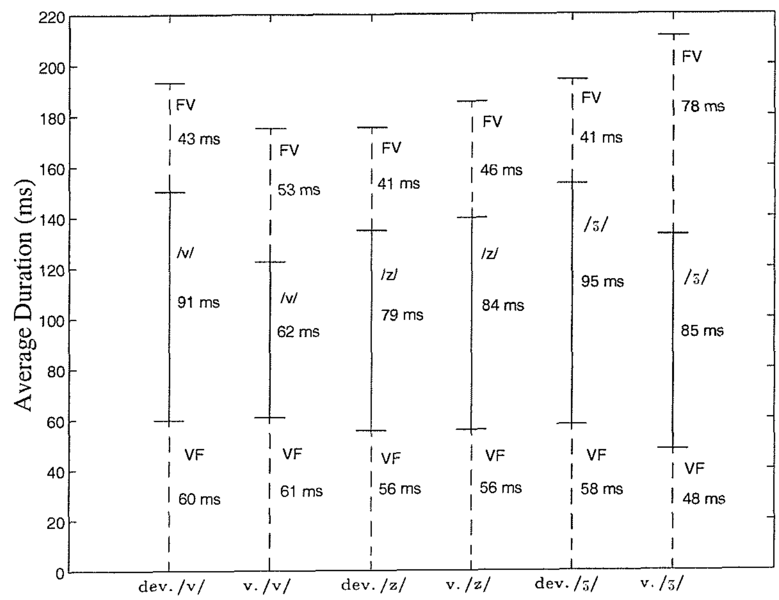

Figure 4.24: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 3 (Speaker LMTJ).

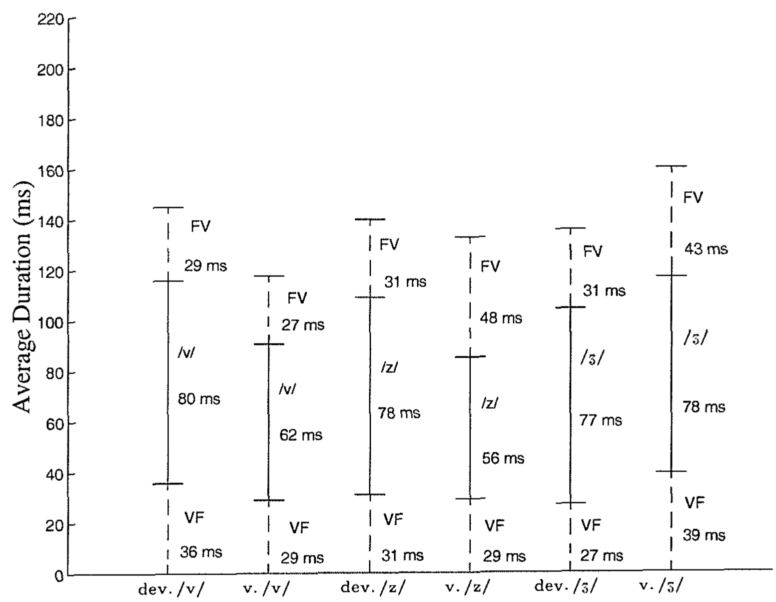

Figure 4.25: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 3 (Speaker CFGA).

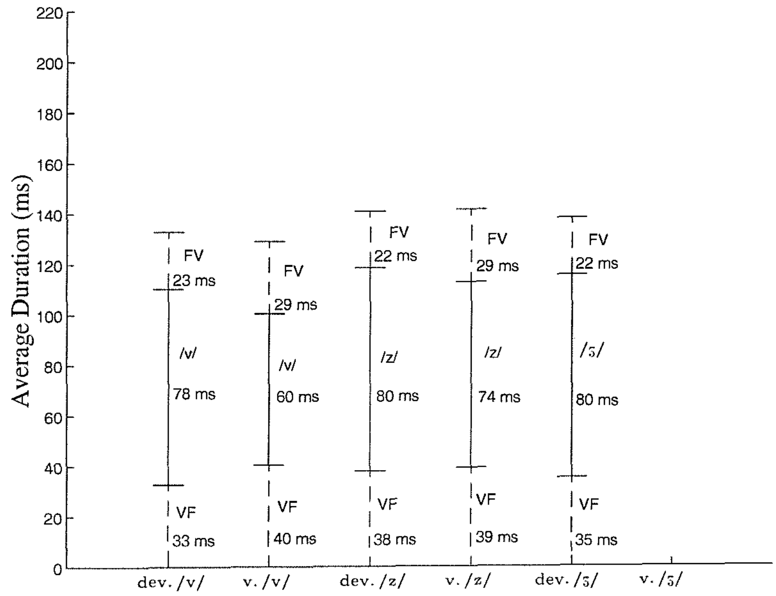

Figure 4.26: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 3 (Speaker ACC).

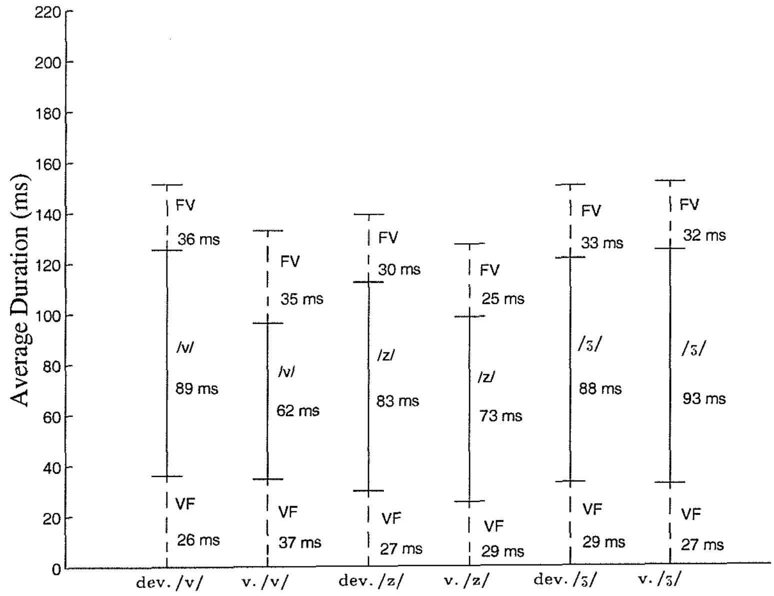

Figure 4.27: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 3 (Speaker ISSS).

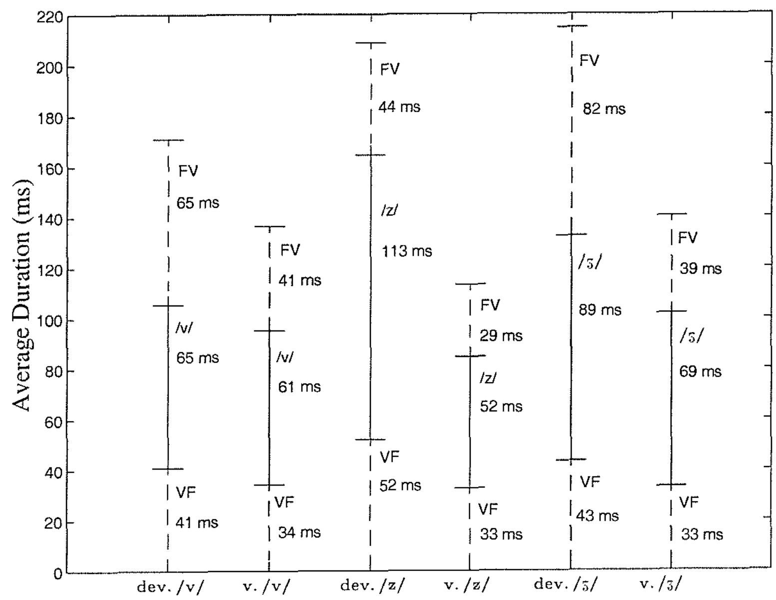

Figure 4.28: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 4 (Speaker LMTJ).

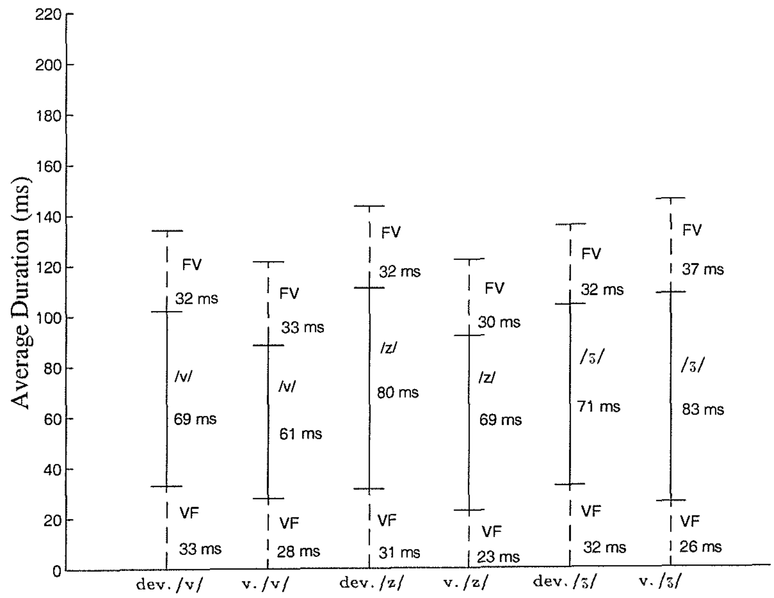

Figure 4.29: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 4 (Speaker CFGA).

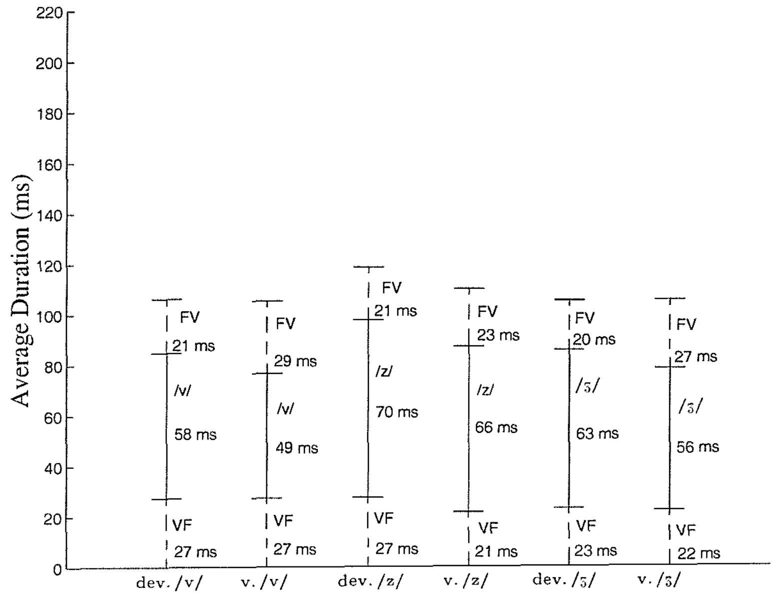

Figure 4.30: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 4 (Speaker ACC).

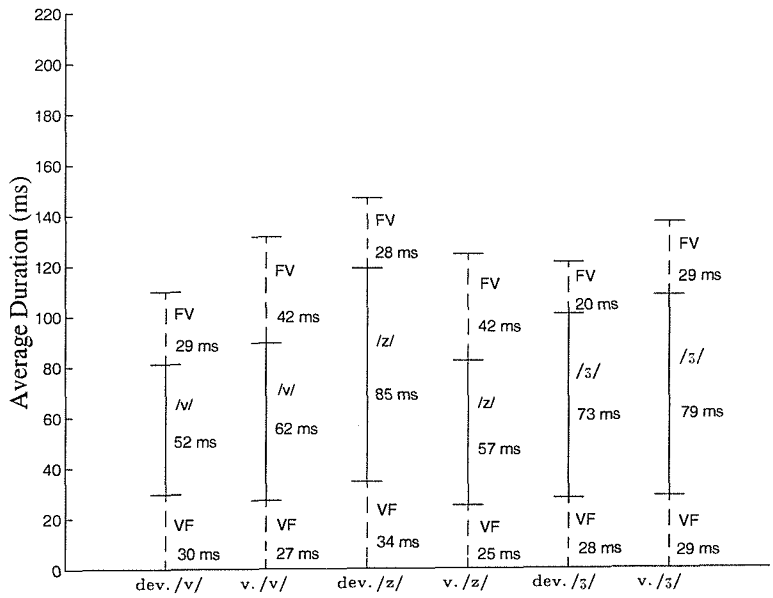

Figure 4.31: Average duration of voiced and devoiced examples of fricatives /v, z, ʒ/, and of VF and FV transitions in Corpus 4 (Speaker ISSS).

4.5 Summary

In this chapter, a discussion of the results from the temporal analysis including durations of the fricatives, and of the VF and FV transitions (Corpus 3 and 4), and a study of devoicing in Corpus 2, 3 and 4, were presented. The results of the automatic measure of devoicing were compared with the manual ones, and the correlation between devoicing and duration investigated.

The mean duration of the unvoiced fricatives is always greater than the mean duration of the voiced fricatives, and the mean duration of the fricative is greater than the mean duration of the VF and FV transitions. These characteristics are not particular of Portuguese, as similar results have been previously reported for the English language. The mean duration of the fricatives in the word corpus (Corpus 3) is quite similar to the mean duration of fricatives from the sentence corpus (Corpus 4).

Devoicing occurs more often in word-final than word-initial position. Devoicing rate by fricative differs between the two measures, and between Corpus 2 and 3, but it is generally very high, especially when compared with studies of other languages. It is thought that this is an important characteristic of European Portuguese, which would have to be incorporated in any production model to obtain more natural-sounding synthetic speech.

Other factors that might be correlated with devoicing were investigated using Corpora 3 and 4 for two of the subjects, LMTJ and ACC. First, there is no consistent pattern between duration of the fricatives /v, z, ʒ/ and percentage of devoicing. Second, there seems to be no particular vowel context that is primarily associated with devoicing. We note that the speakers all produced a large number of repeated tokens of nonsense words in one breath (more than 12 tokens). This high rate of speech (compared with previous recordings of similar corpora by French, American English and German speakers) could be one of the reasons why there are so many devoiced examples.

A preliminary evaluation of the automatic criterion for devoicing showed great potential for the use of this technique in future work. The percentage of examples from Corpus 3 and 4 which were classified in the same category using the two methods (manual and automatic) is quite high (64-93%).

Analysis of variance showed that devoicing was significantly more likely for word-final fricatives and posterior place of articulation. Also, results of linear regression analysis showed that as duration of fricatives increased the amount of devoicing increased proportionately.

Relationships between duration of fricatives and the spectra to be presented in Chapter 5 will be investigated in Chapter 6, and so the results discussed in this chapter constitute a set of time - domain acoustic characteristics that will be related to frequency - domain parameters in the following chapters.

Chapter 5

Results of the Spectral Analysis

5.1 Introduction

In the following sections, a detailed study of the time-averaged spectra of sustained fricatives (Corpus 1a and 1b) and of fricatives in real words (Corpus 3 and 4), together with a study of the ensemble-averaged spectra of nonsense words (Corpus 2), is presented. It includes a discussion of the main spectral peaks and troughs, and some considerations on the amplitude of the spectra, and the influence of vowel context, word position and stress.

Substantial differences are found between spectra of voiced and unvoiced, same-place fricatives: not only are the voiced spectra lower in amplitude, as expected (see Section 3.3), but differences in spectral shape occur. Other comparisons of effort level in sustained fricatives to position in the word are added to a knowledge of the interaction of voice and frication source in Portuguese.

Spectral peaks are due to the poles of the vocal tract frequency response. Spectral troughs are due to the zeros of the vocal tract frequency response. Moving the articulators alters the shape of the vocal tract which in turn changes its frequency response. Shadle and Scully (1995, pp. 59-60) discussed these issues in detail:

> A first-order model of fricative production consists of a noise source independent of the tract transfer function, exciting it. The

> vocal tract resonances, modelled as poles, appear as peaks in the spectrum, but intermediate position of the source in the tract generates zeros as well, which may appear as troughs or may simply reduce the amplitude of the peaks. The zeros occur in three groups: one zero always occurs at very low frequency, causing a low amplitude at low frequencies for unvoiced fricatives; one set occurs at frequencies approximately equal to the frequencies of all back cavity poles, thus effectively cancelling those poles; and one set occurs at frequencies related to the distance between the constriction and the source location. The most obvious spectral peaks and troughs are then the low-frequency zero, the front cavity poles, and the constriction-source zeros.
>
> ...
>
> An overall increase or decrease in spectral amplitude, then, would be most simply ascribed to the source characteristics; small shifts in peak frequencies, to changes in the lengths of the corresponding cavities; radical changes in relative amplitude of peaks, to changes in the distance between the constriction and source location.

Since spectral troughs do not show clearly in the spectra of speech (because of effects such as pole-zero cancellation, window leakage from the analysis technique and noise floor superimposition) spectral peaks tend to be the most prominent feature. They will therefore be referred to as one of the most important acoustic features, and the following three "categories of peaks" as illustrated in Figure 5.1, will be considered: peaks, medium bandwidth peaks and broad peaks.

The term bandwidth is usually used to describe a single pole, but the medium bandwidth peaks observed in the fricative spectra probably result from more than one pole and zero. Peaks and troughs can be observed in the spectra which are related to underlying poles and zeros, resulting in an extremely complex transfer function. Shifts in any pole or zero frequency affect all peaks and troughs, sometimes substantially (Stevens 1998, pp. 130-137), and radiation impedance increases bandwidth as frequency increases, but much more for front cavity resonances (Stevens 1998, pp. 152-156). The fact that there is noise excitation during the production of fricatives increases the difficulty of determining cavity affiliation of peaks and troughs (Shadle et al. 1991). However, peaks and troughs which appeared consistently were used to describe the spectra of Portuguese fricatives.

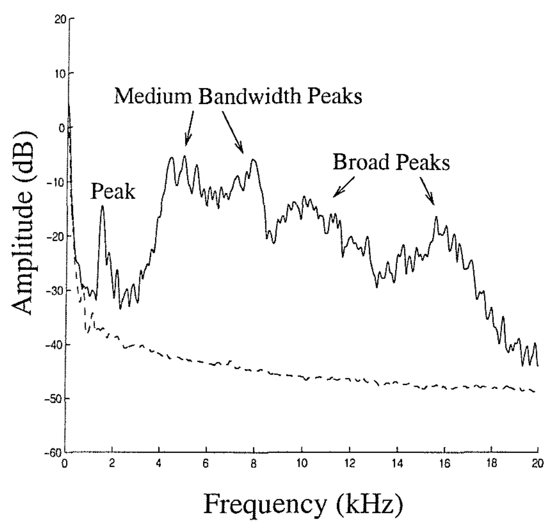

Figure 5.1: Spectrum of /s/ from ['pusɨ], Corpus 2 (Speaker ACC).

Due to the extreme difficulty of picking the troughs and peaks "automatically", all of the data presented in the following sections resulted from a careful examination of the fricative spectrum and a manual measurement of peak and trough frequencies and amplitudes. This time-consuming research method was deemed necessary for a reliable characterization of Portuguese fricatives in the frequency domain, and of crucial importance to the parameterisation of the spectra in Chapter 6. The data presented in Section 5.3 are the result of a partial, but significant and reliable, analysis of the fricative spectra: Speaker LMTJ – Corpus 1a, 1b, 2, 3 and 4; Speaker CFGA – Corpus 1a and 1b; Speaker ACC – Corpus 1a, 1b, 2 and 4; ISSS – Corpus 1a and 1b. But before presenting these data, we consider the results of past analysis of fricatives.

5.2 Previous Spectral Studies of Fricatives

One of the first detailed studies of the spectra of fricatives was that of Hughes and Halle (1956). Their subjects were two male and one female American English speakers. The spectra of /s, z/ had peaks at consistently higher frequencies than /ʃ, ʒ/, and in most of the /f, v/ spectra no prominent peak could be observed (below 10 kHz). /ʃ/ almost never had a strong concentration of energy above 4 kHz, while /s/ and /f/ had their most prominent peaks above this frequency.

Scully (1971) studied the aeroacoustic and spectral characteristics of /s, z/ in nonsense words, spoken by a British English male subject. Spectrograms showed a more intense and longer frication noise in /s/.

Lindblad (1980) studied the acoustics of speech production, the articulation and the perception of Swedish sibilants (see definition in page 33). The spectra of fricatives /s/ and /ʃ/ were relatively flat and with no clear formants, but characterised by a wide band of strong high frequency energy which was limited abruptly below a sharp cut-off frequency (highest in [siː] and lowest in [suː]). In Swedish, there is a wide geographical phonetic variation of /ʃ/, e.g., [ʂ] in central Sweden dialects (a phonemic variant of which has also been reported by Mateus (1996) for European Portuguese as spoken in Viseu).

Manrique and Massone (1981) analysed the acoustic properties and perceptual cues for the recognition of Argentine Spanish fricatives, in different vowel contexts. Spectral peaks were observed at 1.5 - 2.2 kHz (mean = 1.7 kHz) and 8.5 kHz (mean) for /f/, 1.5 kHz (mean) for /ð/, 4.3 - 5.6 kHz (mean = 5 kHz) and 8 kHz (mean) for /s/, 2.2 - 3.4 kHz (mean = 2.8 kHz) and 5 kHz (mean) for /ʃ/, 3 kHz (mean) and 5.5 kHz (mean) for /ʒ/.

Behrens and Blumstein (1988) presented a spectral analysis of American English voiceless fricatives. The context used was nonsense words of the form /FV/, where the fricative F was followed by three vowels V = /i, a, u/. Results revealed an increase in amplitude (10 dB for /f/ and /θ/, and 20 dB for /s/ and /ʃ/) from the beginning to the middle of the fricative. Major peaks occurred in the following frequency ranges: /f/ and /θ/ – 1.8 - 8.5 kHz; /s/ – 3.8 - 8.5 kHz; /ʃ/ – 2.3 - 7 kHz.

Hoole et al.'s (1989) analysis of the spectra of /saːsaː/, /siːsiː/, /ʃaːʃaː/ and /ʃiːʃiː/, produced by two male English speakers, showed a major peak at 5 kHz for /s/, and revealed that the spectra of /ʃ/ had higher overall amplitude than /s/.

Badin's (1989) study of French voiceless fricatives included the spectral analysis of sustained /f, s, ʃ/, and the replication of natural speech using a model of the vocal tract area functions (Badin and Fant 1984). For natural speech spectra, Badin (1989) observed the following: /f/ – peaks around 500 Hz, 1.5 kHz, 2.5 kHz and 3.8 kHz; /s/ – peaks around 500 Hz and 1.8 kHz, and broad peaks around 4.5 kHz and 8 kHz; /ʃ/ – a peak around 400 Hz, a trough around 1 kHz, and broad peaks around 1.8 kHz, 2.5 kHz and 3.5 kHz. Fricative /ʃ/ exhibited the greatest spectrum tilt change, whereas /f/ showed the smallest and least regular changes. This appeared to reflect variations in the source spectrum rather than in the transfer function.

Beautemps et al. (1993, 1995) measured, on a male French subject, a set of midsagittal profiles and the corresponding spectral peaks : /f/ – 397 Hz, 1282 Hz, 2504 Hz and 3665 Hz; /θ/ – 386 Hz, 1456 Hz, 2677 Hz and 3869 Hz; /s/ – 400 Hz, 1500 Hz, 2647 Hz, 4276 Hz and 5062 Hz; /ʃ/ – 450 Hz, 1710 Hz, 2230 Hz and 2952 Hz.

Badin (1991) reported that the /ʃ/ spectra of a female American English speaker exhibited fewer pole/zero pairs in the low frequency region than a male French speaker, because of lesser amount of coupling between front and back cavities (the female speaker produced constriction areas half as large as the male speaker).

Shadle, Badin and Moulinier's (1991) study of the ensemble-averaged spectra of sustained fricatives produced by two speakers included a detailed discussion on cavity affiliation of peaks and troughs, shown in Figure 5.2. Results showed spectral peaks located at the following frequencies:

- /s/ (female American English speaker) – 500 Hz, 1.5 kHz, 2.5 kHz, 4 kHz and 5 - 5.5 kHz;
- /s/ (male French speaker) – 500 Hz, 1.5 kHz, 2.5 kHz, 3.8 kHz and 5.5 kHz;
- /ʃ/ (female American English speaker) – 500 Hz, 1.8 kHz, 2.8 kHz, 3.5 kHz and 5 kHz;
- /ʃ/ (male French speaker) – 500 Hz, 1.5 kHz, 2.5 kHz, 3.8 kHz and 5.5 kHz.

The spectral shape of /ʃ/ had a region of high energy between 1.5 kHz and 6 kHz for the male speaker, and between 2.5 kHz and 7 kHz for the female speaker. There was an abrupt drop in amplitude at 6 kHz in the male speaker's spectra, but the amplitude level fell off steadily for the female speaker. The spectra for the female speaker presented the first four peaks evenly spaced, while the male speaker had the second, third and fourth peaks, clustered together. The lowest-frequency high-amplitude peak was the second peak for male speaker and the third for the female speaker.

Shadle, Dobelke, and Scully (1992) studied the time and ensemble averaged spectra of fricatives in nonsense words produced by a female American

English speaker. They reported fricative peaks in /pafa/ around 1.5 kHz, 2.7 kHz, 4 kHz and 5.2 kHz, as shown in Figure 5.3. The spectra at the centre of fricatives /θ/ and /f/ were very similar, but in the VF and FV transitions overall amplitude was higher and formant frequencies were shifted relative to those of /f/. The authors proposed the following explanation for this phenomenon:

> A possible explanation is that since the tongue tip is required to form a constriction for /θ/ but not for /f/, formants shift during /θ/ transitions. The direction of shift will depend on the vowel. The timing differences in boosting of high-frequency energy are likely due to sequencing differences that result when the tongue has a long versus a short distance to travel.

The fricative in /pasa/ had peaks around 800 Hz, 1.5 kHz, 2.8 kHz, 4.5 kHz, 5.2 kHz and 6.5 kHz. The overall amplitude of /s/ at the low frequencies ($\leq$ 4.5 kHz) was the same at the beginning, middle and end of the fricative. For higher frequencies (5 kHz to 17 kHz), the spectrum at the end of the fricative had the lowest amplitude, at the start it was 20 dB higher than at the end, and at the middle the overall amplitude of /s/ was the highest (5 dB higher than at the start). This study was extended to include also a male French speaker (Shadle, Moulinier, Dobelke, and Scully 1992; Shadle and Scully 1995) with the following spectral characteristics. For /pafa/, peaks were approximately at 800 Hz, 1.8 kHz, 2.5 kHz, 3 kHz, 4 kHz and 5 kHz, and a broad peak around 11 kHz. The same spectral differences between /f/ and /θ/ observed for American English were also noted for French. For /pasa/, peaks were at 500 - 600 Hz, 1.5 - 2 kHz, 2.5 - 2.7 kHz and 3.5 - 4 kHz, and broad peaks at 4.4 - 5.5 kHz, 7.5 - 11.5 kHz and 15 kHz, as shown in Figure 5.4. Fricative /z/ spectra displayed fairly similar patterns, with an overall amplitude approximately 10 dB lower than the corresponding spectra of /s/, as shown in Figure 5.5. For /paʃa/, a peak around 1.8 kHz, and broad peaks around 2.2 kHz, 3.5 kHz, 4.5 kHz and 11 kHz, as shown in Figure 5.6. The overall amplitude of /ʃ/ in /paʃa/, at low frequencies ($\leq$ 4 kHz), was the same at the beginning, middle and end of the fricative. For higher frequencies, the spectrum at the end of the fricative had the lowest amplitude; at the start it was higher than at the end, and at the middle it was the highest.

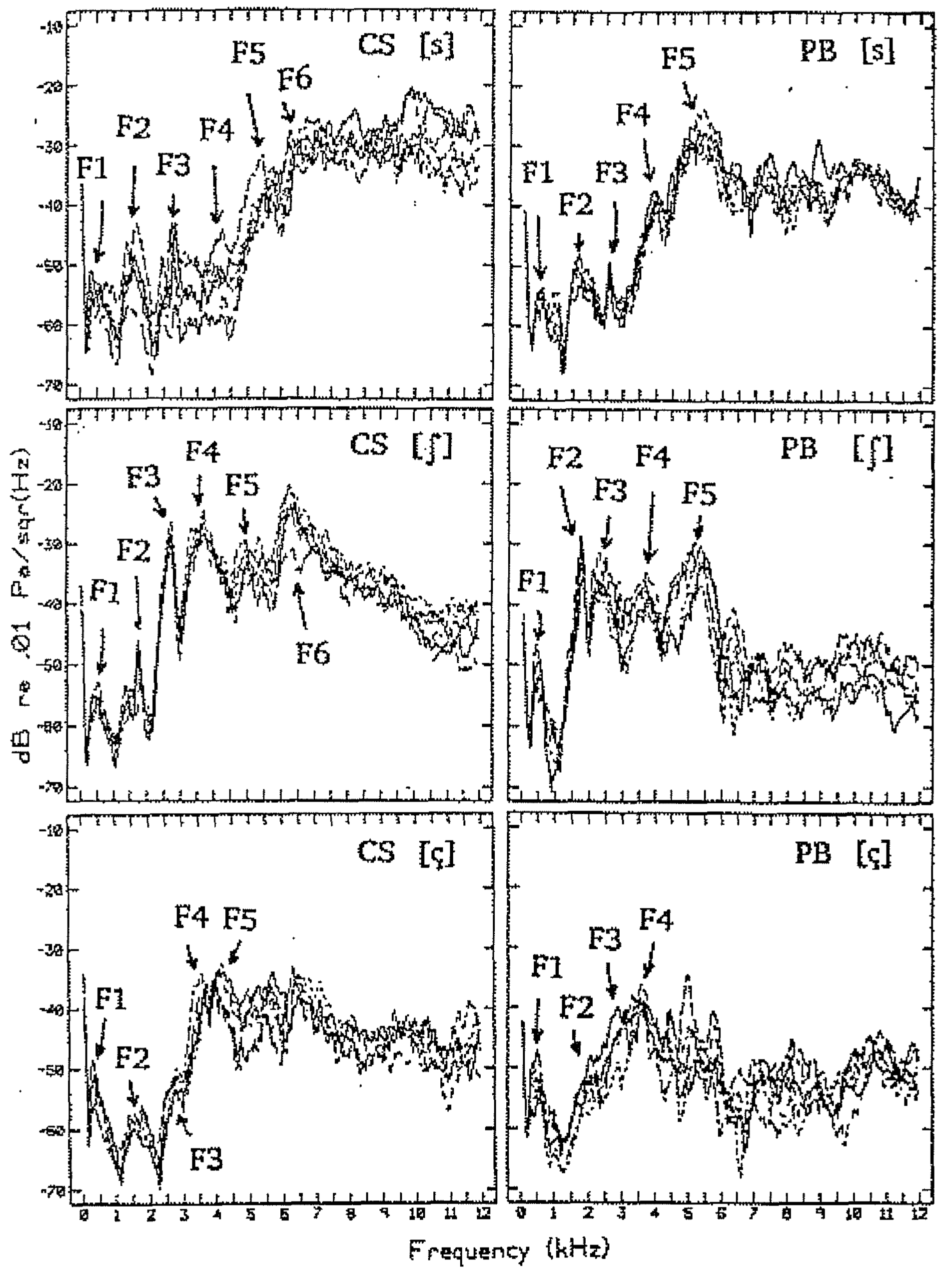

Figure 5.2: Averaged power spectral density of six sustained tokens of fricatives /s, ʃ, ç/ produced by a female American English speaker CS and a male French speaker PB. Formants F_1, F_2, F_3, F_4, F_5 and F_6 were identified by transfer function measurements. From Shadle, Badin, and Moulinier (1991).

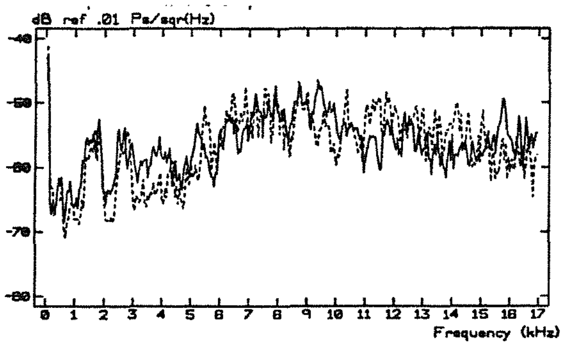

Figure 5.3: Ensemble-averaged spectra at the centre of the fricatives in /pafa/ (solid line) and /paθa/ (dotted line). From Shadle, Dobelke, and Scully (1992).

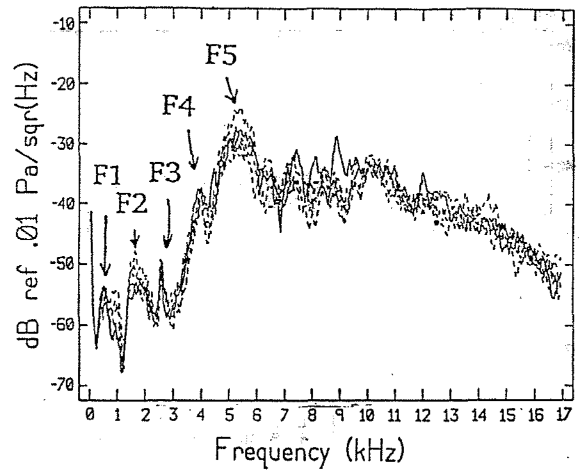

Figure 5.4: Time-averaged spectra of six sustained tokens of fricative /s/ produced by a male French speaker PB. Formants F_1, F_2, F_3, F_4 and F_5 were identified by transfer function measurements. From Shadle and Scully (1995).

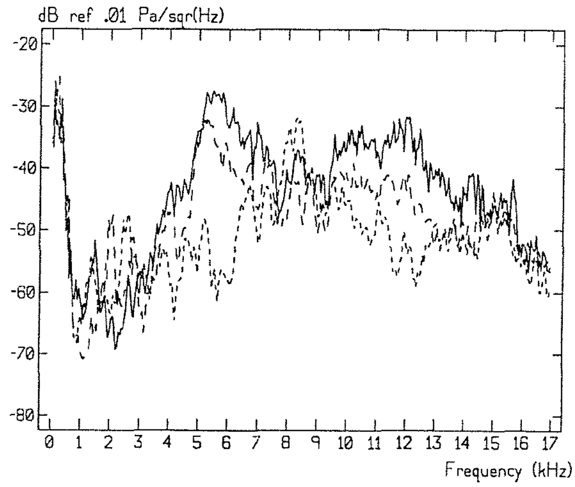

Figure 5.5: Ensemble-averaged spectra of the fricative in /paza/ (solid line), /pizi/ (long dashes) and /puzu/ (short dashes), produced by a male French speaker PB. From Shadle and Scully (1995).

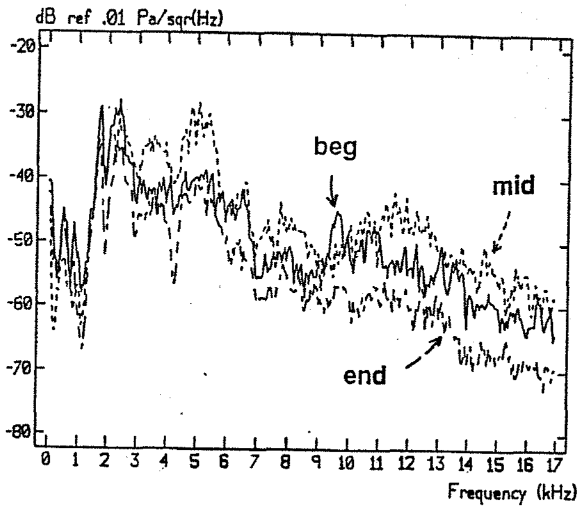

Figure 5.6: Ensemble-averaged spectra at the beginning (beg), middle (mid) and end of the fricative in /paʃa/, produced by a male French speaker PB. From Shadle, Moulinier, Dobelke, and Scully (1992).

Wilde (1993) studied American English fricatives in nonsense words, measuring formant frequencies and onset times, and analysing $F_2 - F_1$ plots. Results showed that voicing onset time and formant structure provided important place information. Wilde (1995a, 1995b) also analysed amplitudes in restricted frequency regions of the fricative and quantified them with respect to the neighboring vowel. Labiodental and dental fricatives had relatively weak and flat spectra, alveolars had a peak above 4 kHz, and postalveolars had a broad peak that fell in the frequency region 2 - 4 kHz, which confirmed results first reported by Hughes and Halle (1956). The variation of the amplitude over the duration of the fricative was calculated by subtracting the amplitude value at the end of the fricative from the amplitude value at the middle of the fricative. For the 1 - 4 kHz range there was greater amplitude for /f/, and the end was stronger than the middle for all fricatives. For /f/ the 4 - 8 kHz band amplitude at the end was greater than the amplitude at the middle, and for /s/ and /ʃ/ the 4 - 8 kHz band amplitude at the end was lower than the amplitude at the middle. The maximum fricative amplitude above 2 kHz, normalized by subtracting the first formant amplitude in the following vowel from the maximum amplitude peak above 2 kHz in the fricative, were 15 - 20 dB higher for /s/ than for /ʃ/. Fricative /f/ showed greater overall amplitude variability than /s/ and /ʃ/, a characteristic that had also been previously observed by Badin (1989) for French. Fricatives were weaker and shorter before reduced vowels, at times appearing stop - like in manner.

Narayanan's (1995) study of American English fricatives, produced by two male and two female native speakers, included a detailed acoustic analysis of time - averaged power spectra. Strident fricatives [1] exhibited a dynamic range about 15 - 30 dB greater than nonstridents. The overall amplitude of voiced fricatives was 5 dB lower than their unvoiced counterparts. There was a considerable inter - speaker variability in the frequency and bandwidth of the main spectral peaks, which was mainly due to differences in the front cavity dimensions across subjects. However Narayanan (1995) could still observe a broad high frequency peak around 10 kHz for labiodentals; a broad peak at 9 - 10 kHz for dentals; a broad peak at 5 - 6.6 kHz and significant secondary peaks at 1.6 - 1.8 kHz, 2.5 - 2.9 kHz and 4.6 - 4.8 kHz for alveolars (with a free zero at 2 - 3.7 kHz arising from the cavity between the source and the oral constriction for /s/); and a broad peak at 5 - 7 kHz and peaks at 1.5 - 1.8 kHz and 2 - 3.5 kHz for postalveolars.

Jongman et al. (2000) studied spectral peak location of fricatives produced by 20 American English speakers in nonsense words. Results showed that the mean frequency of the highest - amplitude spectral peak of fricatives de-

[1] Strident – speech sound produced by a relatively complex stricture and marked by relatively high frequency and intensity (Crystal 1997); /s, z/ and /ʃ, ʒ/ are examples, and in some systems /f, v/. This is a phonetic classification based on the source features of the sound (see also definition of *sibilant* on page 33).

creased as the place of articulation moved back: /f/ – 7.7 kHz; /v/ – 7.5 kHz; /θ/ – 7.8 kHz; /ð/ – 7 kHz; /s, z/ – 6.8 kHz; /ʃ, ʒ/ – 3.8 kHz.

Vowel context has been shown to affect the spectra of fricatives. Komshian and Soli (1981) reported that, for American English alveolar fricatives in nonsense words, the highest amplitude spectral peak was located at approximately 4 kHz in /su/ and at 5 kHz in /si/ and /sa/. The fricative peak in the frequency range 1 - 2.5 kHz in /si/ was 60 - 180 Hz higher than in /sa/ and /su/. Soli (1981) examined the effects of anticipatory vowel coarticulation on the spectra of American English fricatives spoken in isolation and initial nonsense word position. Results from the analysis of the LPC mean spectra revealed:

- /s/ – a peak around 2 kHz; /si/ – peaks around 1.8 kHz, 2.8 kHz and 3.5 kHz; /sa/ – peaks around 1.5 kHz and 3.7 kHz, and a trough around 1 kHz; /su/ – a peak around 1.5 kHz;
- /z/ – a peak around 1.7 kHz; /zi/ – peaks around 1.8 kHz and 3.5 kHz; /za/ – a peak around 3.5 kHz; /zu/ – a peak around 1.6 kHz;
- /ʃ/ – a peak around 2.4 kHz; /ʃi/ – peaks around 1.9 kHz and 2.5 kHz; /ʃa/ – a peak around 1.7 kHz; /ʃu/ – peaks around 1.7 kHz and 2.4 kHz, and a trough around 3 kHz;
- /ʒi/ – peaks around 1.9 kHz and 2.4 kHz; /ʒa/ – peaks around 1.8 kHz and 2.4 kHz; /ʒu/ – peaks around 1.8 kHz and 2.4 kHz, and a trough around 3 kHz.

The spectral peaks of alveolar fricatives were more clearly visible for high vowel contexts, /i/ and /u/, than for the low vowel /a/ context.

Nartey's (1982) study of twelve different world languages revealed that the fricatives most affected by vowel context differed from language to language, and that the effect of /u/ context was not consistent with regards to the various places of articulation, as shown in Tables 5.1 to 5.3. Nevertheless, Trong and Hoole (1993) reported that the frequency of the main spectral peak of French /s/ was shifted down for /u/ vowel context. Wilde (1993) showed that voiceless fricatives were more dependent on vowel context, and that there was less acoustic variability in fricatives in high front vowel context than in back vowel context (Wilde 1995a, 1995b). The 4 - 8 kHz band intensity of alveolar fricatives increased before rounded vowels.

Table 5.1: Peak frequencies in kHz for fricatives /f, v, θ, ð/. For a given place of articulation the same number of columns is used, and if a peak is present its frequency range is given. After Nartey (1982).

	/ifi/	/afa/	/ufu/
Amharic	1.7 - 2; -	1.3 - 1.7; -	1 - 1.3; -
Arabic	1.7 - 2; -	1.5 - 1.7; -	1.7 - 2; -
Hebrew	-; 2.3 - 2.7	-; 2.3 - 2.7	1.1 - 1.3 ; 2.3 - 3.2
Polish	1.7 - 2.3; -	1.3 - 1.7; -	1.1 - 1.3; -
Yoruba	1.7 - 2; -	1.1 - 1.7; -	1.3 - 1.5; -
	/ivi/	/ava/	/uvu/
Hebrew	-	-; 2.3 - 2.7	-
Hopi	-; 1.3 - 2	0.9 - 1.1; -	0.8 - 0.9; -
Papago	-; 1.3 - 1.7	0.9 - 1.3; -	-
Pima	0.9 - 1.1; 1.7 - 2	0.9 - 1.1; 2 - 2.3	-
Polish	-	-; 1.3 - 1.5	-
	/iθi/	/aθa/	/uθu/
Arabic	-; 1.7 - 2; -	0.5 - 0.6; 1.5 - 1.7; -	-; 1.7 - 2; -
	/iði/	/aða/	/uðu/
Pima	0.2 - 0.3; 0.9 - 1.1; 2.7 - 3.7	0.1 - 0.2; 0.9 - 1.1; 2.7 - 3.2	0.2 - 0.3; 1.1 - 1.5; 2.7 - 3.2

Table 5.2: Peak frequencies in kHz for fricatives /s, z/. For a given place of articulation the same number of columns is used, and if a peak is present its frequency range is given. After Nartey (1982).

	/isi/	/asa/	/usu/
Amharic	-; 1.5-2; -; 7.7-9.5	-; 1.5-1.7; 4.4-6.4; -	-; 1.5-1.7; 3.2-4.4; 7.7-9.5
Arabic	-; 2-2.3; 3.7-4.4; -	-; 1.5-2; -; 7.7-9.5	-; 1.5-1.7; -; 6.4-7.7
Hebrew	-; 2.3-2.7; -; 7.7-9.5	-; 2.4-2.7; -; 7.7-9.5	-; -; 3.7-4.4; 7.7-9.5
Hopi	0.8-0.9; 1.7-2; 4.4-5.3; -	-; 1.3-1.5; 4.4-5.3; -	-; 1.5-1.7; 3.7-4.4; -
Japanese	-; 1.5-2; 5.3-6.4; -	-; 1.1-2; -; 7.7-9.5	-; 1.7-2.3; -; 7.7-9.5
Korean	-; -; 4.4-5.3; 6.4-7.7	-; 2-2.3; -; 6.4-9.5	-; 1.7-2.3; 3.7-4.4; -
Navajo	-; 1.7-2; -; 6.4-7.7	-; 1.7-2.3; -; 6.4-7.7	-
Papago	-; 1.7-2; -; -	-; 1.1-1.7; -; 6.4-7.7	-; 1.3-1.7; -; 6.4-7.7
Pima	-; 1.7-2; -; 6.4-7.7	-; 1.7-2; -; 6.4-7.7	-; 1.7-2; 3.2-3.7; 6.4-7.7
Polish	1.1-1.3; 1.7-2.3; -; 7.7-9.5	1-1.1; 1.5-2; -; 6.4-9.5	-; 1.5-1.7; 3.7-4.4; 7.7-9.5
Yoruba	-; 1.7-2; -; 6.4-7.7	-; 1.3-2; -; 6.4-7.7	-; 1.5-1.7; 3.2-3.7; 6.4-7.7
Zuni	-; 2-2.3; -; 6.4-7.7	-; 1.7-2; -; 6.4-7.7	-; 1.7-2; -; 6.4-7.7

	/izi/	/aza/	/uzu/
Amharic	1.5-2; -; 7.7-9.5	1.3-1.7; -; 6.4-9.5	1.3-1.5; 3.7-4.4; 6.4-7.7
Arabic	-	1.3-1.7; -; 6.4-7.7	-
Hebrew	-	2.4-2.7; -; 7.7-9.5	-
Japanese	1.3-1.7; 5.3-6.4; -	1.7-2; -; 6.4-7.7	1.7-2; 5.3-6.4; -
Navajo	-	1.7-2.3; -; 6.4-7.7	-
Polish	1.3-1.5; 2-2.3; 6.4-7.7	1.3-1.7; -; 6.4-9.5	1.3-1.5; 3.2-3.7; 6.4-7.7

Table 5.3: Peak frequencies in kHz for fricatives /ʃ, ʒ/. For a given place of articulation the same number of columns is used, and if a peak is present its frequency range is given. After Nartey (1982).

	/iʃi/	/aʃa/	/uʃu/
Amharic	1.7 - 2; 3.2 - 3.7; -	1.7 - 2; 2.3 - 2.7; -	1.5 - 2; -; -
Arabic	-; 2 - 2.3; 3.7 - 4.4	1.7 - 2; -; 3.7 - 4.4	1.7 - 2; -; -
Hebrew	-; 2.7 - 3.2; -	1.7 - 2.4; 3.2 - 3.7; -	-; 2.3 - 3.2; -
Japanese	1.5 - 2; -; 4.4 - 6.4	1.3 - 1.7; -; 4.4 - 5.3	-; -; 3.7 - 4.4
Navajo	-; 3.7 - 4.4; -	-; 3.2 - 4.4; -	-
Polish	2 - 2.3; 3.7 - 4.4; -	1.5 - 1.7; 3.2 - 3.7; -	1.3 - 1.5; 4.4 - 5.3; -
Yoruba	-; 2.7 - 3.7; -	1.5 - 2; 3.1 - 3.7; -	1.7 - 2; -; -
Zuni	2 - 2.3 ; 3.7 - 4.4; -	-; 3.7 - 4.4; -	-; 2.7 - 3.7; 6.4 - 7.7
	/iʒi/	/aʒa/	/uʒu/
Amharic	-	-; 1.7 - 2; 2.3 - 3.2	-
Navajo	-	-; -; 3.2 - 4.4	-
Polish	-	0.8 - 0.9; 1.5 - 1,7; 3.2 - 3.7	-

Results from Shadle, Moulinier, Dobelke, and Scully (1992) and Shadle and Scully (1995) showed that all the peaks except the second and fourth peaks, and all the broad peaks of fricative /s/ were shifted down in frequency for the nonsense word /pusu/, and the overall amplitude for high frequencies ($\geq$ 4 kHz) was lower, relative to /pisi/ and /pasa/ contexts. The overall amplitude of the spectra of /s/ in /pisi/ and /pasa/ was very similar, but in /pusu/ it was 10 - 25 dB lower, except for the second broad peak around 7.5 kHz. A possible explanation would be a different source mechanism for the fricative in /pusu/, suggestive of a breathy whistle. In Shadle and Mair (1996) the most striking spectral change due to vowel context also occurred for /pusu/: "a fairly narrow high - amplitude peak with "shoulders" considerably lower in amplitude than" in /pasa/ and /pisi/ contexts.

Shadle, Mair, Carter, and Millner (1995) studied the fricative in /pufi/ and /puθi/. They observed that /i/ vowel context moved the place of constriction anteriorly, increasing the front cavity resonances, and that the rounded vowel context /u/ decreased formant frequencies and bandwidths. Shadle, Mair, and Carter (1996, p. 194) also showed that for one of their two subjects (a male French speaker) the frequency of the fricative peak in /pufi/ at 1.5 - 2 kHz, increased from beginning, through middle, to end:

> A spectral peak corresponding in frequency range to the second formant ... is highest in frequency for the [i - i] context, and lowest for the [u - u] context. The higher peak in the [i - i] context is accompanied by the lowest amplitude low - frequency *trough* at approximately 1 kHz; one explanation of this is that the trough corresponds to a zero with a frequency related to source and constriction location, which do not change; the zero becomes more visible.

The spectra of fricatives /f/ and /θ/ were the most variable, and the most noticeably affected by vowel context.

From the results of the studies reported in this section, it can clearly be seen that the detailed spectral characteristics of fricatives are language dependent, but also that there are some effects of context, such as the downward shift of peak frequencies due to rounding, that could be observed for a variety of speech material. The data collected from all the different bibliographical sources referred to in this section will be compared with some specific spectral characteristics observed for Portuguese.

5.3 Results of the Analysis of Spectral Peak and Trough Frequencies, and of Spectral Amplitude – Labiodental, Alveolar and Postalveolar Fricatives

5.3.1 Fricative /f/

Fricative /f/, for all subjects and all corpora, has spectral peaks at 1.4 - 1.9 kHz, 2.2 - 2.8 kHz, 3.1 - 3.9 kHz and 4.1 - 5.3 kHz, and a broad peak at 8 - 11.5 kHz. The first peak and the broad peak tend to be shifted down in frequency to lower ranges for back vowel contexts, an effect reported previously for various other languages (see Literature Review in Section 5.2).

The first peak is located in the same frequency range as the one observed for Argentine Spanish (Manrique and Massone 1981); for Amharic, Arabic, Polish and Yoruba (Nartey 1982); for American English in various studies by

Shadle et al.; and for French (Badin 1989; Beautemps et al. 1993; Beautemps et al. 1995). A similar range to that found for the second peak of Portuguese /f/ has been previously reported for Hebrew (Nartey 1982), American English (Shadle, Moulinier, Dobelke, and Scully 1992; Shadle and Scully 1995), and French (Badin 1989; Beautemps et al. 1993; Beautemps et al. 1995). The third and fourth peaks have been reported before only for French and American English in studies by Shadle and various colleagues. The broad peak seems to be also a spectral characteristic of Argentine Spanish (Manrique and Massone 1981) and French (Shadle, Moulinier, Dobelke, and Scully 1992; Shadle and Scully 1995).

Analysis of Speaker LMTJ's fricative /f/ spectra has revealed that, for most examples, the amplitude of the first peak is 4 - 10 dB higher than the amplitude of the second peak, as shown in Figure 5.7.

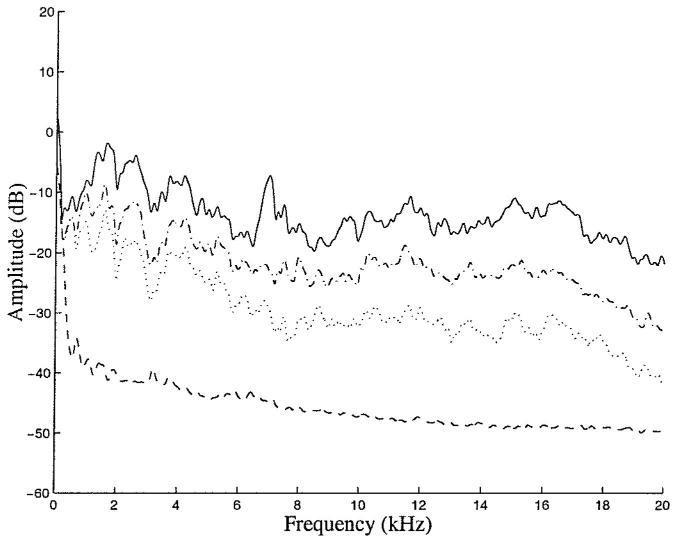

Figure 5.7: Time - averaged spectra of fricative /f/ sustained at three different effort levels: soft (dotted line), medium (dash - dotted line) and loud (solid line). The dashed curve is the averaged spectrum of the room noise. Corpus 1b (Speaker LMTJ).

In Corpus 2, for Speaker LMTJ, when there is no stress, or the stress is placed in the syllable containing the fricative, the amplitude at the beginning and end of the fricative decreases by 5 - 15 dB. When the stress is placed in the syllable before the fricative, the spectrum at the beginning, middle and end are quite similar in amplitude remaining within 5 dB of each other. For Speaker ACC, there is a 20 - 30 dB difference between the amplitude of the first peak and that at 20 kHz. The high frequency amplitude ($\geq$ 14 kHz) is lower for /pufu/.

5.3.2 Fricative /v/

Peaks at 1.3 - 1.7 kHz, 2.2 - 3 kHz, 3.6 - 4 kHz and 4.5 - 5.3 kHz have been observed for fricative /v/, for all subjects and all corpora. Although no high frequency broad peak could be observed for /v/, its frequency ranges of the peaks are otherwise quite similar to those found for /f/. However, the overall amplitude of /v/ was 5 - 20 dB lower than for /f/. Narayanan (1995) observed that the overall amplitude of American English /v/ was 5 dB lower than its unvoiced counterpart. Nartey's (1982) study also revealed a first peak in a similar frequency range as Portuguese /v/ for Hopi, Papago, Pima and Polish; and an equivalent second peak for Hebrew.

In Corpus 1b, ACC's fricative /v/ spectral amplitude is identical in the low frequencies ($\leq$ 5 kHz) for all effort levels, as shown in Figure 5.8. The amplitude difference between high and medium effort level spectra is 10 - 20 dB, and the amplitude of medium and low effort levels is identical, at higher frequencies.

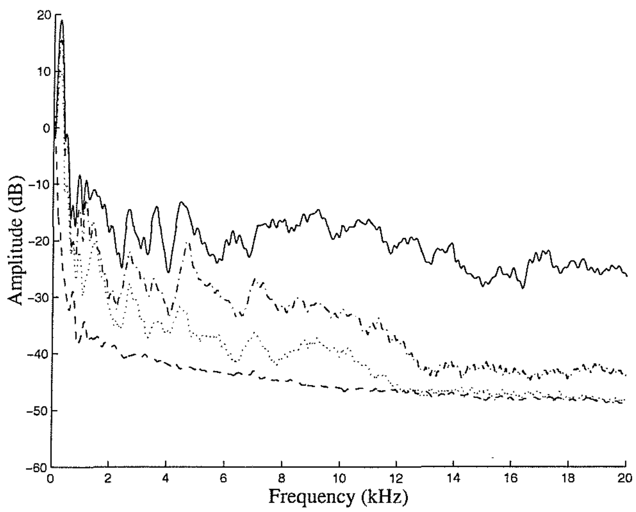

Figure 5.8: Time - averaged spectra of fricative /v/ sustained at three different effort levels: soft (dotted line), medium (dash - dotted line) and loud (solid line). The dashed curve is the averaged spectrum of the room noise. Corpus 1b (Speaker ACC).

5.3.3 Fricative /s/

Fricative /s/ has peaks at 1.4 - 2 kHz, 2.3 - 3.2 kHz and 4 - 5 kHz, for all subjects and all corpora. There are also broad peaks at 5.5 - 8.5 kHz (usually shifted down to 4 - 5 kHz for back vowel contexts) and 9.5 - 16 kHz. All the peaks and the first broad peak in Portuguese /s/ were also observed in the spectra of the twelve world languages studied by Nartey (1982). In studies of French spectra, peaks and broad peaks were reported to be located in the same frequency ranges as given above for Portuguese (Badin 1989; Shadle, Badin, and Moulinier 1991; Shadle, Moulinier, Dobelke, and Scully 1992; Shadle and Scully 1995; Beautemps, Badin, and Laboissière 1993; Beautemps, Badin, and Laboissière 1995). American English studies showed the same spectral characteristics, but the highest frequency broad peak was not visible (Shadle, Dobelke, and Scully 1992; Wilde 1995a; Wilde 1995b; Narayanan 1995).

The amplitude difference of /s/ in Corpus 1b (Speaker LMTJ) between low and medium effort level spectra, and between medium and high effort level spectra is about 15 dB at the first broad peak and about 30 dB at higher frequencies (above 5.5 - 8.5 kHz). For Speaker ACC, the amplitude difference between low and medium effort level spectra, and between medium and high effort level spectra, is the same as for LMTJ at the first broad peak and at higher frequencies.

In Corpus 2, above 5 kHz the spectral amplitude is highest mid - fricative and lowest end - fricative, for speakers LMTJ (see Figure 5.9) and ACC. These results are the same as those reported for American English by Shadle, Dobelke, and Scully (1992).

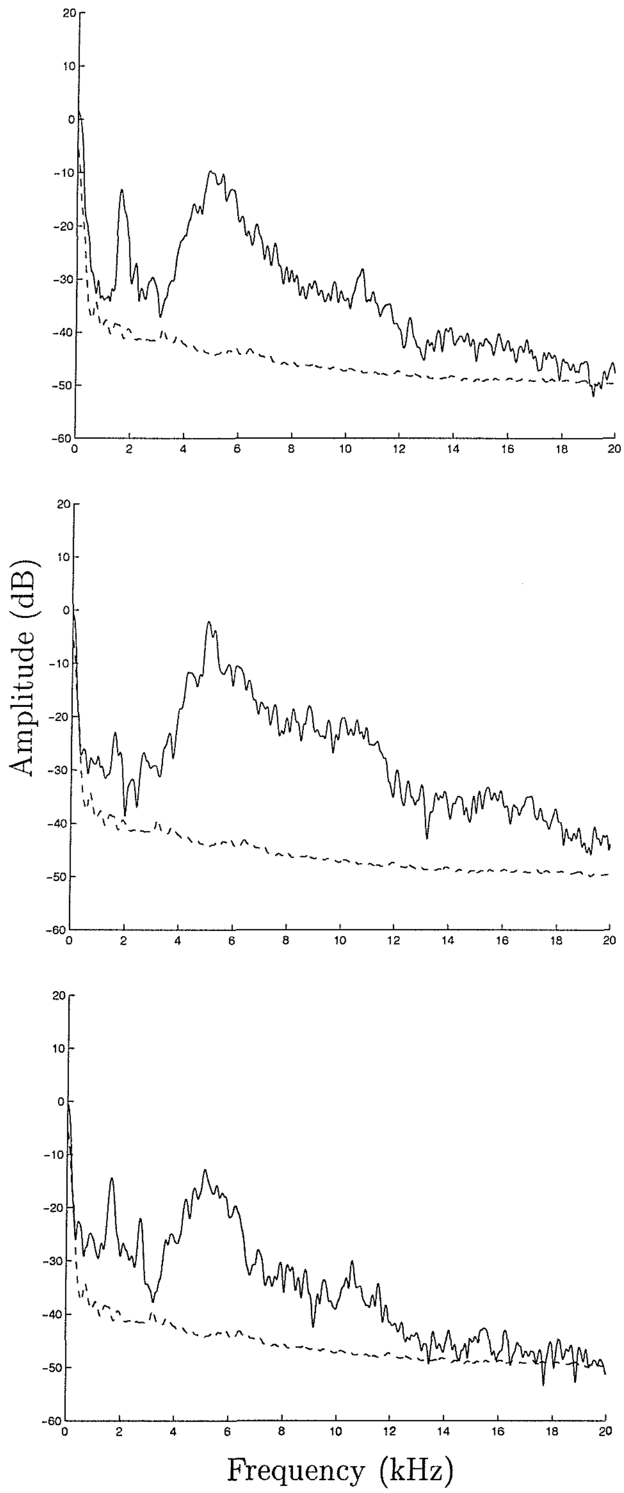

Figure 5.9: Ensemble-averaged spectra of fricative /s/ in /pɨ'sɐ/; Top: beginning of the fricative; Middle: centre of the fricative; Bottom: end of the fricative. The dashed curve is the time-averaged spectrum of the room noise. Corpus 2 (Speaker LMTJ).

The spectra of Corpus 4 (Speaker LMTJ) /s/ present a 20 - 35 dB drop in amplitude from the first broad peak to the second broad peak. There are some vowel and consonant contexts where the first broad peak has a reduced bandwidth, see Figure 5.10 – (a), due to the close presence of the second broad peak, as shown in Figure 5.10 – (b). For the words of ACC (Corpus 3 and 4), the most "reliable" spectral characteristics of /s/ are the first two peaks and a trough at 2.3 - 2.6 kHz (also observed for American English by Narayanan (1995)).

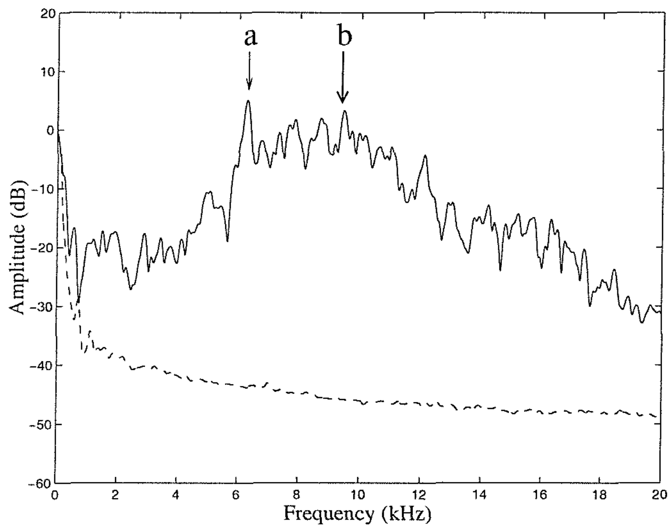

Figure 5.10: Time - averaged spectrum of fricative /s/ in *sala* ([a 'salɐ]): (a) – First broad peak; (b) – second broad peak. The dashed curve is the averaged spectrum of the room noise. Corpus 4 (Speaker ACC).

5.3.4 Fricative /z/

The spectrum of fricative /z/ has peaks at 1.1 - 1.8 kHz, 2.3 - 3.5 kHz and 4 - 5 kHz; broad peaks at 5.3 - 8.5 kHz and 9.8 - 17 kHz, for all subjects and all corpora. The first broad peak is shifted down for back vowel contexts, as shown in Figure 5.11. The first peak tends to be located at higher frequencies (1.6 - 1.8 kHz) for Speaker ISSS. The most prominent peaks and broad peaks of /z/ are located in the same frequency ranges as its unvoiced counterpart, but the overall amplitude of /z/ is 10 - 15 dB lower than that of /s/, as previously observed for French (Shadle and Scully 1995). There are also references to some low frequency peaks, in the same range as the Portuguese results, in studies of various other languages (Komshian and Soli 1981; Nartey 1982).

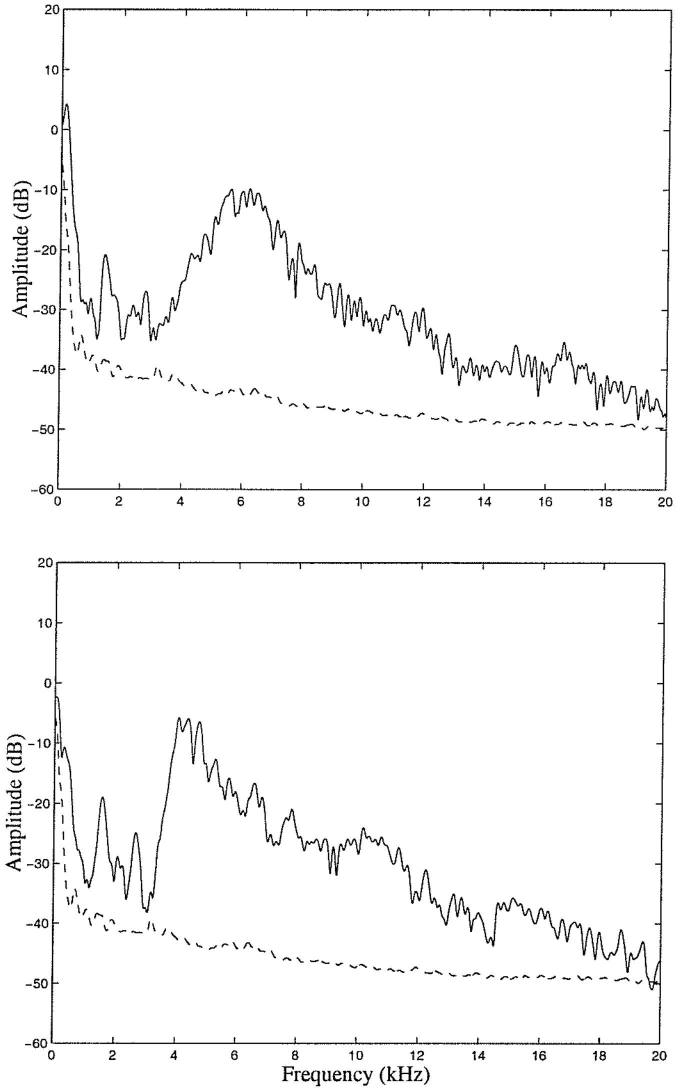

Figure 5.11: Time-averaged spectra of fricative /z/ in *zelar* /ˈzɨlar/ (top) and *zona* /ˈzonɐ/ (bottom). The dashed curve is the averaged spectrum of the room noise. Corpus 3 (Speaker LMTJ).

In Corpus 1b, for speakers LMTJ and ACC, there is a 10 - 30 dB difference between the amplitude of the first broad peak and that at 20 kHz. The spectral amplitude is identical in the low frequencies ($\leq$ 4 kHz) for all effort levels. The amplitude difference at higher frequencies (4 to 20 kHz) varies between 5 and 20 dB, as shown in Figure 5.12 for Speaker LMTJ.

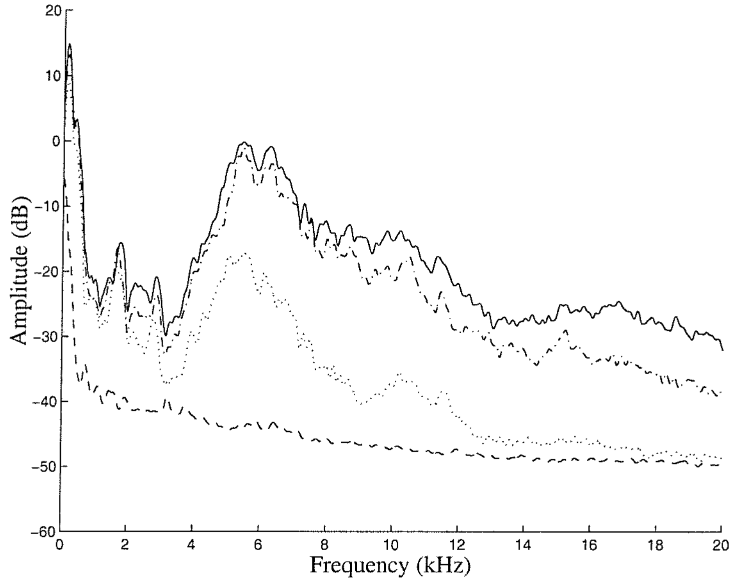

Figure 5.12: Time-averaged spectra of fricative /z/ sustained at three different effort levels: soft (dotted line), medium (dash-dotted line) and loud (solid line). The dashed curve is the averaged spectrum of the room noise. Corpus 1b (Speaker LMTJ).

For the words produced by Speaker ACC, a trough at 2.3-2.6 kHz could also be observed in the /z/ spectra (equivalent to the one in the /s/ spectra). Another characteristic of the alveolar fricatives /s, z/ of Speaker ACC is that the bandwidth of the third peak increases and the bandwidth of the two broad peaks decreases (Stevens 1998, pp. 130-137), resulting in a spectrum where there is a "third category of peaks" (designated in Figure 5.1 as "medium bandwidth peaks"). It is thought this is due to the proximity of poles and zeros of the transfer function (as referred to in the introduction).

For back vowel contexts /ɔ, o, u/, in Corpus 4 (Speaker LMTJ), the difference in amplitudes between the first broad peak and the first two peaks is larger. The broad peak is generally sharper and narrower for back vowel contexts. The lip rounding present in all back vowel contexts /ɔ, o, u/ results in lowering of front cavity peak frequencies and decreases their bandwidth which is strongly influenced by the radiation impedance. This decreased bandwidth contributes to enhancing the prominence of the broad peak (Stevens 1998, pp. 290-294). For back cavity resonances, that result in the spectral peaks at 1.1-1.8 kHz, 2.3-3.5 kHz and 4-5 kHz for Portuguese, deduced by comparison with results on cavity affiliation of peaks and troughs by Shadle et al. (1991), the bandwidths are only weakly influenced by the radiation impedance (Stevens 1998, pp. 152 - 156).

5.3.5 Fricative /ʃ/

The spectrum of fricative /ʃ/ has a peak at 1.4 - 2.2 kHz, and broad peaks at 2.5 - 4.5 kHz, 6 - 9.5 kHz and 12.1 - 16.9 kHz, for all subjects and all corpora. The peak and the first broad peak tend to be shifted down for back vowel contexts. There is also a trough at 0.6 - 1.3 kHz in the spectra of speakers LMTJ and ACC. Speaker ISSS's peak tends to be located at a higher frequency range (1.9 - 2.2 kHz), as shown in Figure 5.13.

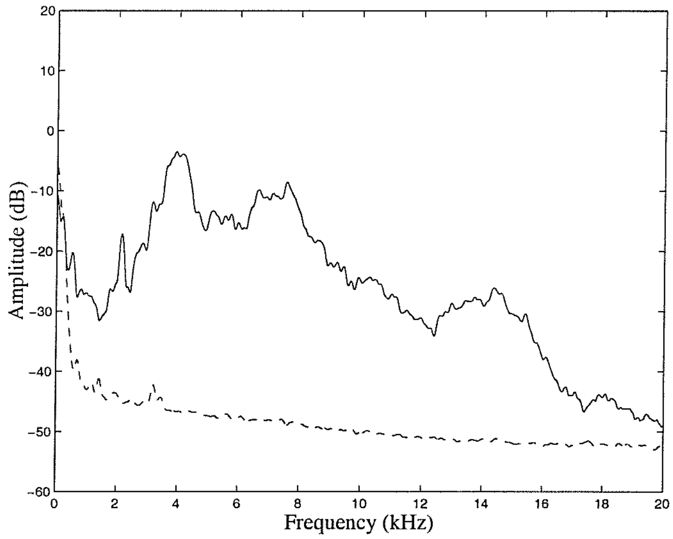

Figure 5.13: Time - averaged spectrum of sustained /ʃ/. The dashed curve is the averaged spectrum of the room noise. Corpus 1a (Speaker ISSS).

French /ʃ/ spectra presents a peak and a trough in the same frequency range as Portuguese, but the broad peaks are located at lower frequencies (Badin 1989; Shadle, Moulinier, Dobelke, and Scully 1992; Shadle and Scully 1995). American English (Shadle, Badin, and Moulinier 1991; Narayanan 1995) and various other world languages (Nartey 1982), seem to have fewer broad peaks than Portuguese, with quite different peak and broad peak locations.

In Corpus 1b, LMTJ's /ʃ/ spectrum has a fairly similar falloff at all levels; the difference between high and low effort level amplitudes at high frequencies is about 30 dB. For Speaker ACC, the amplitude difference between low and medium effort level spectra, and between medium and high effort level spectra is 5 - 20 dB. There is a 20 - 30 dB difference between the amplitude of the first broad peak and that at 20 kHz.

Word final fricative /ʃ/ produced by Speaker LMTJ has lower amplitudes above 7 kHz than LMTJ's other /ʃ/ spectra in Corpus 3 and 4. In Corpus 4, this fricative's spectra present a 25 - 40 dB drop of amplitude from the first broad peak to the second broad peak (approximately 2.5 to 9.5 kHz).

5.3.6 Fricative /ʒ/

The spectrum of fricative /ʒ/ has a peak at 1.2 - 2.1 kHz, and broad peaks at 2.3 - 4.7 kHz, 6 - 8 kHz and 9.5 - 16.3 kHz, for all subjects and all corpora, as shown in Figure 5.14. There is also a trough at 0.8 - 1.1 kHz for speakers LMTJ and ACC. For Speaker ISSS, the first peak tends to be located at higher frequencies (1.9 - 2.1 kHz). The first two broad peaks have a much narrower bandwidth (previously designated as "medium bandwidth peaks") for speakers LMTJ and ACC than for speakers CFGA and ISSS. The overall amplitude difference between /ʒ/ and /ʃ/ is quite variable for different speakers and corpora (0 - 30 dB).

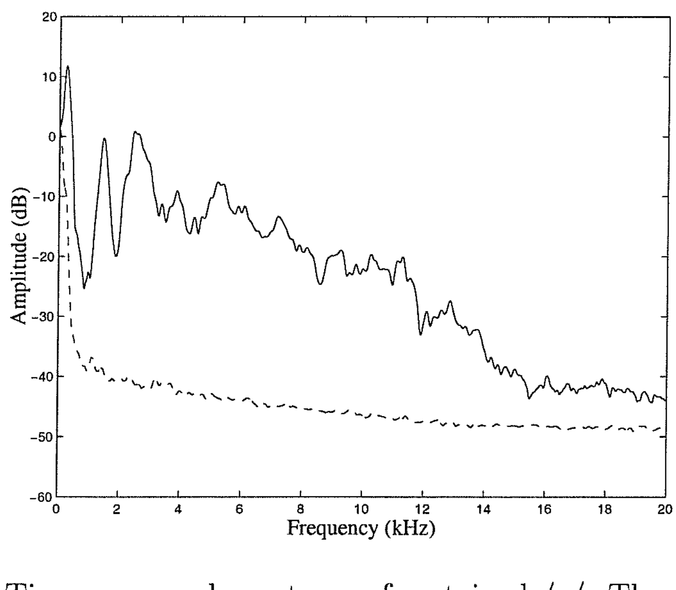

Figure 5.14: Time - averaged spectrum of sustained /ʒ/. The dashed curve is the averaged spectrum of the room noise. Corpus 1a (Speaker CFGA).

In Corpus 1b, for Speaker LMTJ, generally there is less amplitude difference at high frequencies between medium and high effort, than between low and medium effort. For Speaker ACC, the amplitude difference between effort levels for /ʒ/ is 5 - 10 dB, less than for /ʃ/. There is a 30 dB drop of amplitude on the spectra from the first broad peak to 20 kHz.

In the Corpus 4 spectra of Speaker LMTJ, there is a 20 - 30 dB drop in amplitude from the first broad peak to the last broad peak, and the overall amplitude of /ʒ/ varies over a similar range ($\sim$ 40 dB) to that of /ʃ/.

5.3.7 The Effect of Effort Level

The different effort levels tend to be associated with a shift in frequency of one or even two broad peaks of fricative /s/ for Speaker LMTJ, of fricative /z/ for speakers LMTJ and CFGA, and of fricative /ʃ/ for all speakers. The "direction" of such a frequency shift (up or down) varies among fricatives and speakers, which probably is the result of disparate strategies among speakers in the somewhat unnatural task of producing sustained fricatives at the three effort levels.

5.3.8 The Corpus 1a Spectra of Speakers LMTJ and ACC

The peaks in the spectra of Speaker ACC are much broader and their tops much flatter than for Speaker LMTJ, as can be seen in Figure 5.15. The overall amplitude of the spectra of fricatives produced by Speaker ACC is $\simeq$ 10 - 20 dB higher than for Speaker LMTJ. The "voice bar" (first peak in the spectra) of fricatives /v, z, ʒ/ for Speaker ACC is $\simeq$ 10 dB higher than for Speaker LMTJ.

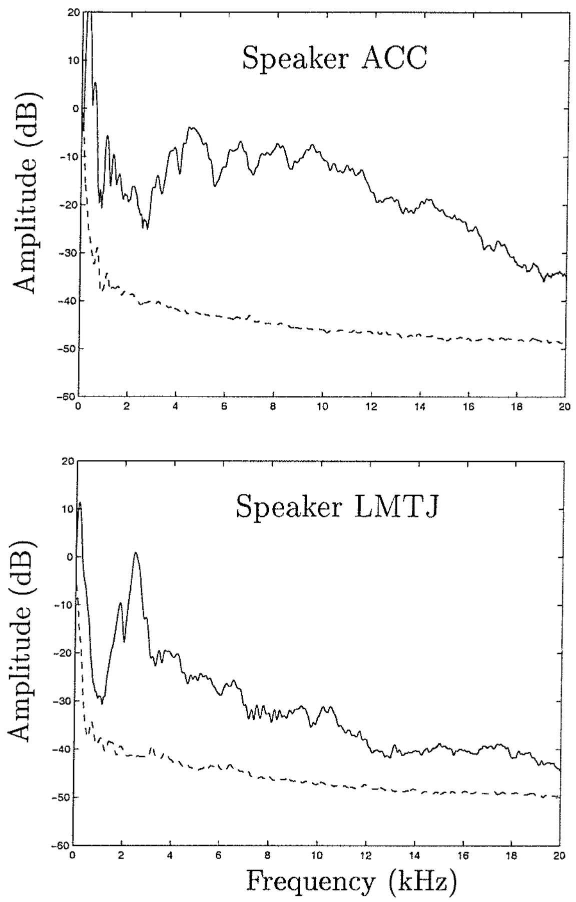

Figure 5.15: Time-averaged spectra of sustained fricative /ʒ/ in /ɐ-ʒʒʒ-ɐ/. The dashed curve is the averaged spectrum of the room noise.

5.3.9 Effects of Word Boundaries

The two examples of fricative /v/ in sentence 12 have very similar spectral peak frequencies (~ 1.6 kHz, 2.5 kHz and 3.6kHz). Durations of the fricative, and VF and FV transitions are also quite similar. It seems that /ɨ've/ whether contained within a word (as in *dever* /dɨ'veɾ/) or split across word boundaries (when a new word starts with /v/ as in *meche ver* /'mɛʃɨ veɾ/) have the same characteristics. The two pairs of fricative /s/ in sentences 11 and 12, with the same vowel context within a word and across word boundaries, are also quite similar in terms of VF, F, and FV durations, and spectral peak frequencies. The two examples of fricative /ʃ/ in sentence 11 are identical (time and frequency domain).

5.4 Syllable Stress and Effort Level – Labiodental, Alveolar and Postalveolar Fricatives

It was expected that varying the effort level of sustained fricatives, as in Corpus 1b, would have the same effect as varying the stress of the syllable containing the fricative in Corpus 2 and 3, and possibly also position within the word in Corpus 3. Results show that the spectral shapes and amplitudes of unstressed fricatives are similar to the soft effort level, and stressed fricatives are similar to the medium effort level. No fricatives from Corpus 2 or 3 resemble their high-effort-level Corpus 1b counterparts.

Figures 5.16 to 5.22 contrast spectra for /ʃ/ with similar vowel context from Corpora 2 and 3 (Speaker LMTJ). For four of the seven pairs-by-vowel-context the Corpus 2 spectral amplitude was slightly higher than that of Corpus 3 for frequencies above 6 kHz (see Figures 5.16 to 5.19). This amplitude difference across the frequency range corresponds strongly with a difference in stress between the two fricatives. In each of these four cases, the Corpus 3 fricative was in an unstressed syllable, and the Corpus 2 fricative was in a stressed syllable.

The other three such cases showed a different pattern, with the spectral amplitudes differing at the main peak but approximately equal above 3 kHz (see Figures 5.20 to 5.22). Two of these cases matched in stress (one pair, both stressed, shown in Figure 5.20; the other, both unstressed, shown in Figure 5.22); the third case, shown in Figure 5.21, did not match, and the amplitude difference at the peak was the largest (stressed Corpus 2 is 15 dB above unstressed Corpus 3). The same set of 7 examples was also analysed for Speaker ACC, but the overall spectral amplitude of fricatives from Corpus 2 and Corpus 3 was approximately the same.

These points taken together (see summary in Table 5.4) give us information needed to model the fricative. Corpus 2 is better controlled and easier to analyse than Corpus 3 or 4; validating its use gives an important advantage. However the inconclusive results for Speaker ACC raised the question of how general (different fricatives and speakers) the previous discussion was. Therefore we concluded that this line of research, with the current limited number of speakers, would not bear fruit.

Table 5.4: Spectral amplitude comparisons; paired by vowel context; stress of syllable containing the fricative differs. Spectral amplitudes > 2 kHz. Speaker LMTJ.

Corpus 1a e.g. [uʃ... ʃu]	Amplitude	Corpus 2 and 3 e.g. [puʃu], [kɐ'puʃ]
high effort		no equivalents in Corpus 2 or 3
medium	≈	all stressed
low	≈	all unstressed

Corpus 2	Amplitude	Corpus 3
stressed; [pi'ʃɐ]	>	unstressed; ['biʃɐ]
stressed; [pɐ'ʃu]	>	unstressed; ['taʃ]
stressed; [puʃu]	>	unstressed; ['moʃ]
stressed; [puʃu]	>	unstressed; [kɐ'puʃ]
stressed; [pɐ'ʃɐ]	≈	stressed; [ɐ'ʃaɾ]
stressed; [pɐ'ʃɐ]	≈	unstressed; [bu'laʃɐ]
unstressed; ['puʃɐ]	≈	unstressed; ['tɔʃɐ]

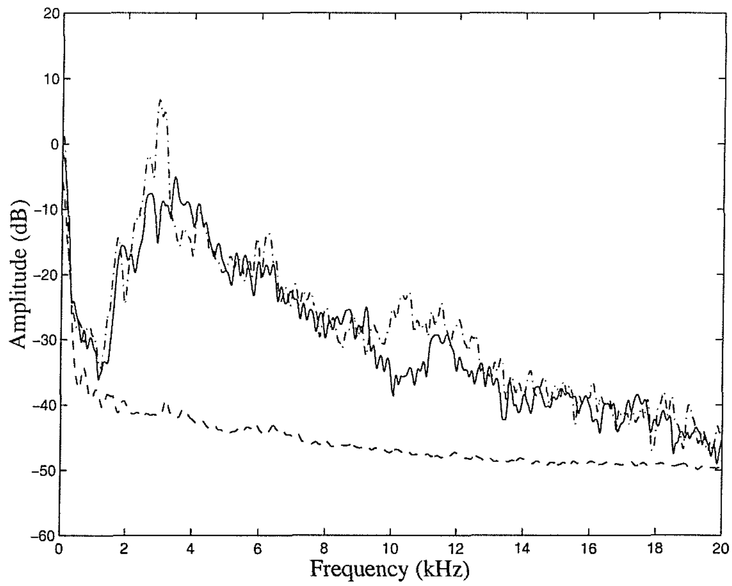

Figure 5.16: Averaged power spectra of fricative /ʃ/ in ['biʃɐ] from Corpus 3 (solid line), and in [pi'ʃɐ] from Corpus 2 (dash-dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

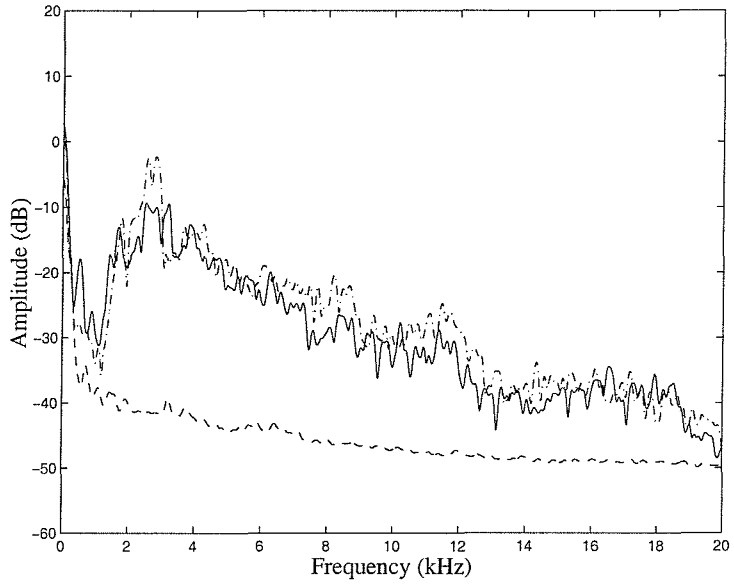

Figure 5.17: Averaged power spectra of fricative /ʃ/ in [ˈtaʃ] from Corpus 3 (solid line), and in [pɐˈʃu] from Corpus 2 (dash-dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

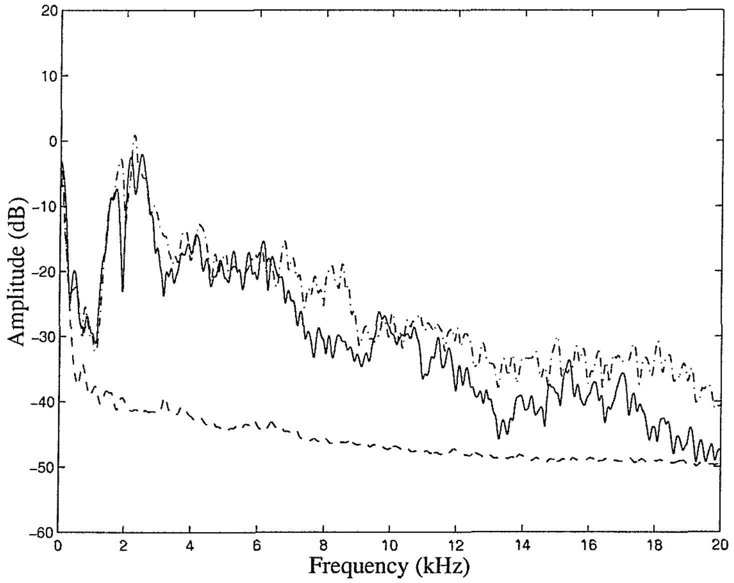

Figure 5.18: Averaged power spectra of fricative /ʃ/ in [ˈmoʃ] from Corpus 3 (solid line), and in [puʃu] (both syllables equal stress) from Corpus 2 (dash-dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

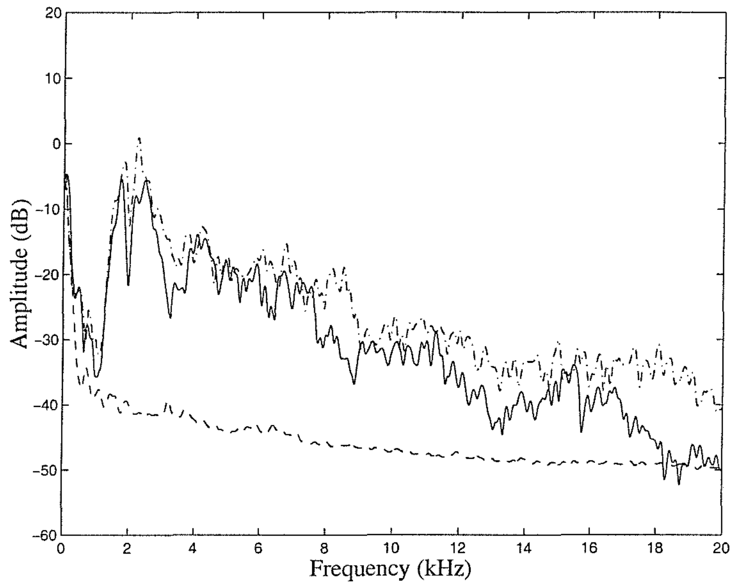

Figure 5.19: Averaged power spectra of fricative /ʃ/ in [kɐ'puʃ] from Corpus 3 (solid line), and in [puʃu] (both syllables equal stress) from Corpus 2 (dash - dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

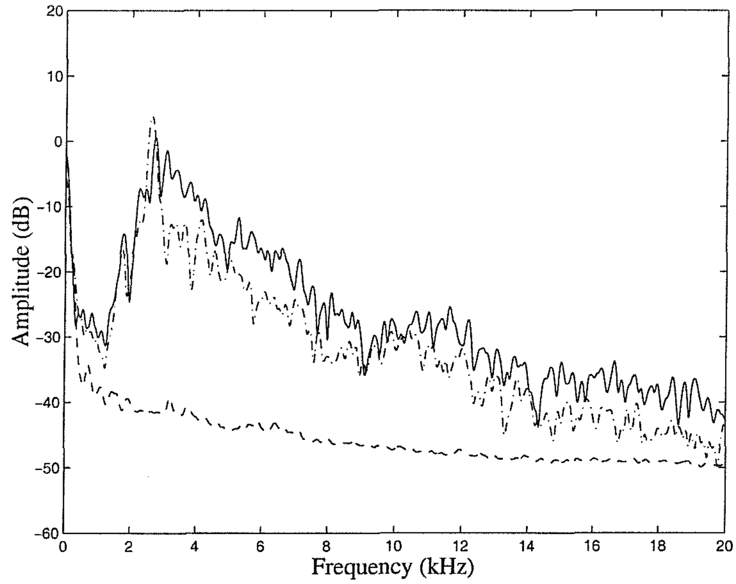

Figure 5.20: Averaged power spectra of fricative /ʃ/ in [ɐ'ʃaɾ] from Corpus 3 (solid line), and in [pɐ'ʃɐ] from Corpus 2 (dash - dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

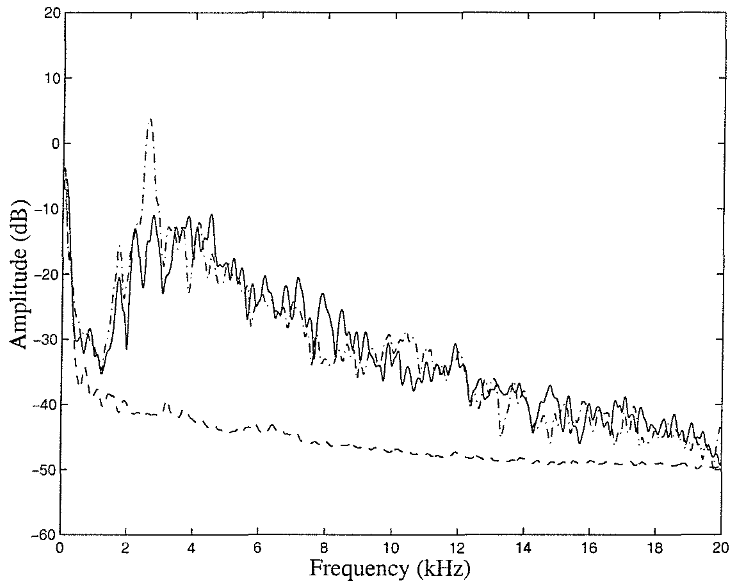

Figure 5.21: Averaged power spectra of fricative /ʃ/ in [bu'laʃɐ] from Corpus 3 (solid line), and in [pɐ'ʃɐ] from Corpus 2 (dash-dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

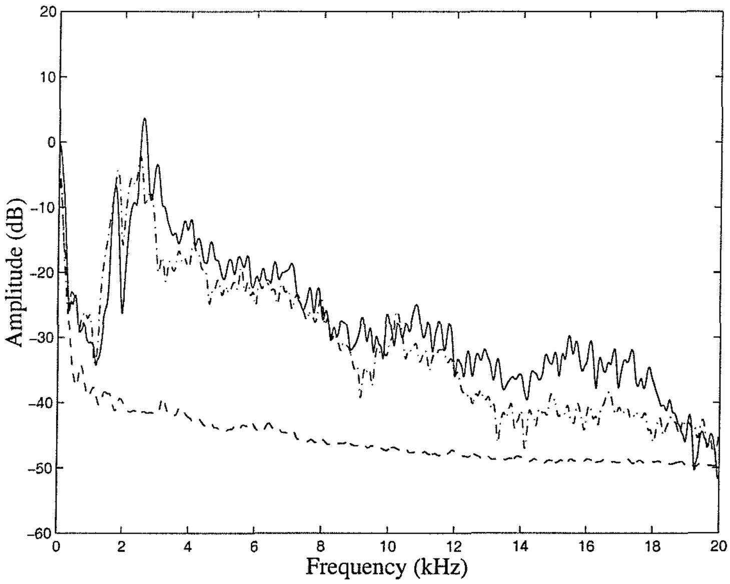

Figure 5.22: Averaged power spectra of fricative /ʃ/ in ['tɔʃɐ] from Corpus 3 (solid line), and in ['puʃɐ] from Corpus 2 (dash-dotted line). The dashed curve is the averaged spectrum of the room noise. Speaker LMTJ.

5.5 Spectral Analysis of Uvular Fricatives and Voiceless Tapped Alveolar Fricatives

The spectrum of fricative [χ] has peaks around 1.2 - 1.8 kHz, 2.4 - 3 kHz and 3.4 - 4 kHz; troughs around 1.9 - 2.2 kHz and 4.1 - 4.7 kHz (not always visible); and a low amplitude broad peak that can be centred from 7.5 kHz to 11.5 kHz (not always visible), for all subjects and all corpora. There is a 30 - 40 dB falloff of amplitude over the 1.5 - 20 kHz frequency range. If we compare the spectrum of [χ] shown in Figure 5.23 (top) with the spectrum of [ʃ] (bottom of Figure 5.23), we can see the main spectral peak being shifted down because the place of articulation moves back from postalveolar ([ʃ]) to uvular ([χ]). A spectrogram of a word with the fricative [χ] is given in Figure 5.24. In our limited inventory we've registered only two productions of fricative [ʁ], as shown in Table E.2, with spectral characteristics very similar to its unvoiced counterparts.

The spectral results for Portuguese are comparable to those of velar and uvular fricatives in various other languages. Jassem (1967) reported spectral peaks for Polish /x/ at 0.5 kHz, 1.4 kHz and 2.1 kHz, and for /χ/ at 0.6 kHz, 1.1 kHz, 2.3 kHz and 3.4 kHz. Delattre's (1971) study of American English, Arabic, French, German and Spanish pharyngeal consonants included a detailed description of the different articulatory gestures used to produce /x/, /χ/ and /ʁ/. Manrique and Massone's (1981) analysis of Argentine Spanish fricatives, in different vowel contexts, revealed spectral peaks at 0.5 - 3 kHz (mean = 1.7 kHz) for /x/. Nartey (1982) studied fricatives of twelve different languages, reporting the spectral peaks shown in Table 5.5 for /x/ and /χ/ in three different nonsense word vowel contexts /i, a, u/.

Table 5.5: Fricative peak frequencies in kHz. For a given place of articulation the same number of columns is used, and if a peak is present its frequency range is given. After Nartey (1982).

	/ixi/	/axa/	/uxu/
Arabic	-; 1.7 - 2.3; -; 6.7 - 7.7	0.5 - 0.6; 1.1 - 1.3; 3.7 - 4.4; -	0.5 - 0.6; -; 3.7 - 4.4; -
Navajo	-; -; 3.2 - 3.7; -	-; 1.1 - 1.3; 3.7 - 4.4; -	-
Polish	-; 2.3 - 2.7; -; -	0.2 - 0.3; 1.5 - 1.7; -; -	0.4 - 0.5; 0.8 - 0.9; -; -
	/iχi/	/aχa/	/uχu/
Hebrew	1.5 - 1.7; 3.2 - 3.7	1.1 - 1.3; 3.2 - 3.7	0.9 - 1.1; 3.2 - 3.7

Alwan's (1986) results from acoustic vocal tract models showed that for uvular fricatives /χ, ʁ/ formant F1 should be a Helmholtz resonance, F2 and F4 front-cavity resonances and F3 a back cavity resonance. These predictions were confirmed by spectral analysis results which revealed the peaks shown in Table 5.6. The waveform envelope of /ʁ/ was lower in amplitude than that of the surrounding vowels and the formants above F1 were very weak.

Table 5.6: Fricative formant frequencies in kHz. The values are averaged across four speakers. After Alwan (1986).

	/χiː/	/χaː/	/χuː/
F1	0.4	0.6	0.4
F2	1.6	1.3	0.9
F3	2.6	2.5	2.6
	/ʁiː/	/ʁaː/	/ʁuː/
F1	0.4	0.5	0.4
F2	1.4	1.2	0.7
F3	2.5	2.6	2.6

Beautemps et al. (1993, 1995) reported spectral peaks for French /x/ at 600 Hz, 1211 Hz, 2180 Hz and 3665 Hz. Shadle et al. (1995) also presented results of a spectral analysis of /x/ produced by American, French and German speakers. The unvoiced velar fricative /x/ had evenly spaced peaks from 1 to 1.5 kHz, front cavity affiliated peaks at 2 and 3.8 kHz, and a trough around 3 kHz.

The spectrum of fricative [ɾ̥] has peaks around 1.4 - 1.7 kHz and 2.5 - 3.1 kHz; a trough around 2 - 2.4 kHz; and a low amplitude broad peak that can be centred from 9 kHz to 12 kHz, for all subjects and all corpora. However, these cues in the spectra are not always visible and there is great variability of the spectral structure (but not the overall amplitude). There is a 15 - 30 dB falloff of amplitude over the 1.5 - 20 kHz frequency range. A spectrogram of a word with the fricative [ɾ̥] is given in Figure 5.25.

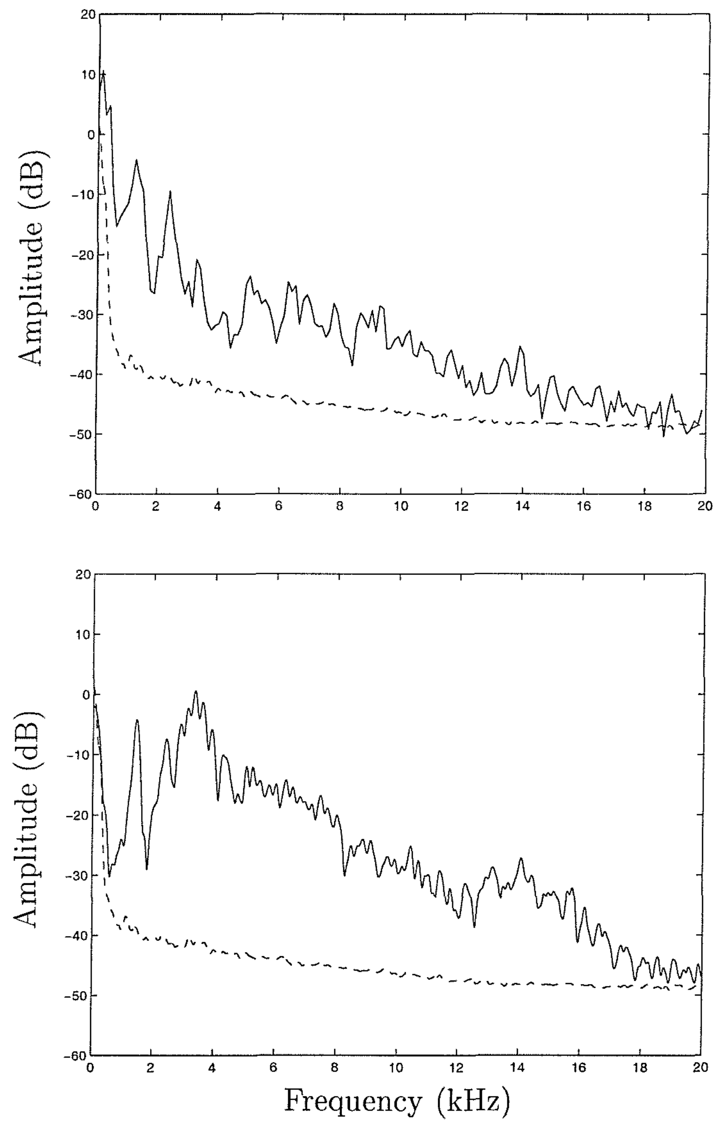

Figure 5.23: Power spectra of [χ] (top) in *ressaca* [ɐ χ'sakɐ] (preceded by the word *diga* ['digɐ]), and [ʃ] (bottom) in *meche* ['mɛʃi]. The dashed curve is the time-averaged spectrum of the room noise. Corpus 4 (Speaker CFGA).

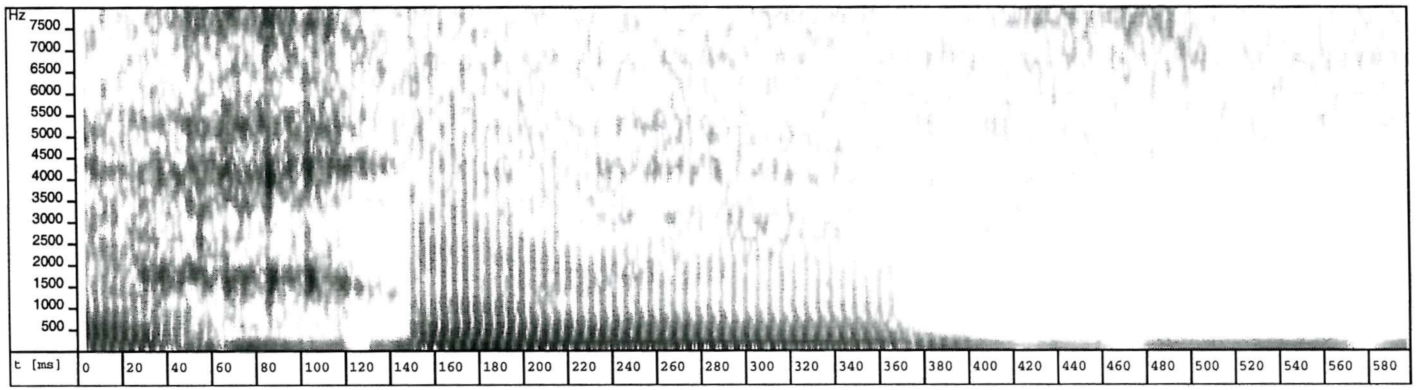

Figure 5.24: Spectrogram of the word *relevo* [ɐ χɨ'lev] (preceded by the word *diga* ['digɐ]). Corpus 3 (Speaker ISSS).

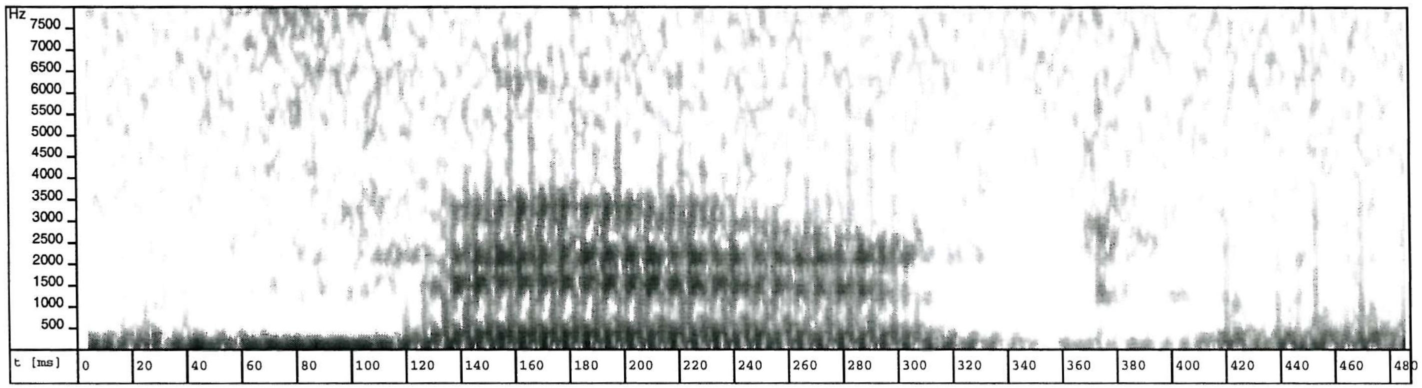

Figure 5.25: Spectrogram of the word *ver* [veɾ̥ u] (followed by the word *o* [u]). Corpus 4 (Speaker CFGA).

5.6 Summary

In this chapter, a study of the time-averaged spectra of sustained fricatives and of fricatives in real words, together with a study of the ensemble-averaged spectra of nonsense words, was presented. The broad spectral envelope was analysed, and then a refined description of specific frequency bands which present significant peaks or troughs was proposed. The analysis of results from different corpora showed that the peak and trough locations that characterize each place of articulation were: labiodentals – peaks at 1.3-1.9 kHz, 2.2-3 kHz, 3.1-4 kHz and 4.1-5.3 kHz, and a broad peak at 8-11.5 kHz; alveolars – peaks at 1.1-2 kHz, 2.3-3.5 kHz and 4-5 kHz, and broad peaks at 5.3-8.5 kHz and 9.5-17 kHz.; postalveolars – peak at 1.2-2.2 kHz, and broad peaks at 2.3-4.7 kHz, 6-9.5 kHz and 9.6-16.9 kHz. Only back vowel context seems to affect some of the peak and trough locations in the spectra of Portuguese fricatives. The peak locations (in the range 1 to 8 kHz) for the spectra of sustained fricatives are identical to those of corresponding fricatives in real words. Some of the broad peaks above 10 kHz observed in sustained fricatives are not visible in the spectra of fricatives from the word corpus.

From the analysis of the time-averaged spectra of different fricatives from Corpus 1b we have observed that the amplitude differences between the three effort levels are smallest at low frequencies. The amount of amplitude difference at high frequencies varies with the fricatives and tends to be smaller for the voiced fricatives. These differences are associated with source type and strength, and are similar to results for American English and French subjects.

All of this detailed spectral information will be used in the parameterisation model proposed in Chapter 6, especially the location of the highest frequency peak of /f, v/ and of the lowest frequency broad peak of /s, z, ʃ, ʒ/.

Chapter 6

Parameterising the Spectral Characteristics of Fricatives

6.1 Introduction

Fricative spectra have been parameterised to aid comparisons across speaker and across corpus, and to gain insight into the production mechanisms underlying the language-specific variations. The parameters spectral slope, frequency of maximum amplitude, and dynamic amplitude, derived from previous studies, were used to analyse changes in effort level, voicing, and sampling time within the fricative. Some combinations of the spectral parameters were also used. Results discussed in Chapter 5 will be related to differences in production mechanisms via these parameters: we expect to have different spectral parameters for each place of articulation; different dynamic amplitude and slope values for the various effort and stress levels; the effect of rounding to be captured by the frequency of maximum amplitude parameter, etc.

6.2 Previous Studies Parameterising Fricatives

While our long-range goal is improved synthesis of Portuguese, the analysis methods described in this thesis are more closely related to other goals that have long proven difficult for fricatives. Distinguishing fricatives in spite of changes caused by context effects is the fundamental task of speech recognition. Identifying differences in production when the intended fricative is known is needed for studies of disordered speech, and changes in speech over time, e.g. after a cochlear implant. Both of these applications require some knowledge of the distinguishing acoustic characteristics of fricatives that hold across speaker, context, and speaking style. We consider previous studies of distinguishing characteristics of fricatives below, grouped by type of parameter.

Locus equations have been used by many authors (Sussman 1994; Fowler 1994). The method consists of finding the slope and the intercept of a line between the same acoustic feature in a vowel and an adjoining consonant. For instance, Sussman (1994) used $F_2(onset) = kF_2(vowel) + c$ on both stops and fricatives. The corpus consisted of a small number of nonsense words produced by four speakers of American English. k and c could be used to distinguish stops, but did not distinguish fricatives well. Jongman et al. (2000) also studied locus equations as cues to place of articulation. The corpus consisted of nonsense words produced by 20 speakers of American English. The slope of locus equations could be used to differentiate labiodentals from the other three places of articulation.

Forrest et al. (1988) used spectral moments to characterise normal speech with the intent of using them on disordered speech. In this method, the spectral envelope was treated as a probability density function and the first four moments of that function were found and used in a cluster analysis. Results showed that spectral moments worked well to classify stops, as shown in Figure 6.1, but could not distinguish all fricatives. However, the authors used a very limited corpus: only 5 words contained fricatives. They were produced by 10 speakers of American English.

Shadle and Mair (1996) used spectral moments, as in Forrest et al. (1988), on a large fricative corpus recorded by one American English and one French native speaker. These researchers examined multiple tokens, varying effort levels, different vowel contexts, and three different locations within a fricative. They showed that the moments that Forrest et al. (1988) found to be the most useful for distinguishing fricatives (especially skewness) were not so useful with their data; the differences across context within place were typically much greater than the differences across place. Jongman et al. (2000) seemingly contradicted this finding in their study using 20 American speakers, and 6 vowel contexts for each fricative. ANOVAs showed that spectral skewness differed significantly by place for all four places, and the other moments were significantly different for two or three place-groups. For these studies, all subjects, voicing, vowel contexts, and location within fricative were lumped together. The implication then is that the distributions for each place overlapped substantially, and classification based on these measures would likely have a high failure rate. Very little information is given about the effect of voicing, vowel context, etc., on the values of each moment.

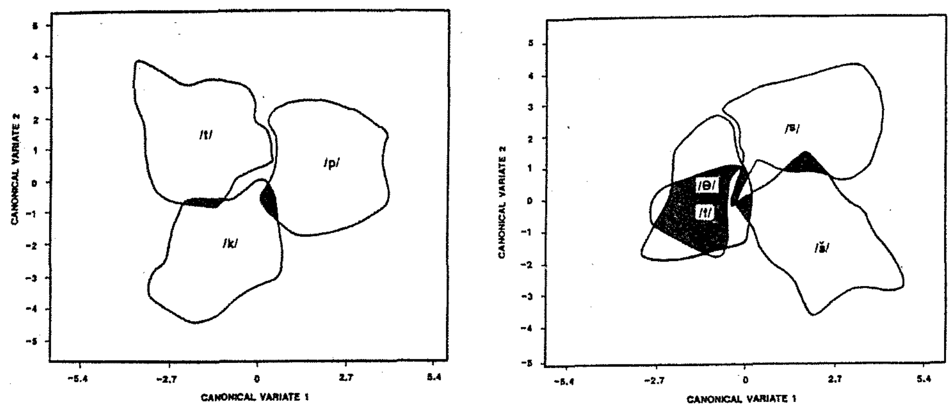

Figure 6.1: Cluster centres, marked by appropriate phoneme, and boundaries enclosing voiceless stops and fricatives. From Forrest et al. (1988).

Wilde (1995a) studied acoustic cues (place and voiced / voiceless categorization) in fricative-vowel boundaries and assessed perceptual importance of various time- and frequency-domain parameters via synthesis. She used nonsense words produced by four speakers of American English. She concluded that temporal and spectral characteristics of voiceless fricatives are more dependent on vowel context than those of voiced fricatives, and what she referred to as "formant onset time of fricatives" provides important place information. She also showed that the amplitudes of fricative noise in restricted frequency regions can distinguish sibilants (see definition on page 33) from nonsibilants.

Funatsu (1995), in a cross-language study of Japanese and Russian fricatives in nonsense words, measured two spectral parameters: the frequency of the most prominent fricative peak (F_F) and the onset frequency of the second formant transition of the following vowel (F_V), both measured over a 0-10 kHz range. Japanese /ʃ/ had a lower F_F and higher F_V than Japanese /s/. Russian /ʃ/ had a lower F_F and the same F_V compared to Russian /s/.

Russian fricatives, when followed by /o/ or /u/, had lower F_F and F_V values than in /a/ vowel context. Japanese fricatives, when followed by /o/, had lower F_F and F_V values than in /a/ vowel context, but when followed by /u/, /s/ had the same F_F and F_V values as in /a/ vowel context and /ʃ/ had the same F_F and higher F_V values than in /a/ vowel context.

Shadle and Mair (1996) defined two parameters, dynamic amplitude and spectral slope, which will be discussed in more detail later. These did not distinguish the fricatives completely but did vary with source location and effort level as predicted.

Evers et al. (1998) tried to distinguish and characterize the fricatives /s, ʃ/ produced by two speakers each of English, Bengali and Dutch (12 real words). 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.

Choo and Huckvale (Choo and Huckvale 1997; Choo 1999) studied the correlations between perceptual and physical spaces of voiceless English fricatives using a multidimensional scaling technique. Results from perceptual tests suggested that a two-dimensional solution was the most appropriate to model the data. Dimension 1, shown to be related to the "peakiness" of spectra (the difference between maximum amplitude and mean amplitude), clearly separated the sibilants from nonsibilants. Dimension 2, related to the centre of gravity of the spectra, separated fricatives according to their place of articulation. These representations were constructed from both perceptual similarity judgments and a Euclidean spectral distance metric obtained from 1/3-octave bandpass filtering.

Jongman et al. (2000) studied the frequency of the highest amplitude spectral peak, and noise duration and amplitude, as cues to place of articulation. Spectral peak location decreased in frequency as place of articulation moved posteriorly. The amplitude of the highest amplitude spectral peak differed significantly for all four places of articulation, for values computed across speakers, voicing and vowel context. Voiced fricatives had smaller amplitude relative to the vowel preceding them than unvoiced fricatives, with a larger difference between voiced and unvoiced for nonsibilants than for sibilants. Noise duration differed significantly for sibilants and nonsibilants.

In a study of acoustic place cues of plosives and fricatives, Chen and Alwan (2000) analysed /f, v, s, z/ in CV syllables where V was one of the vowels /a, i, u/ as produced by 2 male and 2 female American English speakers. Spectral amplitude of frication noise relative to the first formant at vowel onset, and locus equations' y-intercepts, appeared to cue place.

Parameters similar to those used by Shadle and Mair (1996), and a parameter similar to S_a used by Evers et al. (1998), were used in this study 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. In the process we have not only enhanced our understanding of Portuguese fricatives, but have contributed to the methodology for studying fricatives of any language.

6.3 Parameterisation

The mechanical model results of Shadle (1985) were used to define parameters that characterise the fricatives in the present study. These parameters have already been developed as a potential tool for classifying fricatives using real speech (Shadle and Mair 1996). They consist of measures of spectral slope and of the dynamic range of the spectrum, 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 vocal 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, related to the position of the source with respect to the tract.

It can be shown (Shadle 1985; Stevens 1998) that a source located in an intermediate position (i.e. not at the glottal end of the tract) always produces a zero at low frequencies. In a typical fricative configuration, articulators form a small constriction that separates the upstream back cavity from the downstream front cavity. Noise is generated somewhere downstream of that constriction. A set of antiresonances will be generated that nearly cancel back-cavity resonances. Front-cavity resonances are not cancelled, however, and a set of anti-resonances is generated at frequencies related to the distance between noise source and constriction. These anti-resonances result in sharply-defined troughs in the spectrum if the noise source is localized; if the source is distributed, the troughs will be correspondingly smeared.

The spectral prominence of the uncancelled resonances will depend on a number of factors: the particular ordering of resonances and antiresonances in the transfer function as a whole, the losses (particularly radiation losses) and the noise source strength. Above approximately 5 kHz, non-planar modes begin to propagate; the cut-on frequency is inversely proportional to the longest cross-dimension of the tract. The radiation impedance decreases, and losses due to radiation thus decrease, above that frequency.

The noise source spectrum depends on the shape of the constriction, the tract downstream of it, and the flow velocity through it. This source spectrum has been described as having a broad peak, with maximum amplitude at a frequency proportional to the mean velocity through the constriction (Stevens 1971). This is the spectrum of turbulence noise generated by a free jet, and the parameters that allow spectra for any jet diameter or velocity to be collapsed into a single curve have been amply described in the literature (e.g. Goldstein 1976). However, with most fricatives the jet emerging from the constriction cannot freely expand, but impinges on the tract walls downstream, generating additional noise; at speech dimensions and velocities, these sources of noise are generally of much higher amplitude than the self-noise of the jet (Shadle 1990).

Nelson and Morfey (1981) investigated the noise generated by a spoiler in a duct. These spectra likewise can be collapsed onto a single curve, but with an important difference: the spectrum does not show a broad peak, but instead shows a progressive decrease in amplitude as frequency increases. The source spectrum shape was similar to that derived by Shadle (1990) for obstacle-source mechanical models, which were shown to be good models for /s, ʃ/ (Shadle 1991).

Figure 6.2 shows an idealized fricative noise source spectrum. If the tract geometry, including constriction area, remains the same and flow velocity is increased, the spectral envelope of the noise increases in amplitude at all frequencies, *but more so at higher frequencies* (Shadle and Mair 1996; Krane 1999). A greater pressure drop across the constriction (Δp_c) results in a higher velocity through it, and thus an increased source amplitude. For unvoiced fricatives, greater subglottal pressure results in a greater Δp_c; for voiced fricatives, a pressure drop is required to drive the vocal-fold oscillation, and thus the noise source in voiced fricatives is generally weaker on average than in their unvoiced counterparts.

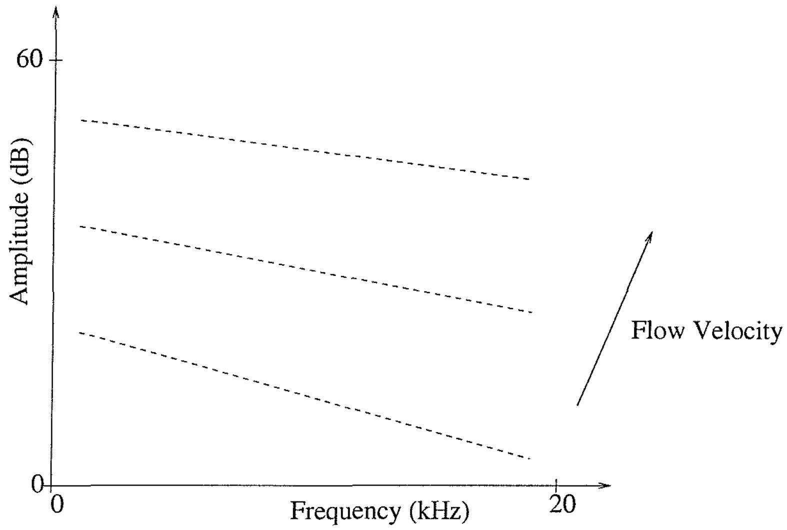

Figure 6.2: Schematic illustration of the effect of increasing airflow velocity on the noise source spectral envelope.

In a sustained fricative, the essentially static tract configuration produces static source and filter characteristics. Increasing the effort level should closely approximate altering the source while keeping the filter constant. For fricatives in vowel context, however, the articulators move at least at the start and end of the fricative, setting up filter and source characteristics simultaneously, and may move during the fricative; for instance, the tongue constriction may be held while the lips round in preparation for the following vowel. While we hope to capture the source differences that occur during a fricative by using ensemble averaging at its beginning, middle and end, we must remember that the filter is changing during this time as well. When we are forced to time-average through a fricative, all of these effects are averaged together.

In this study we are primarily interested in describing the acoustic variation caused by the context or the way in which a particular fricative is spoken, rather than identifying the fricative regardless of its context. We are thus interested primarily in changes in the source spectrum, since it offers clues to the source variations across subject and place. We have devised parameters accordingly. However by restricting ourselves to the far-field acoustic spectrum, it is not always possible to differentiate source and filter characteristics unambiguously.

We originally planned to use F (shown in Figure 6.3) as the endpoint for the spectral slopes. However, particularly for /f, v/, the values of F for the subjects ranged widely, from 4.2 to 7.8 kHz. These differences were not interesting, since the spectra were relatively flat. We therefore computed parameter $\overline{F}$, the average (rounded to the nearest kHz) of the manually calculated values of all sustained tokens (Corpora 1a and 1b) for each place for all four speakers: $\overline{F}_{/\text{f,v}/} = 5\,\text{kHz}$, $\overline{F}_{/\text{s,z}/} = 6\,\text{kHz}$ and $\overline{F}_{/\text{ʃ,ʒ}/} = 4\,\text{kHz}$. In Figure 6.3, F differs from $\overline{F}$, but the spectral slopes resulting appear to characterize the spectral shape fairly, and are computed using the same frequency range (and number of points) for all subjects.

By using $\overline{F}$ we are ignoring changes in the peak frequency with vowel context. In Corpus 1a, in a majority of cases the rounded-vowel context had the lowest F, as expected, but this was not consistent either within subject for a given fricative, or across fricatives.

The dynamic amplitude, A_d, is the difference between the maximum amplitude value of the averaged power spectrum occurring between 500 Hz (lower limit set so that room noise, and the peaks corresponding to the fundamental frequency and its first few harmonics, are not used in the calculation of A_d) and 20 kHz, and the minimum amplitude between 0 and 2 kHz. Two regression lines are fit to the spectrum; S_p' is the slope of the line fit to all the spectral amplitude points from 500 Hz to $\overline{F}$ (shown in Figure 6.3 as a dashed line), and S_p is the slope of the line fit to all the points from $\overline{F}$ to 20 kHz (solid line). This frequency range allowed us to capture relevant variations in the slope of the spectrum which was not possible in previous studies such as the one by Badin et al. (1994), in which spectral tilts were measured only up to 5 kHz.

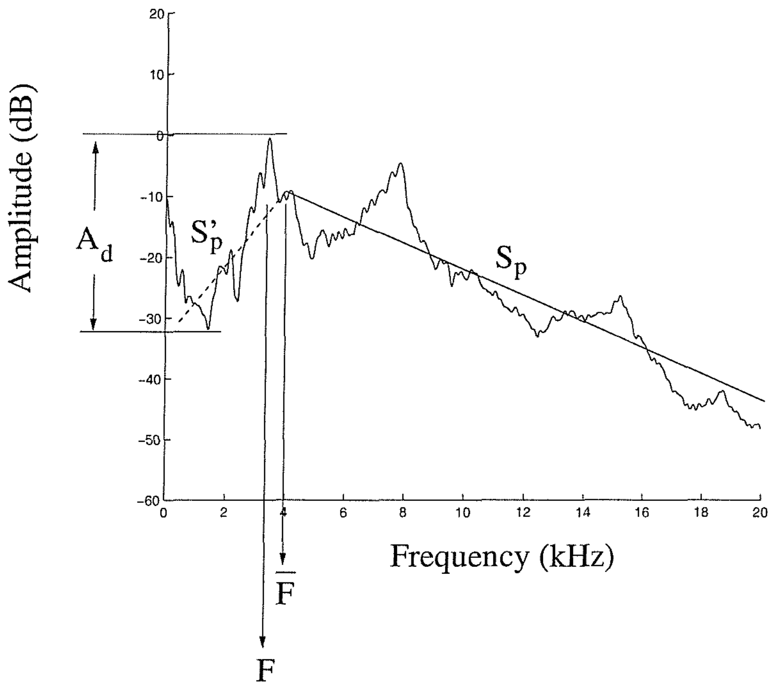

Figure 6.3: Dynamic amplitude A_d, and regression lines used to calculate low frequency (500 Hz to $\overline{F}$ kHz) slope S'_p (dashed line) and high frequency ($\overline{F}$ kHz to 20 kHz) slope S_p (solid line). Sustained fricative /ʃ/ (Corpus 1a) produced by Speaker ISSS.

Given these definitions, and what is known about parameter and articulatory behaviour, we can make the following predictions. The parameter A_d should be maximized for a localized source, and for higher relative noise source strength, as in sibilants and unvoiced fricatives. A_d may not be very large for the weak 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 the transfer function is assumed to 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 (Scully et al. 1992; Shadle and Scully 1995). The behaviour of parameter S'_p should be similar to that of 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 a more posterior place or greater source strength, because a more posterior place (or rounding) lowers F and a greater source strength increases the amplitude of the peak at F. See Table 6.1 for a summary of the predicted effects on parameters.

Table 6.1: Predicted effects on parameters.

Phonetic Class	Aeroacoustics	Predictions
Posterior place; sibilants /s,z,ʃ,ʒ/	Longer front cavity; Localized source; Higher source strength *	F lower; A_d, S_p and S'_p higher
Forward place; nonsibilants /f,v/	Distributed source; Lower source strength	F higher; A_d, S_p and S'_p lower
Unvoiced	Higher source strength *	A_d, S_p and S'_p higher
Voiced	Lower source strength	A_d, S_p and S'_p lower
Loud effort level (relative to medium and soft)	Higher source strength *	A_d, S_p and S'_p higher
Middle of fricative (relative to beginning and end)	Higher source strength *	A_d, S_p and S'_p higher
Stressed syllable (relative to unstressed syllable)	Higher source strength *	A_d, S_p and S'_p higher
Medial word position (relative to initial and final)	Higher source strength *	A_d, S_p and S'_p higher
Rounded (relative to unrounded)	Longer front cavity; Lower source strength †	F lower; A_d higher; ? S'_p and S_p lower
Male subjects	Longer front cavity; Higher source strength *	$\overline{F}$ lower; A_d, S_p and S'_p higher
Female subjects	Shorter front cavity; Lower source strength	$\overline{F}$ higher; A_d, S_p and S'_p lower

* A higher source strength is produced by higher volume velocity for the same constriction area A_c, or a constant volume velocity for a smaller A_c. The net result is a higher particle velocity in the constriction.

† For rounded vowel contexts the lips form a second constriction and so the first constriction (that intrinsic to the fricative) generates a noise source with lower strength (Shadle and Scully 1995).

6.4 Results

6.4.1 Sustained Fricatives

Figure 6.4 shows average regression line fits (from $\overline{F}$ to 20 kHz) to the spectra of the sustained fricatives in Corpus 1b for Subject ISSS. 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. For all subjects, 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. The results were similar for the other subjects.

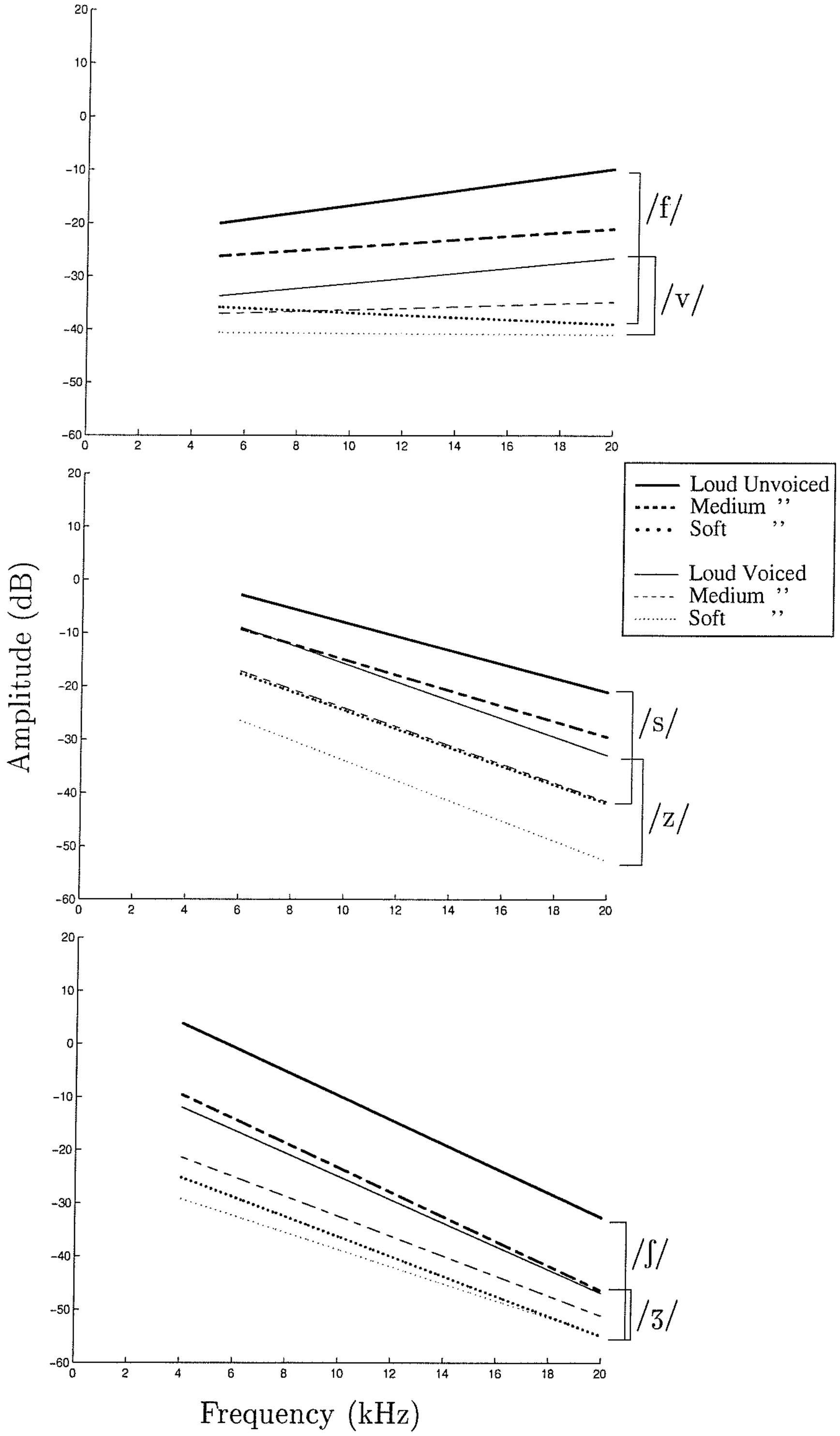

Figure 6.4: Average regression line fits (from $\overline{F}$ to 20 kHz) of sustained labiodental (top), alveolar (middle) and postalveolar (bottom) fricatives from Corpus 1b at loud, medium and soft effort levels. Speaker ISSS.

In a plot of A_d by fricative, as shown in Figure 6.5 for one subject, /s, z, ʃ, ʒ/ have higher A_d than /f, v/ as predicted. This holds for Corpora 1a and 1b for all subjects. A_d also tends to be lower for voiced fricatives than for their unvoiced counterparts, but this is less consistent across subject.

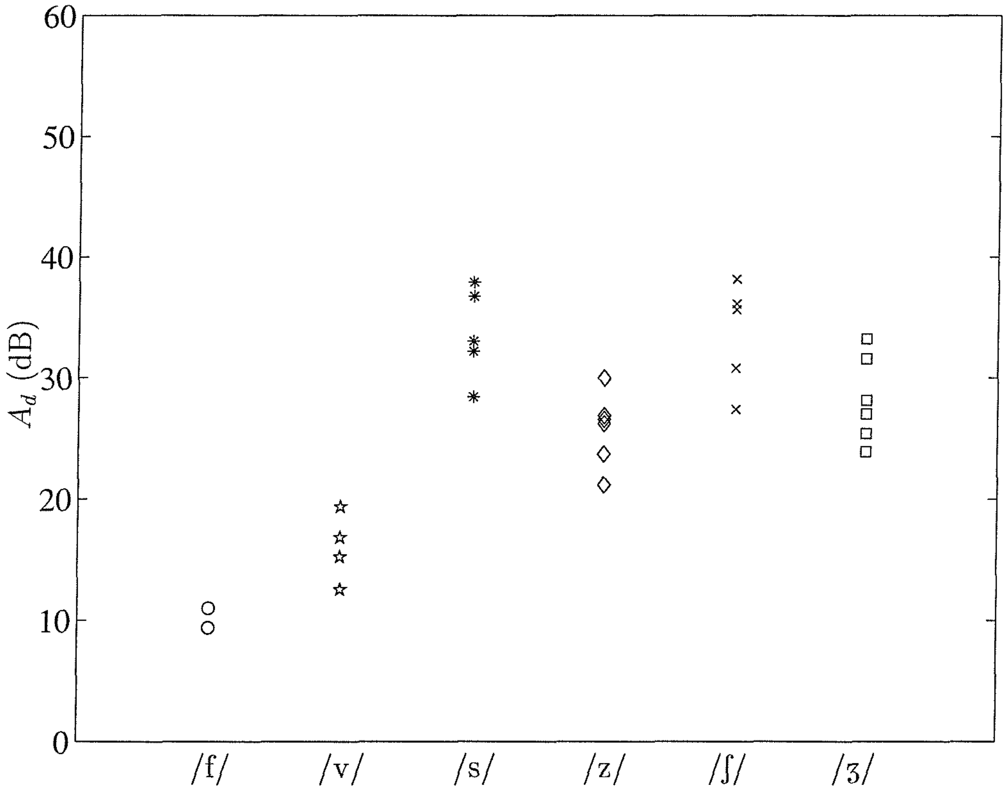

Figure 6.5: Dynamic amplitude of fricatives from Corpus 1a. Speaker LMTJ. Number of tokens per fricative: /f/ – 3 ; /v/ – 4; /s/ – 5; /z/ – 6; /ʃ/ – 5; /ʒ/ – 6.

Figure 6.6 shows S_p vs. effort level for subject ACC. Slope generally increases with increased effort level, though this pattern is much more consistent for unvoiced fricatives. This agrees with results in Shadle and Mair (1996) and the predictions in Table 6.1. However, in a $\overline{F}$ vs. (A_d or S'_p) vs. S_p graph there is no consistent effect of effort level, as shown in Figures 6.7 and 6.8.

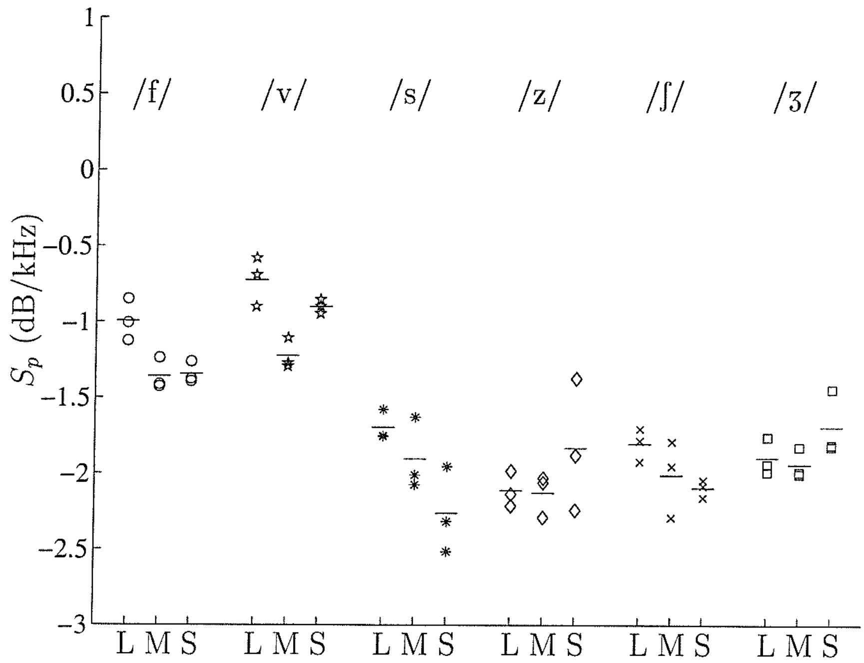

Figure 6.6: Spectral slope of sustained fricatives from Corpus 1b at Loud (L), Medium (M) and Soft (S) effort levels. The horizontal line is the average value of all the examples. Speaker ACC.

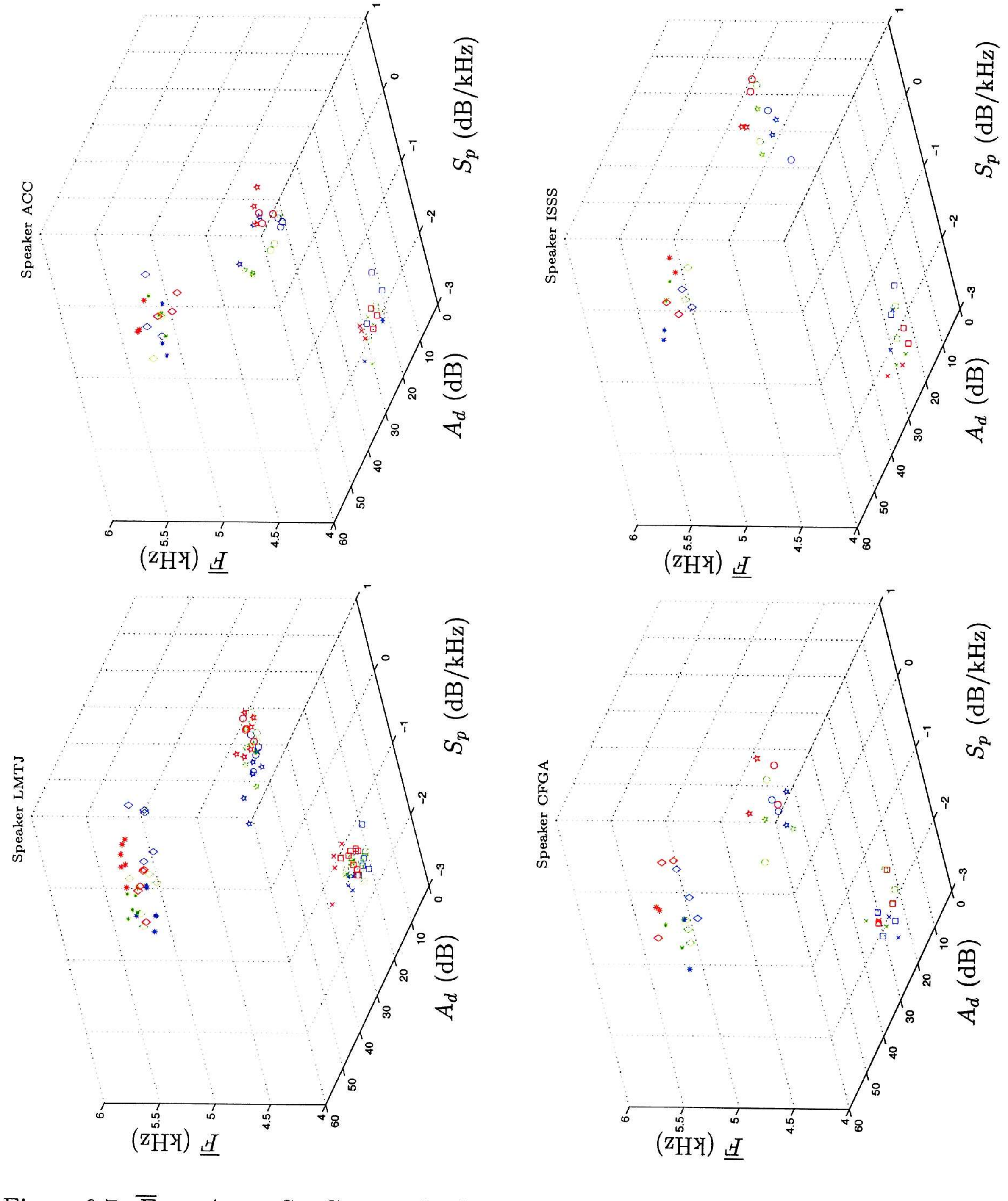

Figure 6.7: $\overline{F}$ vs. A_d vs. S_p. Corpus 1b: fricatives sustained at three different effort levels: soft (blue), medium (green) and loud (red). ∘ – /f/, ⋆ – /v/, ∗ – /s/, ⋄ – /z/, × – /ʃ/ and □ – /ʒ/.

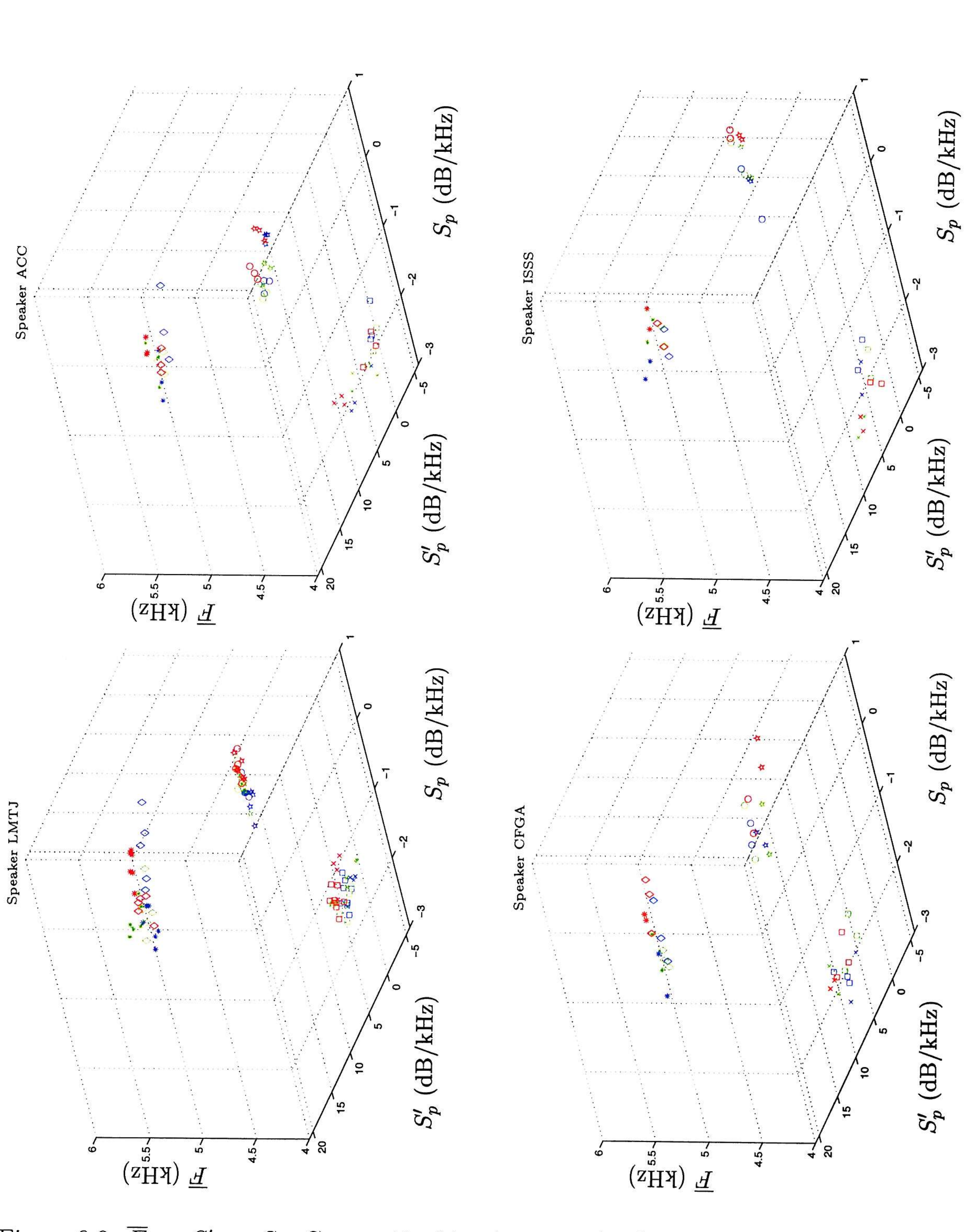

Figure 6.8: $\overline{F}$ vs. S'_p vs. S_p. Corpus 1b: fricatives sustained at three different effort levels: soft (blue), medium (green) and loud (red). $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\Box$ – /ʒ/.

The predictions made for A_d and S_p in Section 6.3 are shown in the schematic diagram of Figure 6.9. This expected clustering of sibilants separately from /f, v/ is borne out by the results for all subjects, as shown in Figure 6.10. If we use the value of $\overline{F}$ as a third dimension then on a $\overline{F}$ vs. A_d vs. S_p plot the fricatives cluster by place (labiodental, alveolar and postalveolar) as shown in Figure 6.11. $\overline{F}$ is not, of course, an independent parameter, but plotting in this way allows us to check data for each place for the influence of other factors.

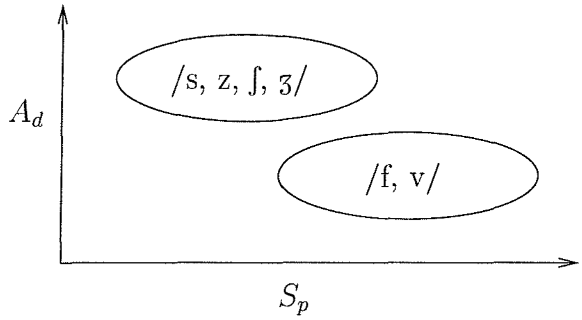

Figure 6.9: Predicted A_d vs. S_p relations for the fricatives.

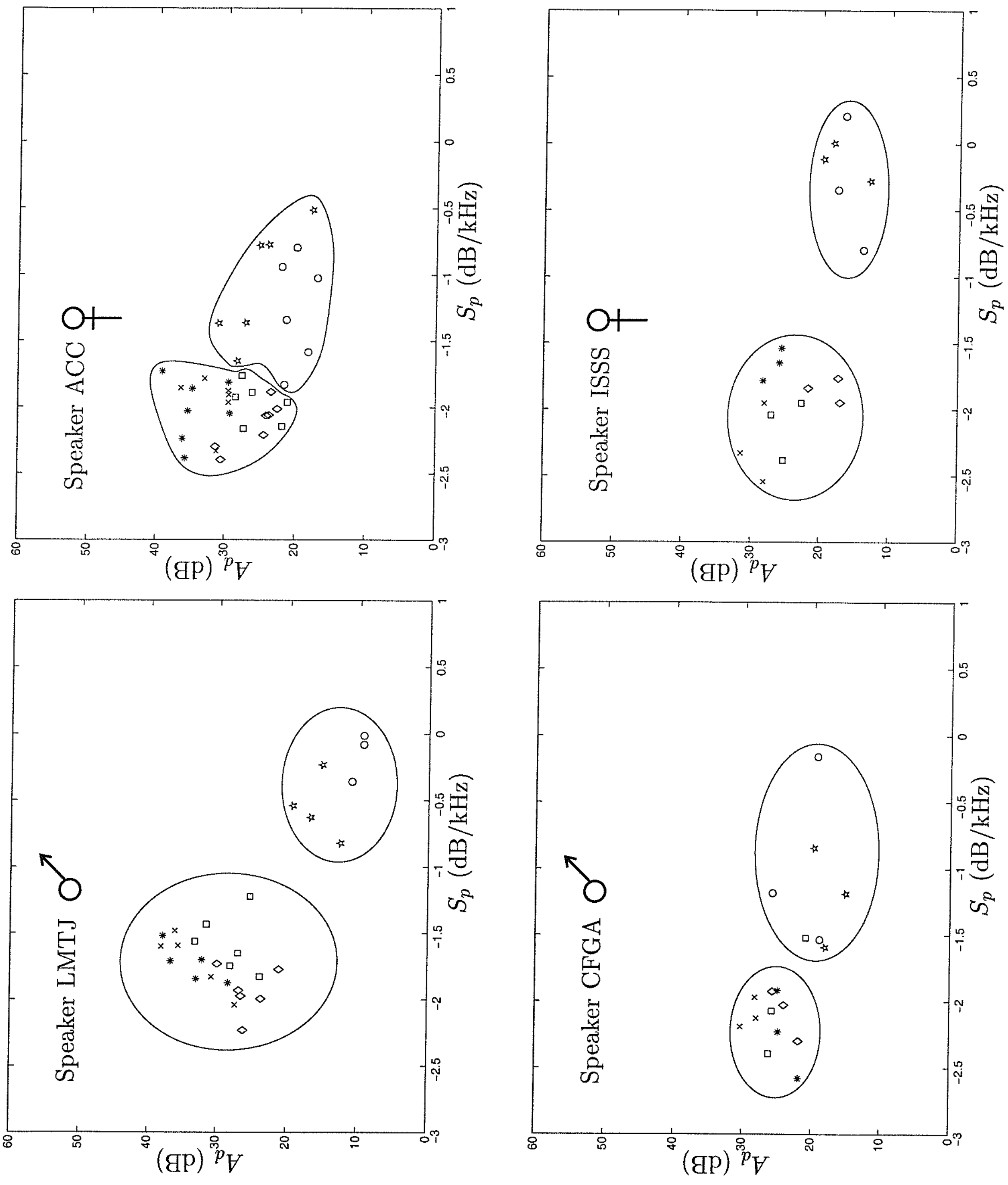

Figure 6.10: Corpus 1a (sustained fricatives), A_d vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/. Three different vowel contexts were used with each fricative.

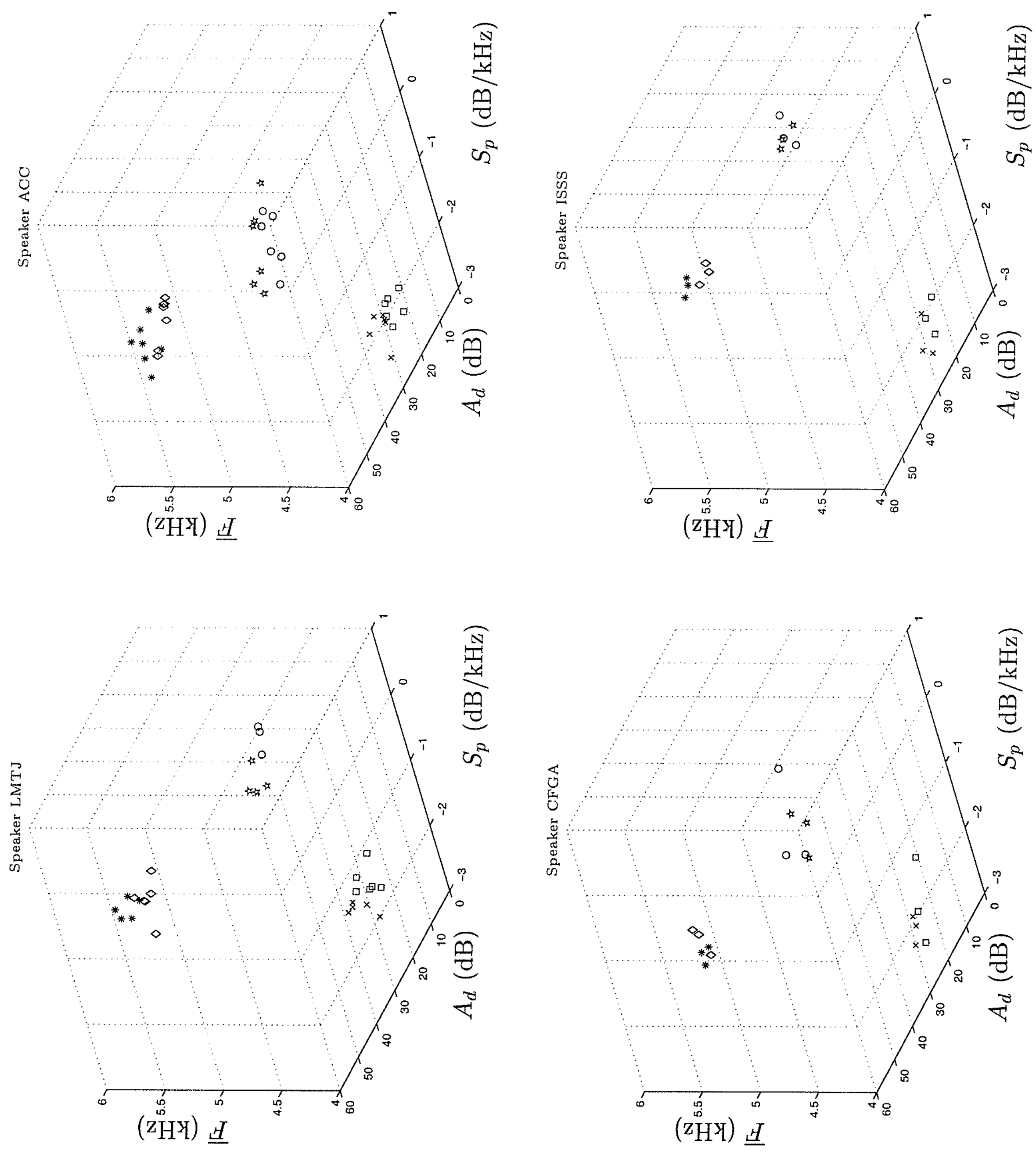

Figure 6.11: Corpus 1a (sustained fricatives), $\overline{F}$ vs. A_d vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/. Same data as plotted in Figure 6.10.

As shown in Figure 6.12, we predict that on a 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. Figure 6.13 shows S'_p vs. S_p values plotted for Corpus 1a. For all speakers except ACC, both predictions were borne out. For ACC the voicing relationship was maintained, but /s, z/ tokens fell in between the /ʃ/ and /ʒ/ tokens. If we plot a $\overline{F}$ vs. A_d vs. S_p graph the fricatives produced by all four speakers cluster by place, as shown in Figure 6.14.

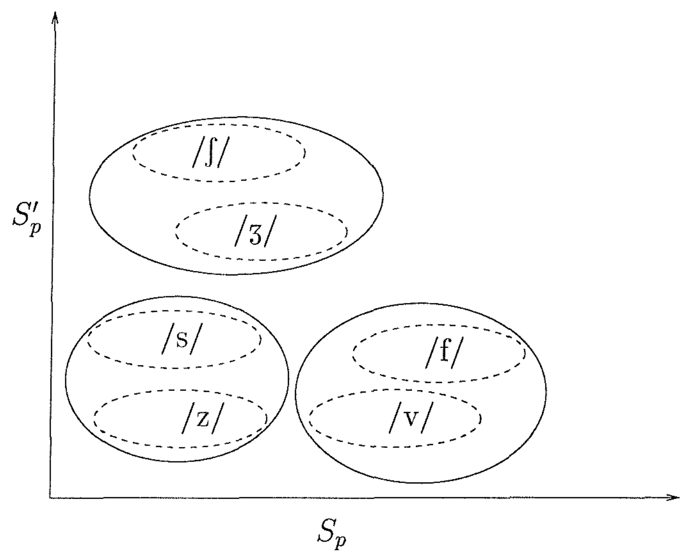

Figure 6.12: Predicted S'_p vs. S_p relations for the fricatives.

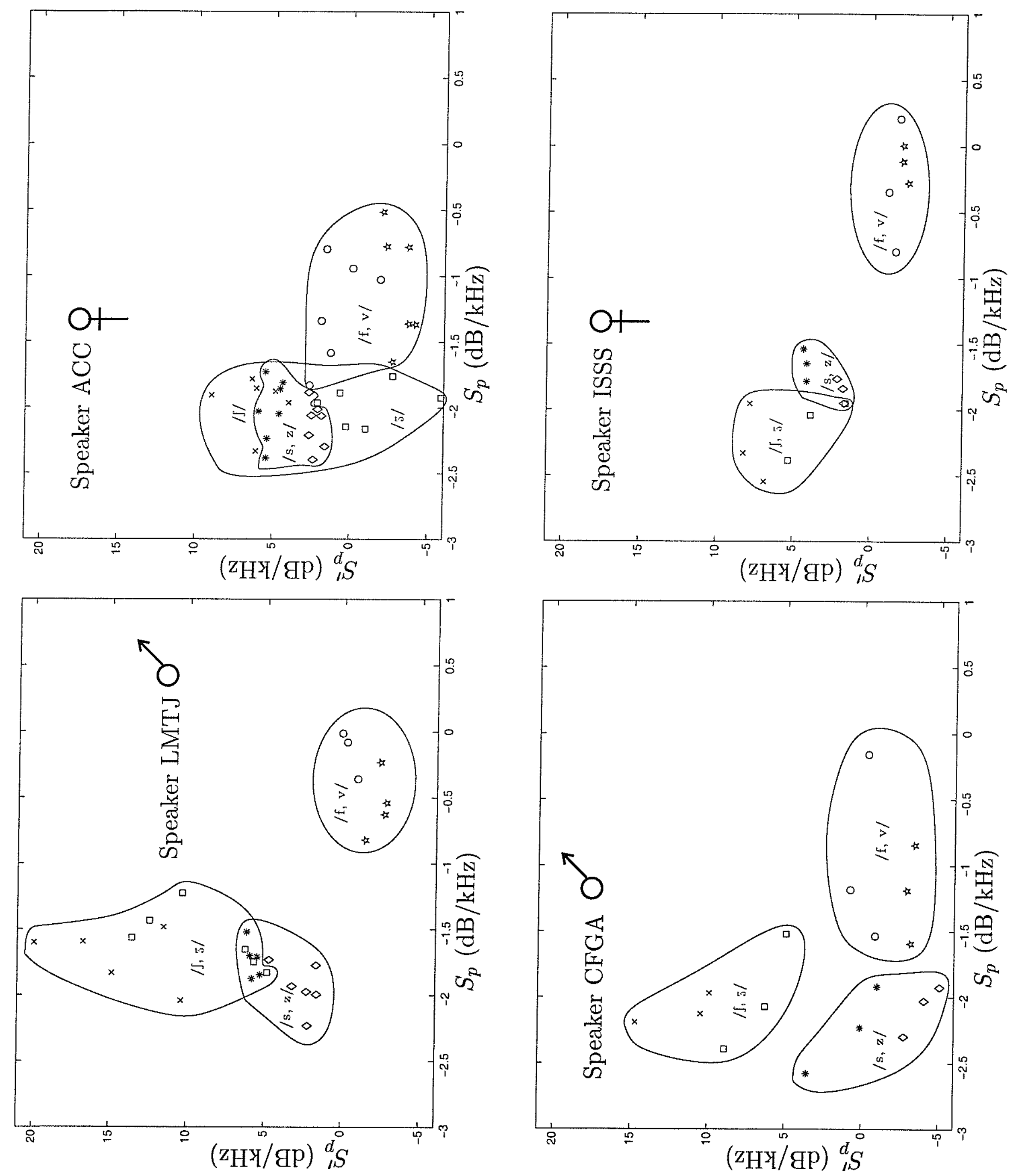

Figure 6.13: Corpus 1a (sustained fricatives), S'_p vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

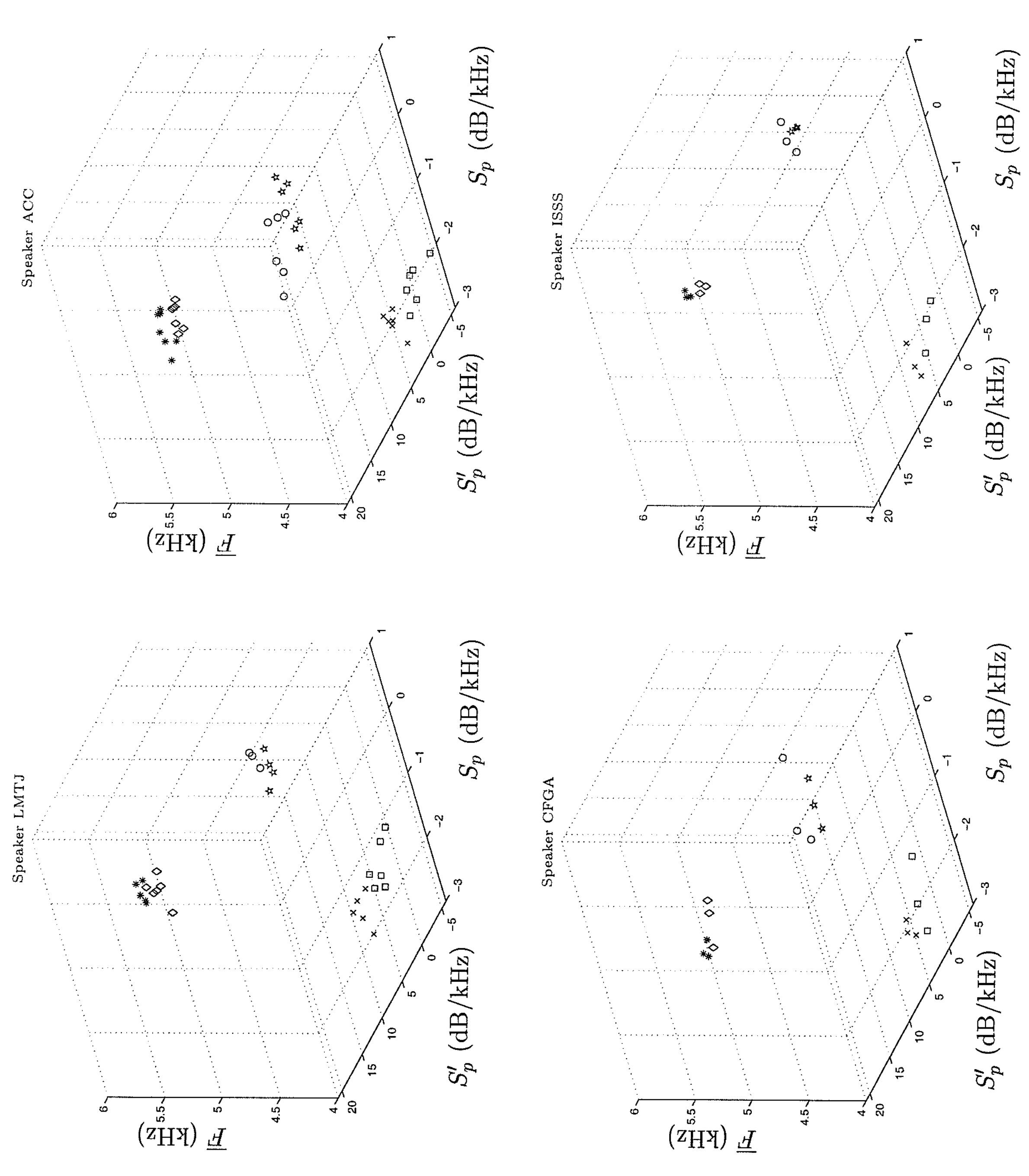

Figure 6.14: Corpus 1a (sustained fricatives), $\overline{F}$ vs. S'_p vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/. Same data as plotted in Figure 6.13.

6.4.2 Fricatives in Context

6.4.2.1 Fricatives in Nonsense Words (Corpus 2)

In Figures 6.15 and 6.16, A_d and S_p are plotted vs. location of the analysis window within the fricative (i.e. beginning, middle, or end) for Corpus 2. For /f, v/ there is no consistent pattern; results in Shadle et al. (1996) indicate that for these 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. S_p tends to be lower for the sibilants than for /f, v/, but has no consistent trend with regard to location of the analysis window within the fricative.

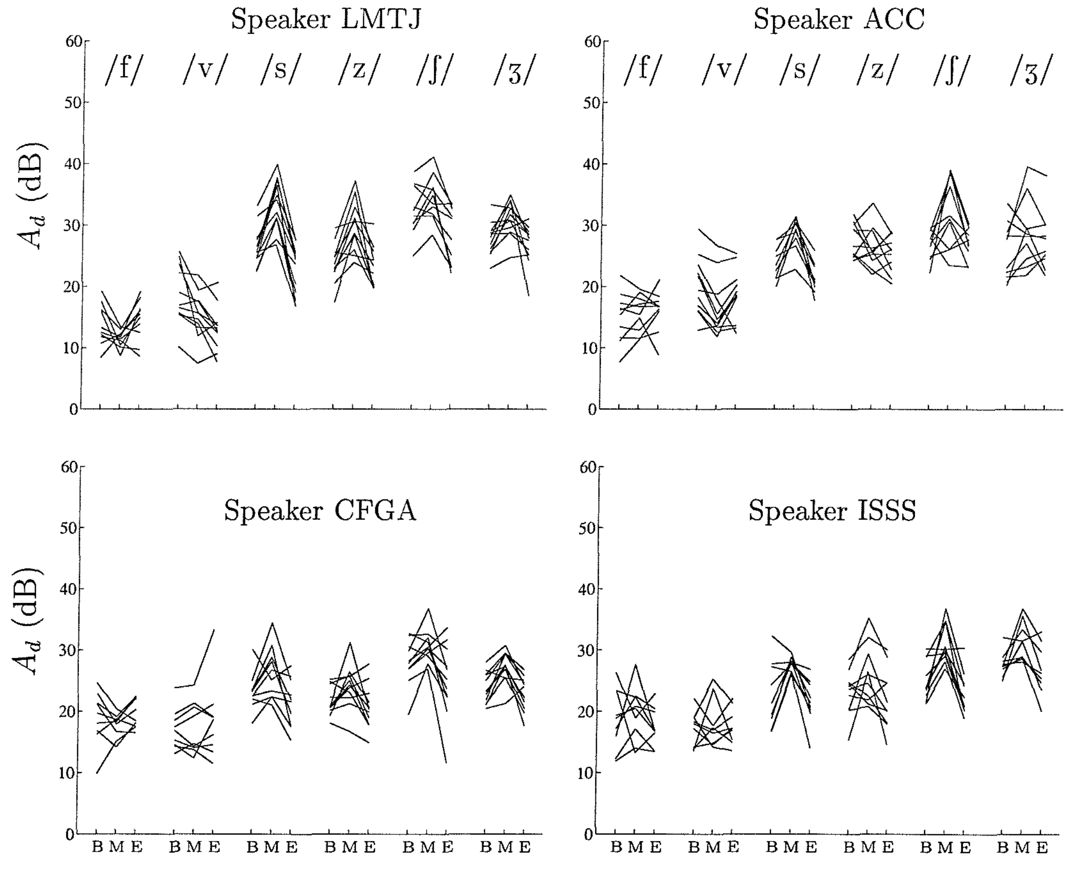

Figure 6.15: Dynamic amplitude of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative.

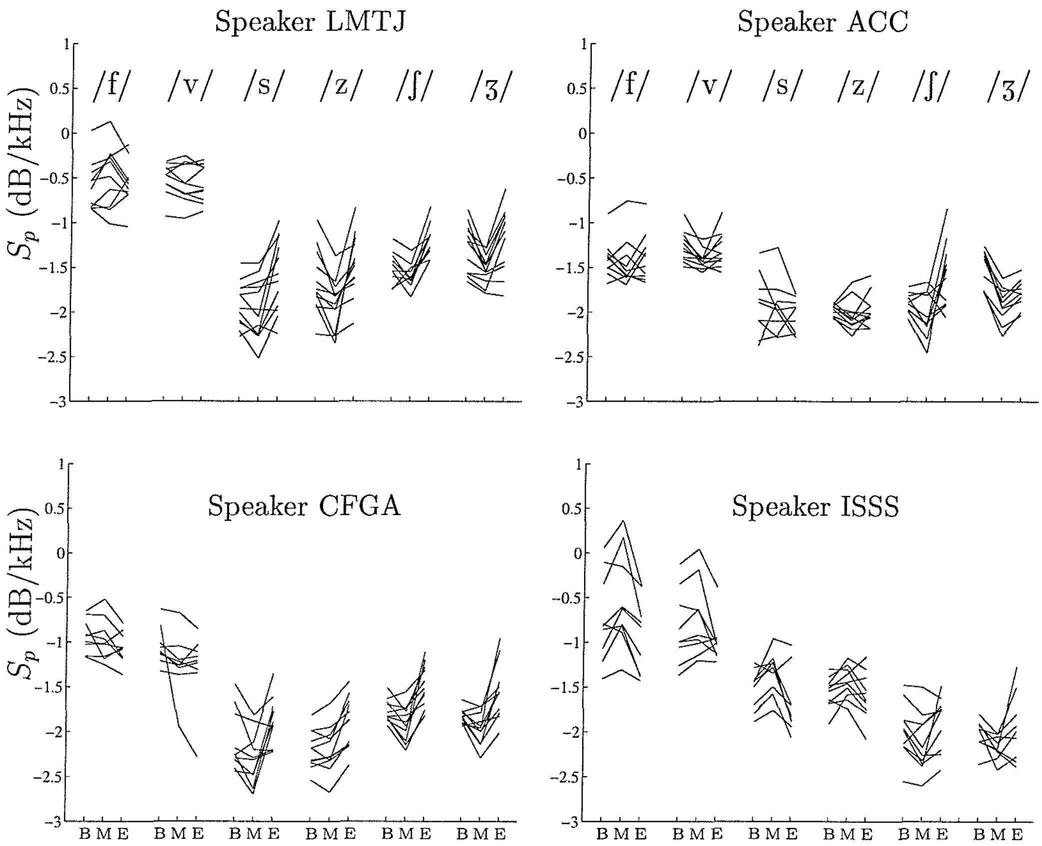

Figure 6.16: Spectral slope of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative.

On an $\overline{F}$ vs. (A_d or S'_p) vs. S_p graph the fricatives in nonsense words (Corpus 2) produced by all four speakers cluster by place as shown in Figures 6.17 and 6.18. When we use place knowledge, i.e. use $\overline{F}$, to plot A_d vs. S_p at the beginning, middle and end the results are also inconclusive, as shown in Figures 6.19 to 6.22.

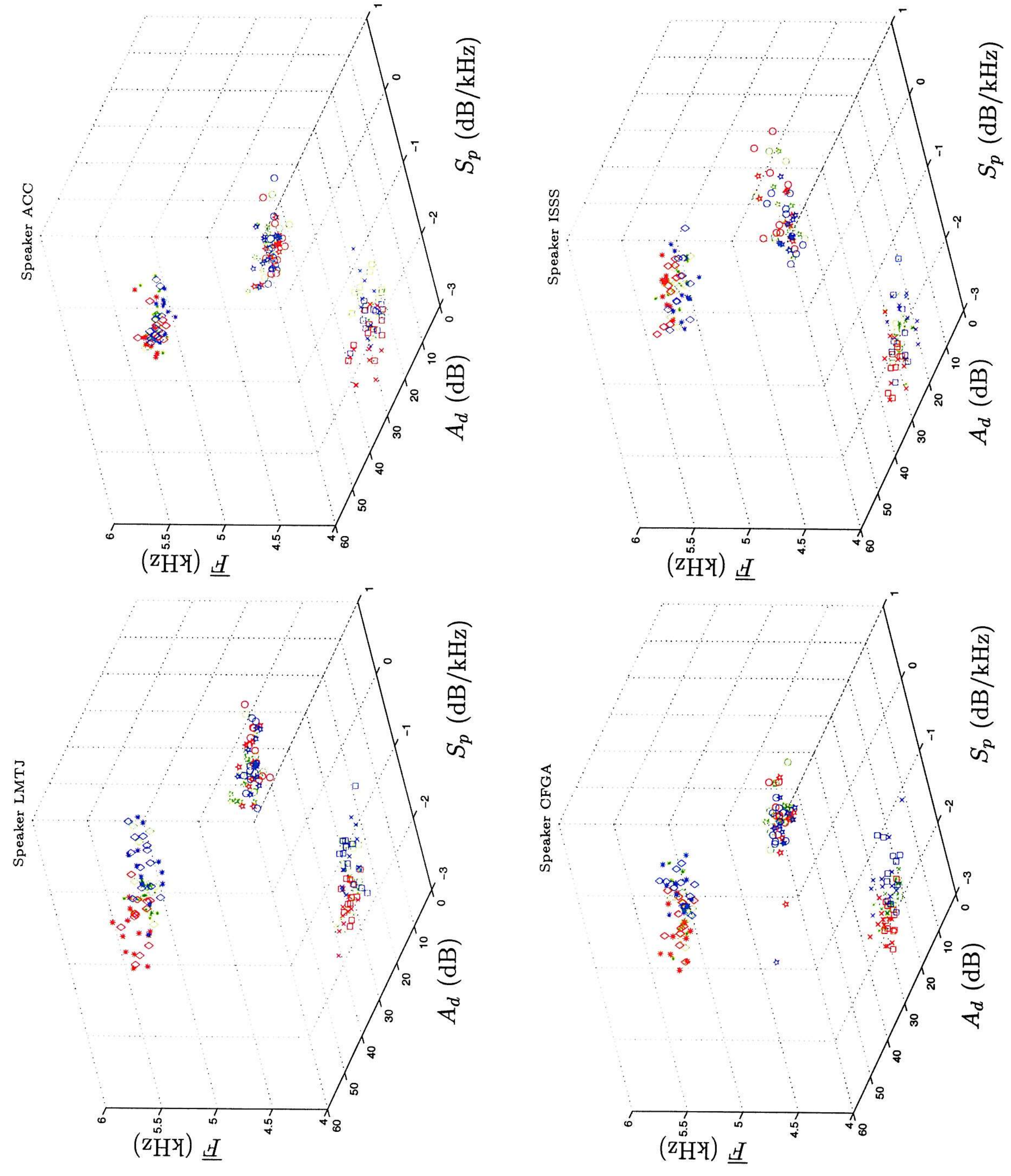

Figure 6.17: Corpus 2, $\overline{F}$ vs. A_d vs. S_p at the end (blue), beginning (green) and middle (red) of the fricative. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

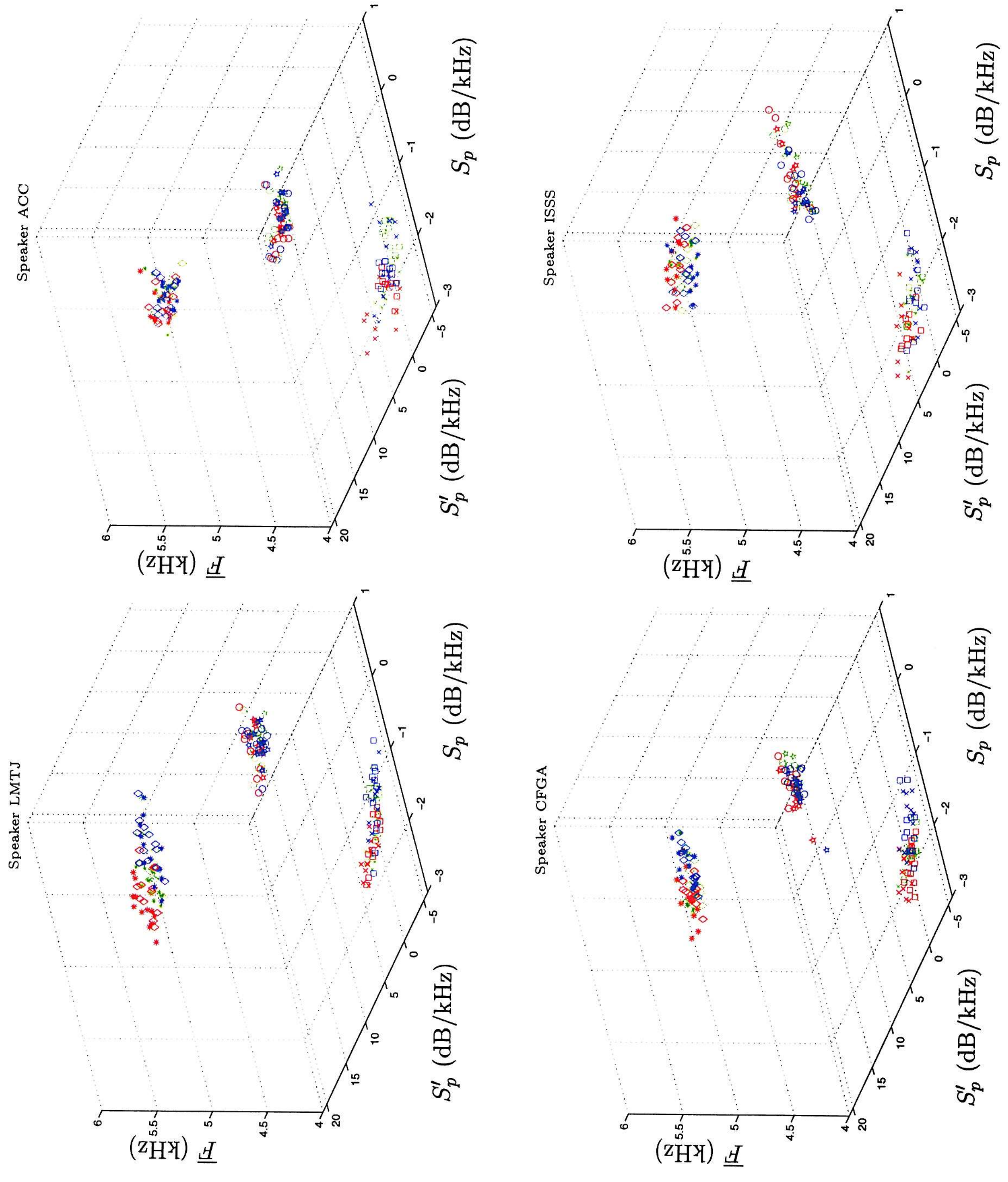

Figure 6.18: Corpus 2, $\overline{F}$ vs. S'_p vs. S_p at the end (blue), beginning (green) and middle (red) of the fricative. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

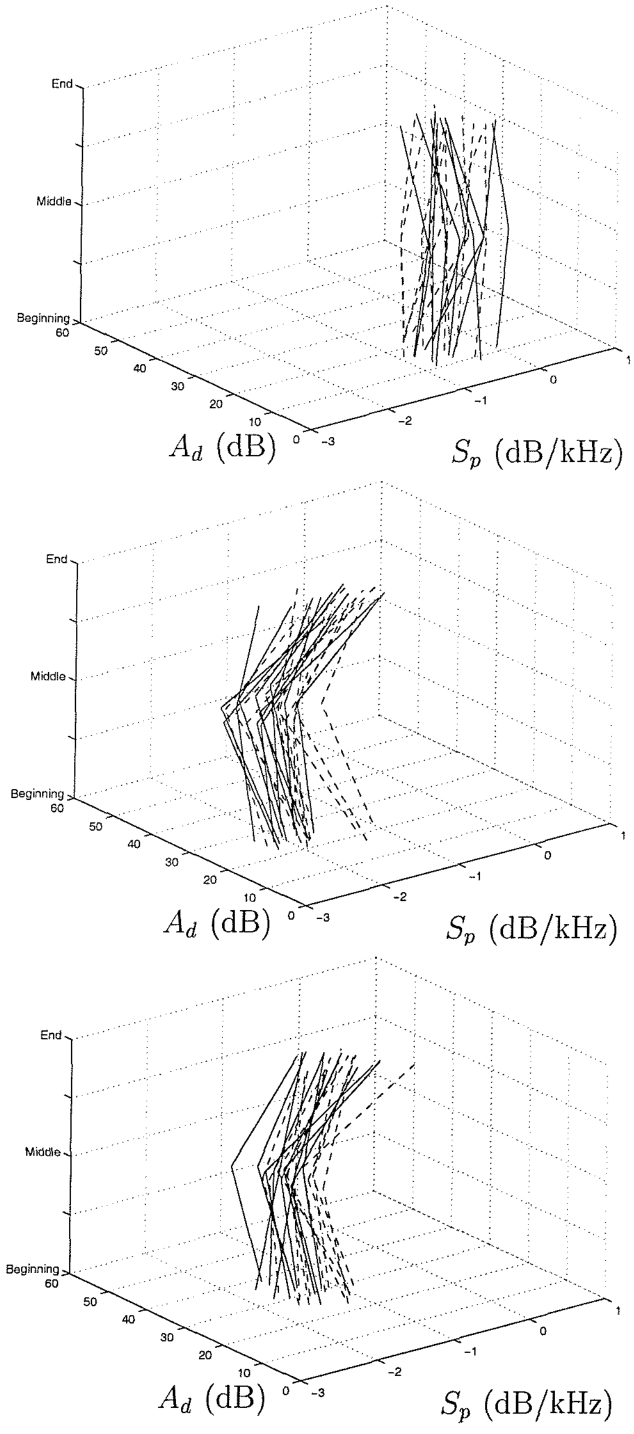

Figure 6.19: A_d vs. S_p at the beginning, middle and end of the fricative. Top: Labiodental (/f/ – solid line, and /v/ – dashed line); Middle: Alveolar (/s/ – solid line, and /z/ – dashed line); Bottom: Postalveolar (/ʃ/ – solid line, and /ʒ/ – dashed line). Corpus 2 (Speaker LMTJ).

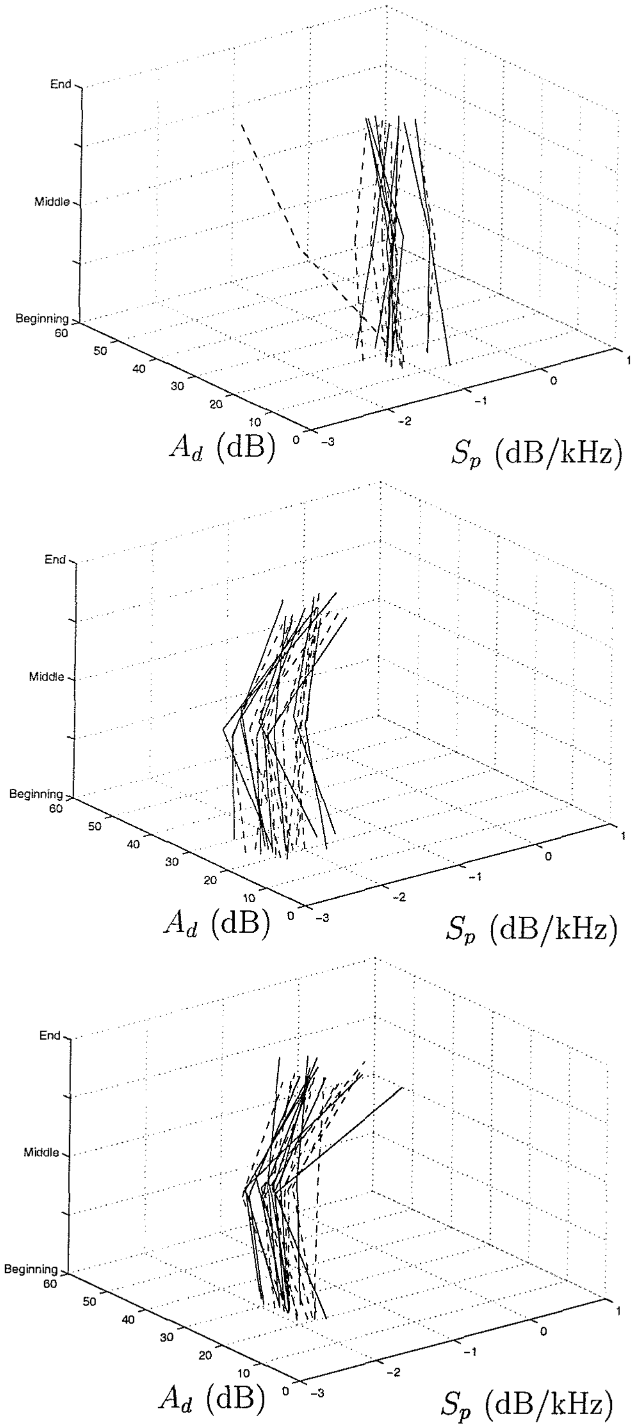

Figure 6.20: A_d vs. S_p at the beginning, middle and end of the fricative. Top: Labiodental (/f/ – solid line, and /v/ – dashed line); Middle: Alveolar (/s/ – solid line, and /z/ – dashed line); Bottom: Postalveolar (/ʃ/ – solid line, and /ʒ/ – dashed line). Corpus 2 (Speaker CFGA).

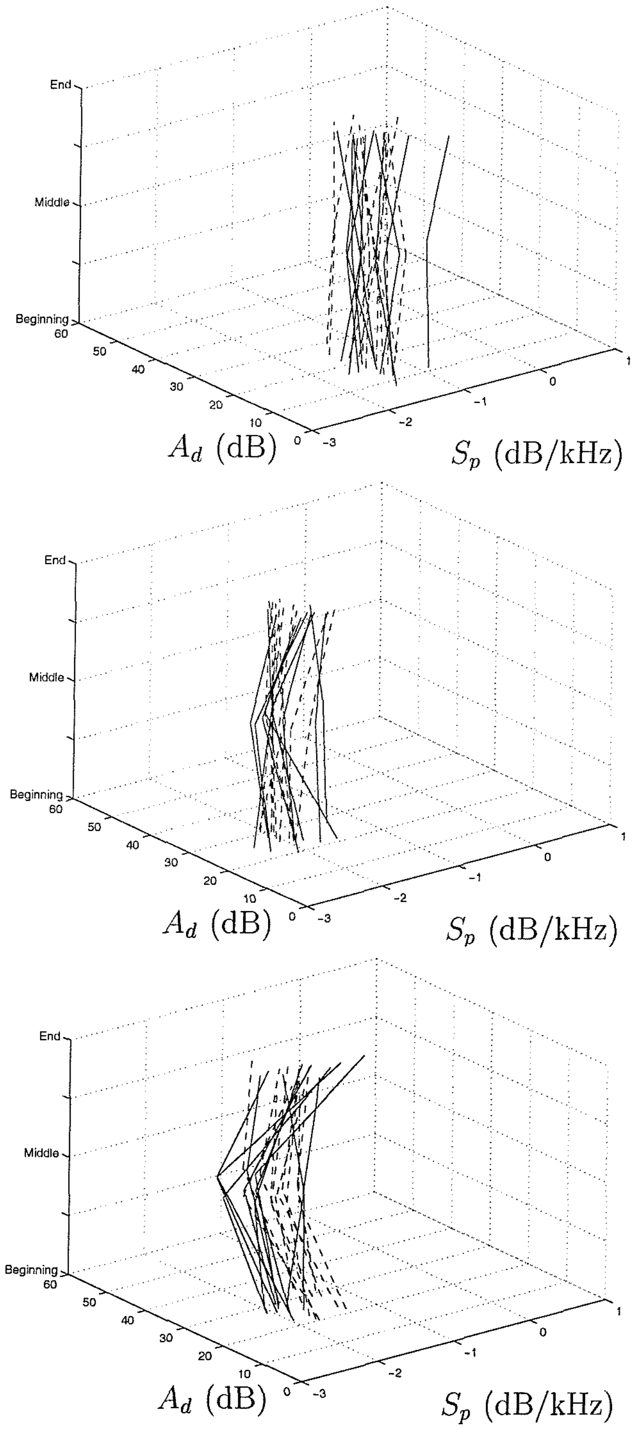

Figure 6.21: A_d vs. S_p at the beginning, middle and end of the fricative. Top: Labiodental (/f/ – solid line, and /v/ – dashed line); Middle: Alveolar (/s/ – solid line, and /z/ – dashed line); Bottom: Postalveolar (/ʃ/ – solid line, and /ʒ/ – dashed line). Corpus 2 (Speaker ACC).

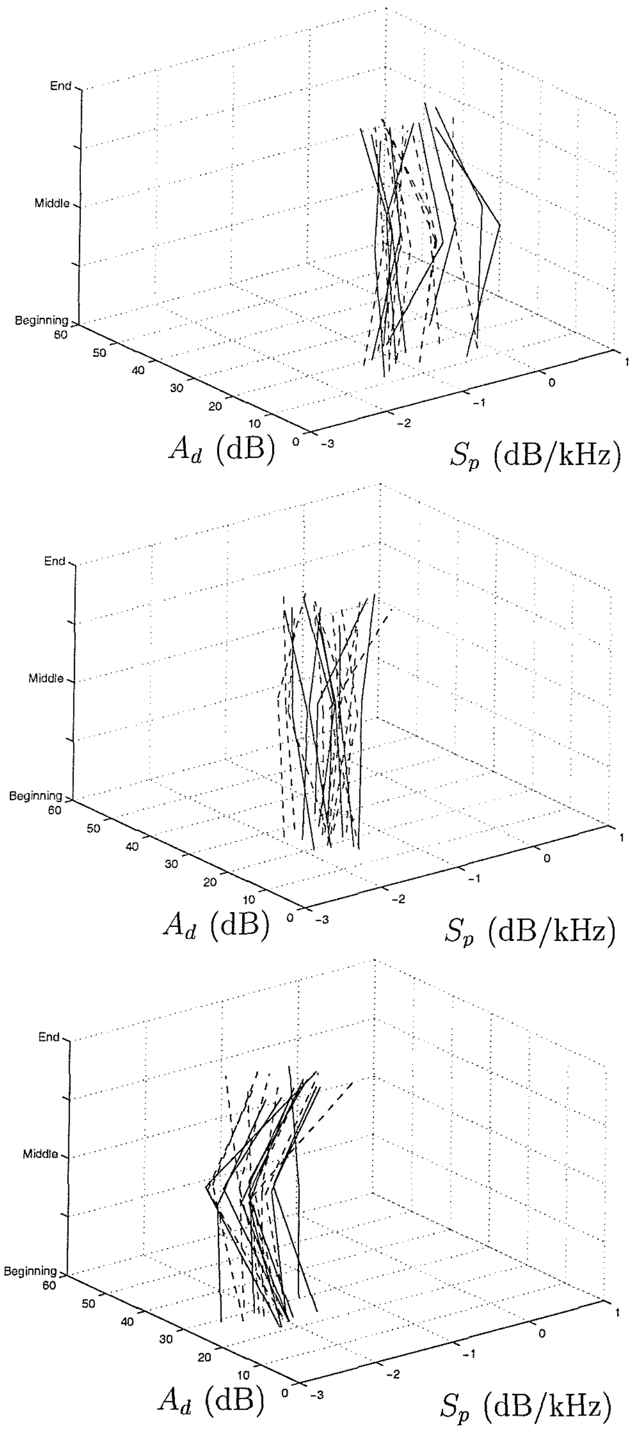

Figure 6.22: A_d vs. S_p at the beginning, middle and end of the fricative. Top: Labiodental (/f/ – solid line, and /v/ – dashed line); Middle: Alveolar (/s/ – solid line, and /z/ – dashed line); Bottom: Postalveolar (/ʃ/ – solid line, and /ʒ/ – dashed line). Corpus 2 (Speaker ISSS).

Comparisons of stressed and unstressed fricatives indicate little or no change in A_d and S_p, as shown in Figures 6.23 and 6.24 for Corpus 2 fricatives, which is not as predicted. We expected stress to act as increased effort level; that is, we expected both A_d and S_p to be higher in stressed than in unstressed syllables. However, it is clear that syllable stress does not affect fricatives in context the way effort level affects sustained fricatives.

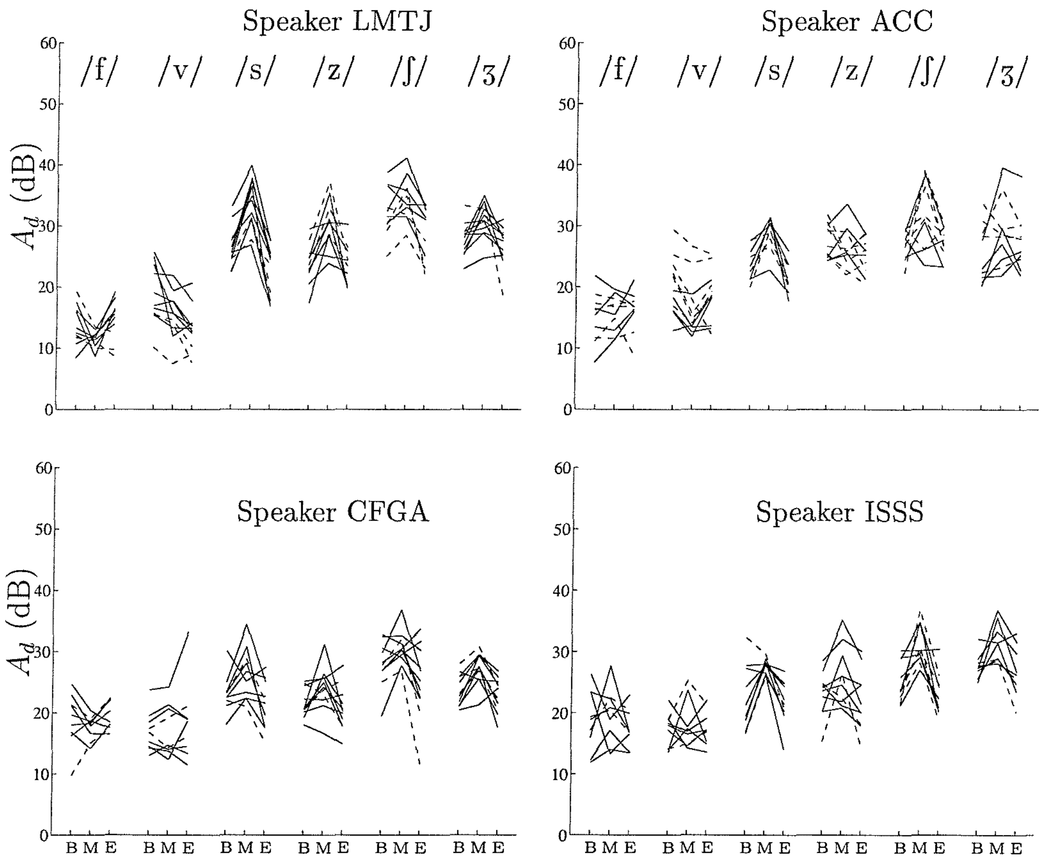

Figure 6.23: Dynamic amplitude of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative. Fricatives in stressed syllables (solid line) and unstressed syllables (dashed line).

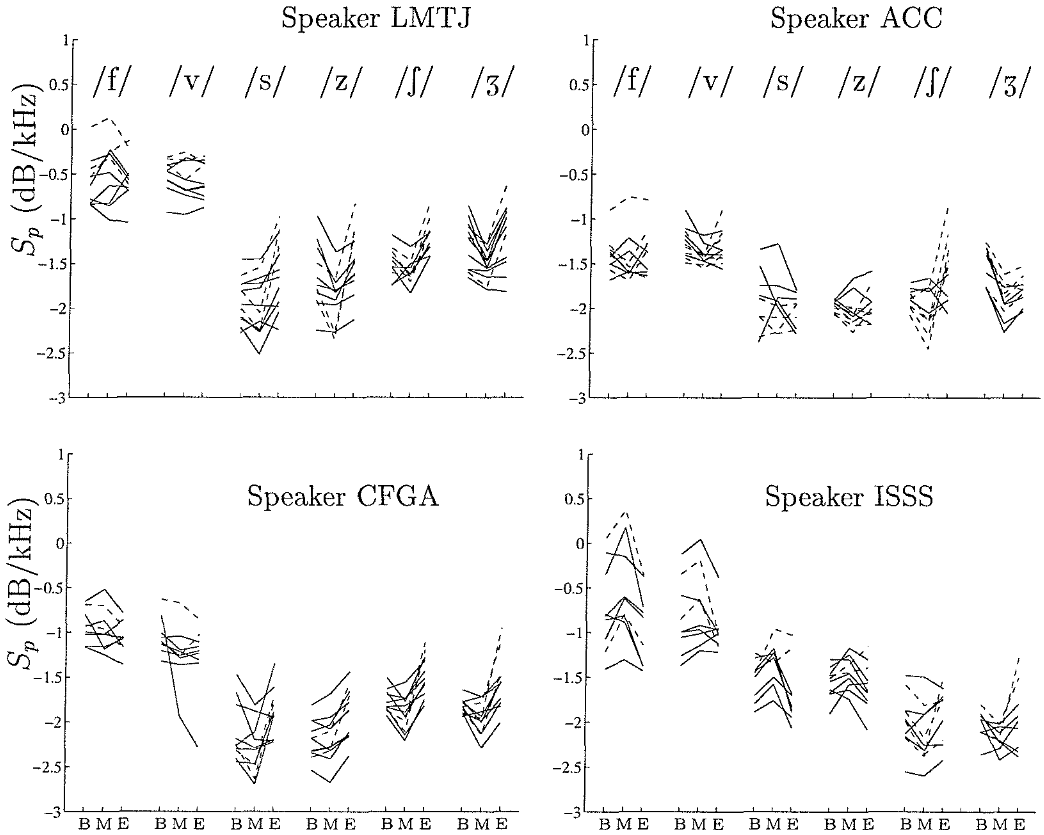

Figure 6.24: Spectral slope of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative. Fricatives in stressed syllables (solid line) and unstressed syllables (dashed line).

Analysis of Corpus 2 fricatives, which occur in rounded and unrounded vowel context, revealed no consistent effect of rounding on the values of A_d and S_p, as shown in Figures 6.25 to 6.28. It is possible that F, if measured for this corpus, would show an effect, but the results from Corpus 1a indicate it unlikely to be a strong or significant one.

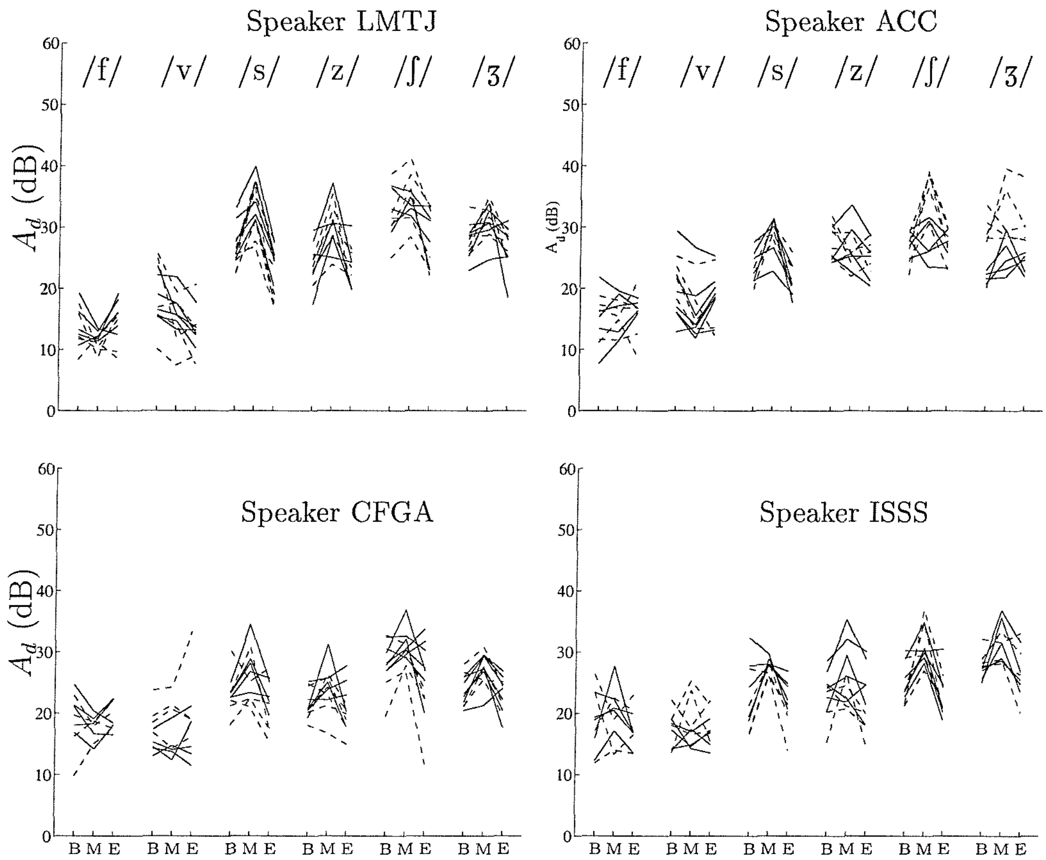

Figure 6.25: Dynamic amplitude of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative. Fricatives in rounded vowel context (solid line) and unrounded vowel context (dashed line).

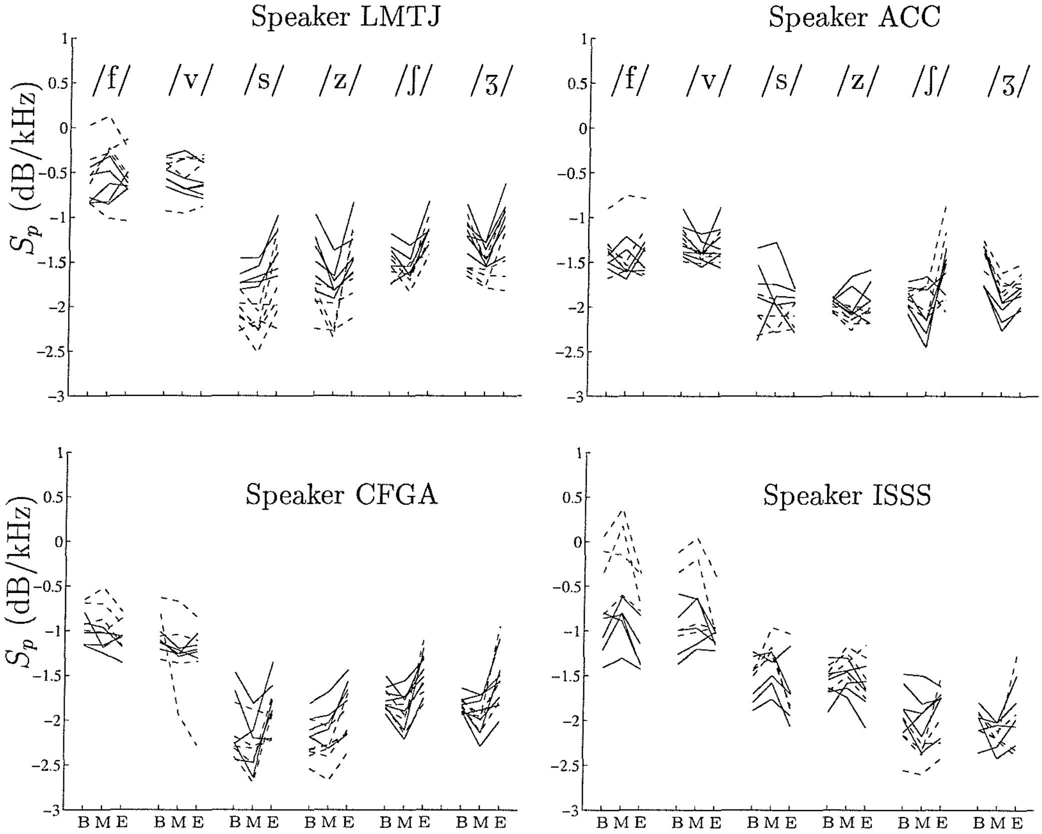

Figure 6.26: Spectral slope of fricatives from Corpus 2, at the Beginning (B), Middle (M) and End (E) of the fricative. Fricatives in rounded vowel context (solid line) and unrounded vowel context (dashed line).

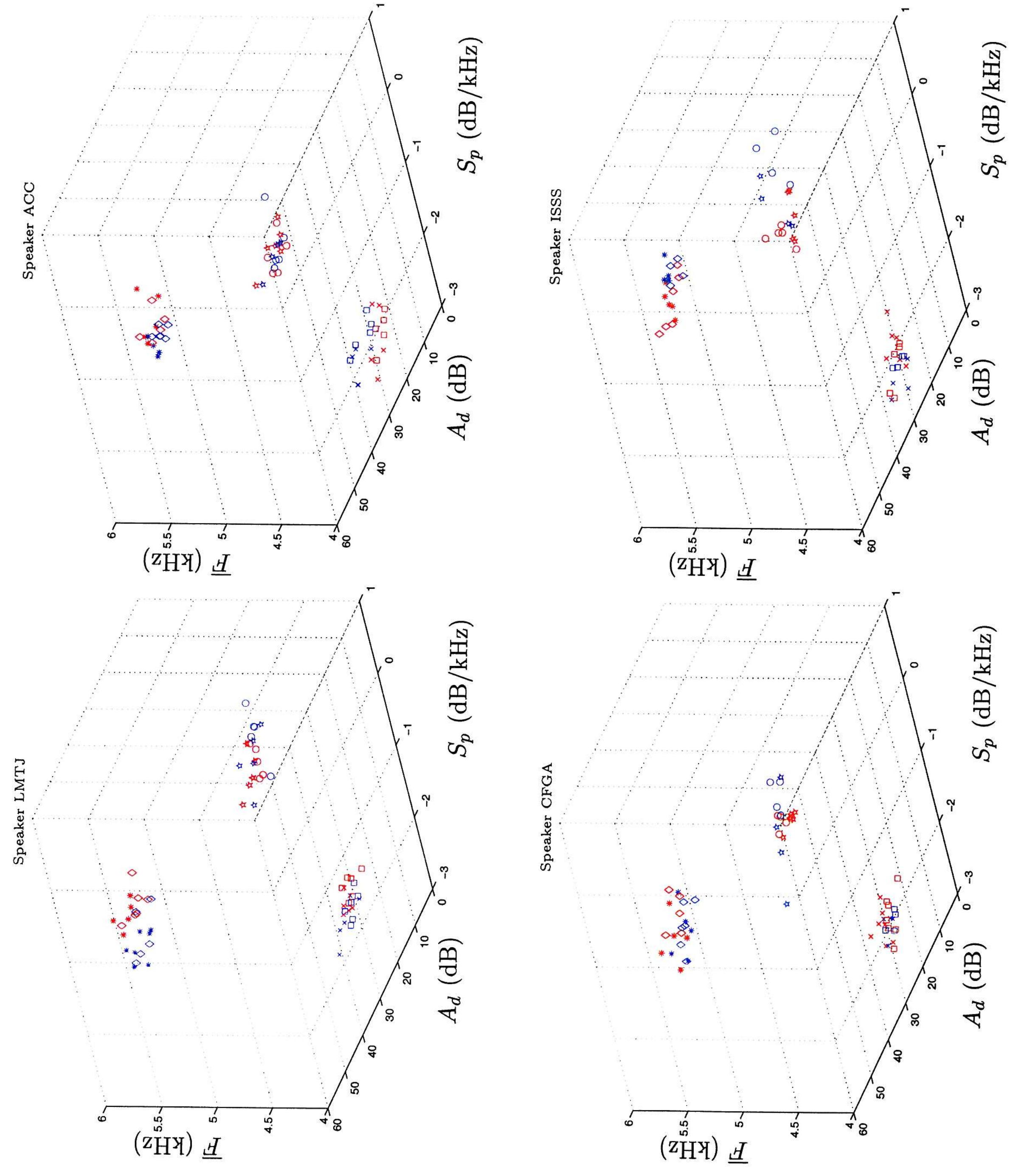

Figure 6.27: Corpus 2, $\overline{F}$ vs. A_d (middle) vs. S_p (middle). Fricatives in rounded vowel context (red) and unrounded vowel context (blue). $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

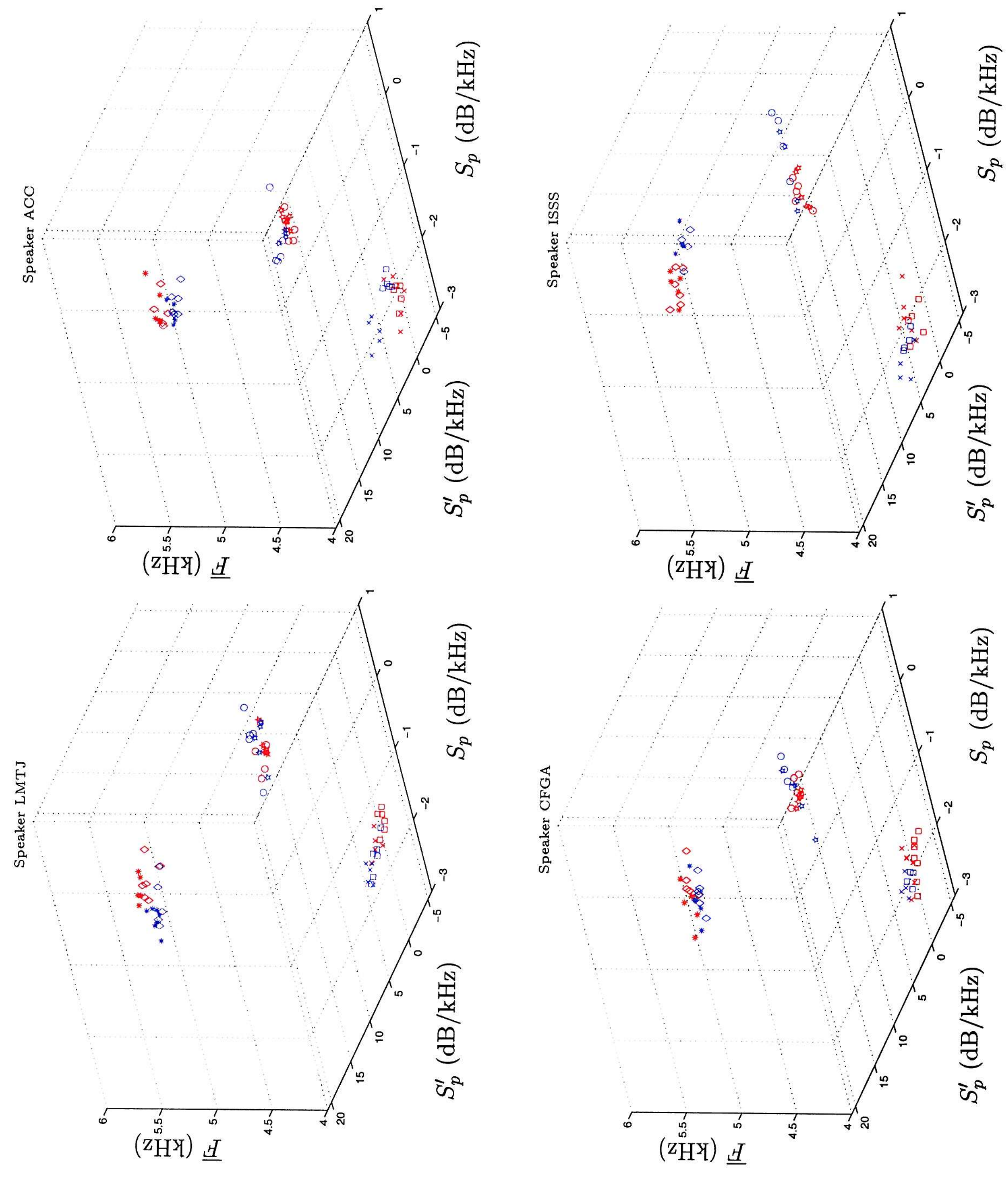

Figure 6.28: Corpus 2, $\overline{F}$ vs. S'_p (middle) vs. S_p (middle). Fricatives in rounded vowel context (red) and unrounded vowel context (blue). $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

We also wished to pursue stress: were the parameters affected noticeably? Did the overall amplitude of the spectrum of a fricative vary noticeably with stress? To answer these questions, we superimposed the ensemble-averaged spectra computed at the middle of all examples of each fricative in Corpus 2 (Speaker LMTJ). The overall amplitude of stressed and unstressed fricatives is the same within ± 5 dB. The only significant difference seems to be the amplitude of the fundamental frequency component of voiced fricatives, which is 10-15 dB higher for stressed than for unstressed examples. It thus appears that stress results in a stronger voicing source, but does not affect the fricative noise source.

6.4.2.2 Fricatives in Real Words (Corpus 3 and 4)

On $\overline{F}$ vs.(A_d or S_p') vs. S_p graphs the fricatives in real words (Corpus 3 and 4), produced by all four speakers, cluster by place as shown in Figures 6.29 to 6.32. Of course, $\overline{F}$ is not an independent parameter, but given the large number of tokens, using $\overline{F}$ as in Figures 6.29 to 6.32 allows us to check for other relationships more easily. Unlike in Corpus 1a, there is no clear relationship between voiced and unvoiced versions of the same place.

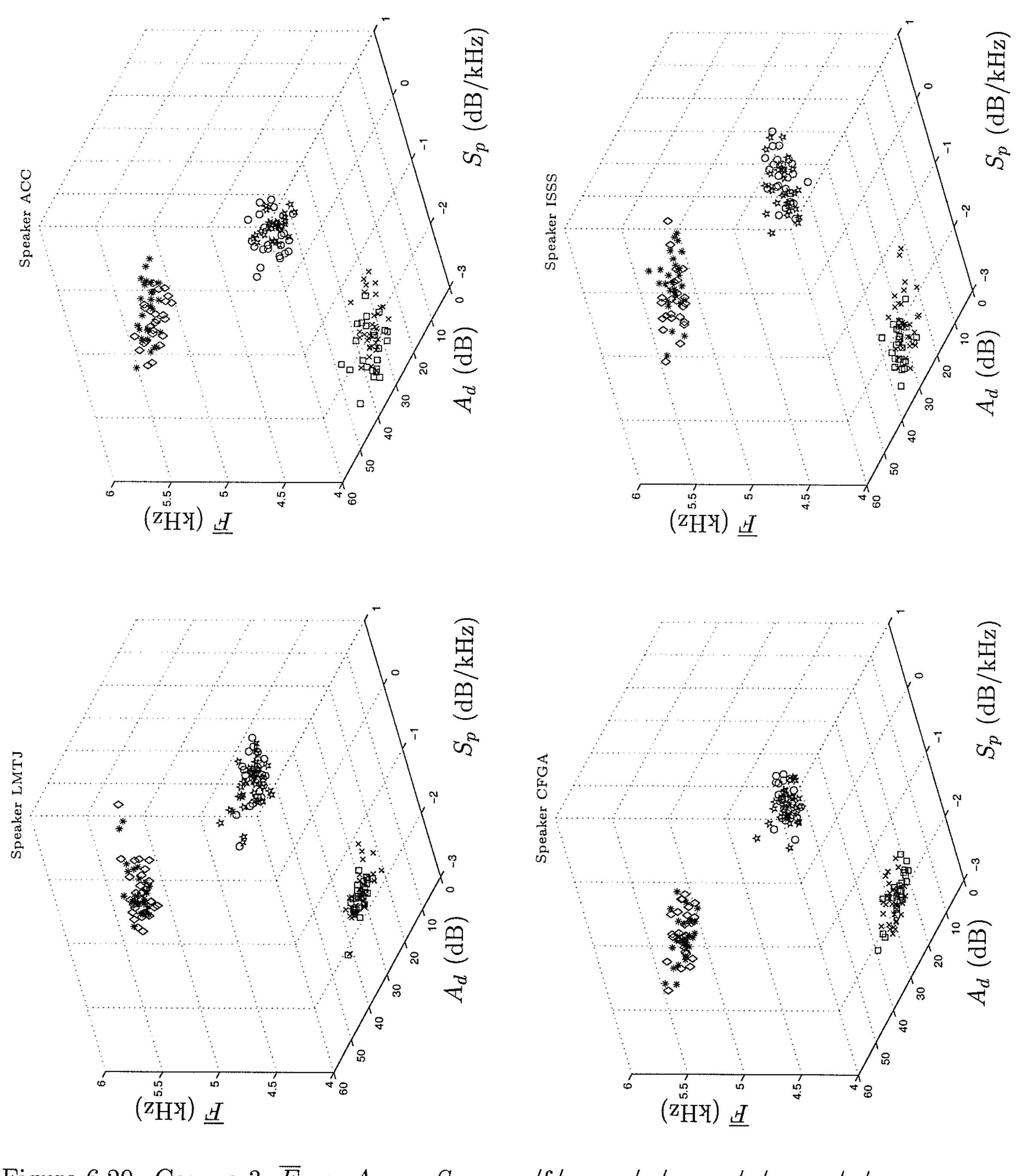

Figure 6.29: Corpus 3, $\overline{F}$ vs. A_d vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

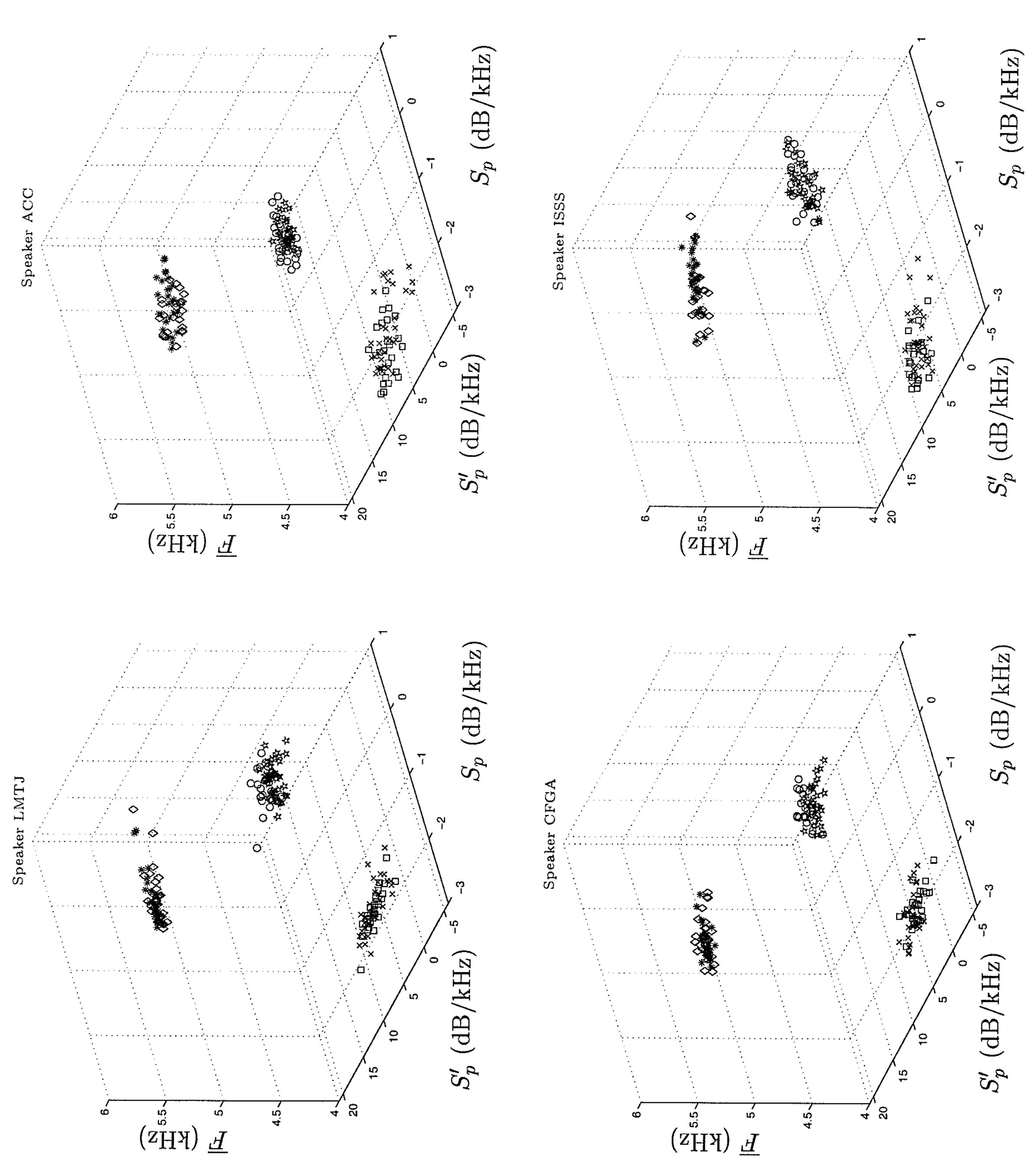

Figure 6.30: Corpus 3, F vs. S'_p vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

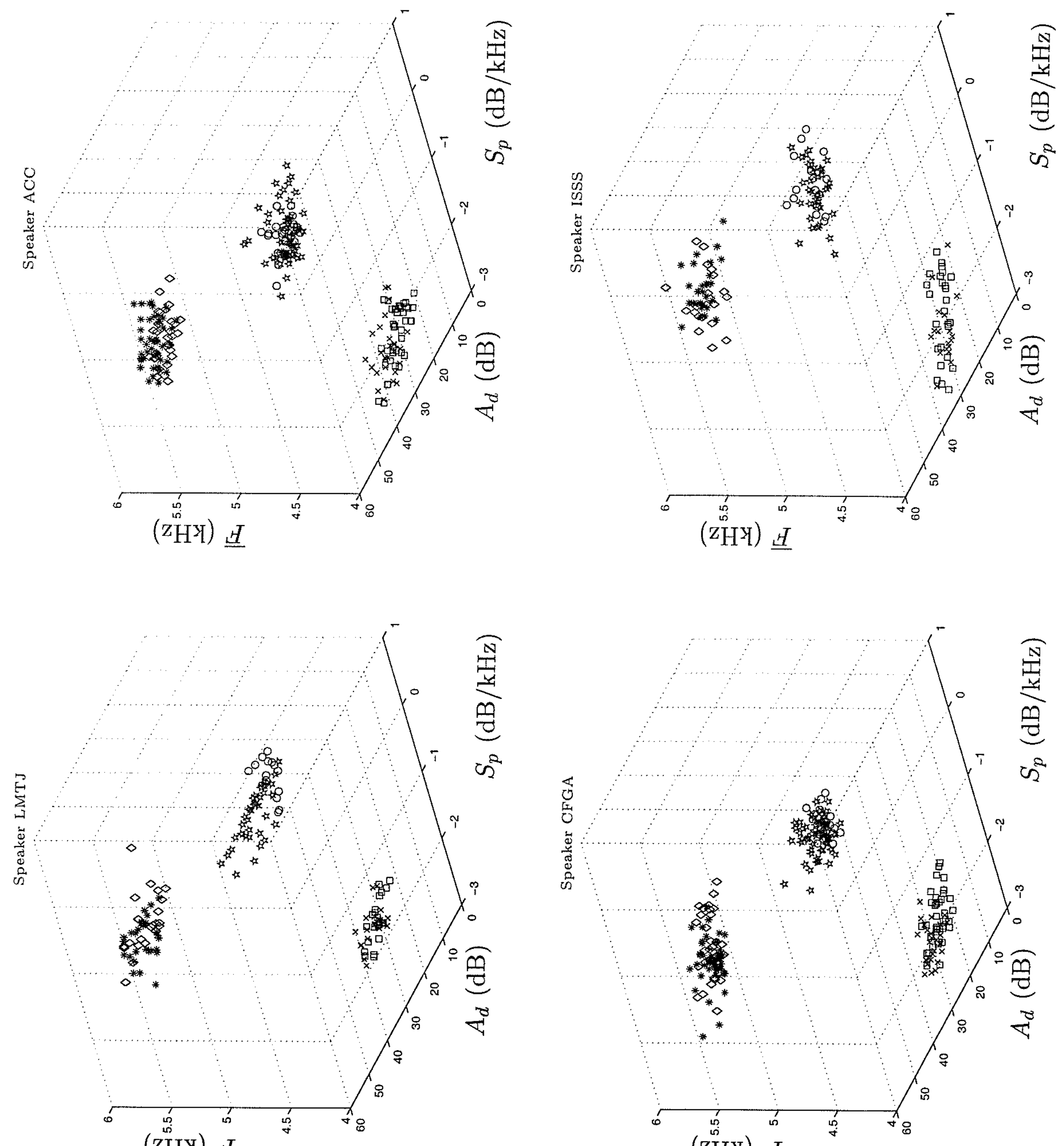

Figure 6.31: Corpus 4, $\overline{F}$ vs. A_d vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

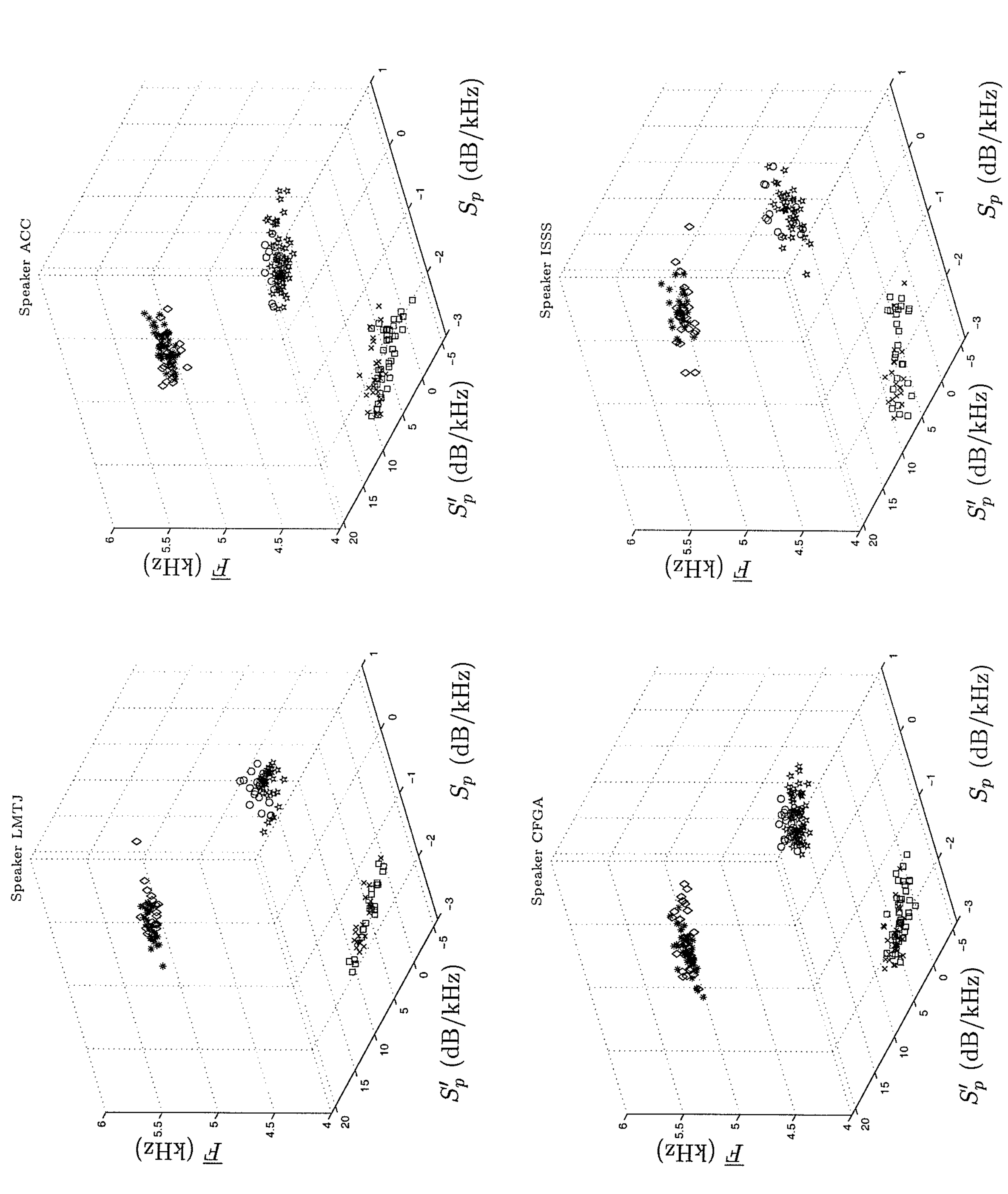

Figure 6.32: Corpus 4, $\overline{F}$ vs. S'_p vs. S_p. $\circ$ – /f/, $\star$ – /v/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

6.4.2.3 Correlations Between Duration, Word Position, Vowel Context, Stress and Devoicing

Duration and devoicing correlations were also examined, as well as the relationship to the parameters A_d and S_p. There does not seem to be any correlation between these two factors (duration and devoicing) and the values of our parameters. There is some evidence, however, that the *vowel to fricative duration ratio* (Hogan and Rozsypal 1980) might be a better candidate than the duration of the fricative to make such a comparison.

We also studied the relationships between the values of A_d, S_p and duration, and various other contextual factors (stressed / unstressed syllables; word - initial / medial / final fricatives; voiced / partially - devoiced / devoiced fricatives), but without identifying any significant trends, as shown in Figures 6.33 to 6.35.

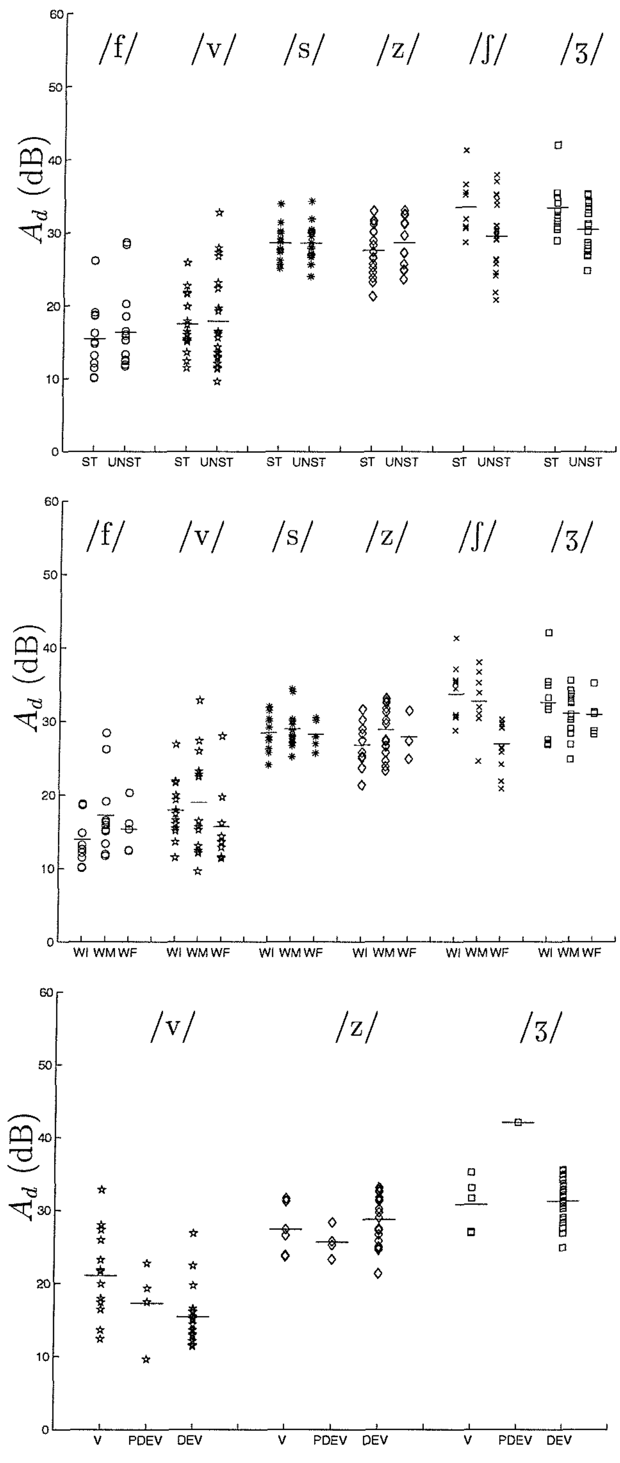

Figure 6.33: Dynamic amplitude for: fricatives in stressed (ST) and unstressed (UNST) syllables; word-initial (WI), word-medial (WM) and word-final (WF) fricatives; voiced (V), partially devoiced (PDEV) and devoiced (DEV) fricatives. Corpus 3 (Speaker LMTJ).

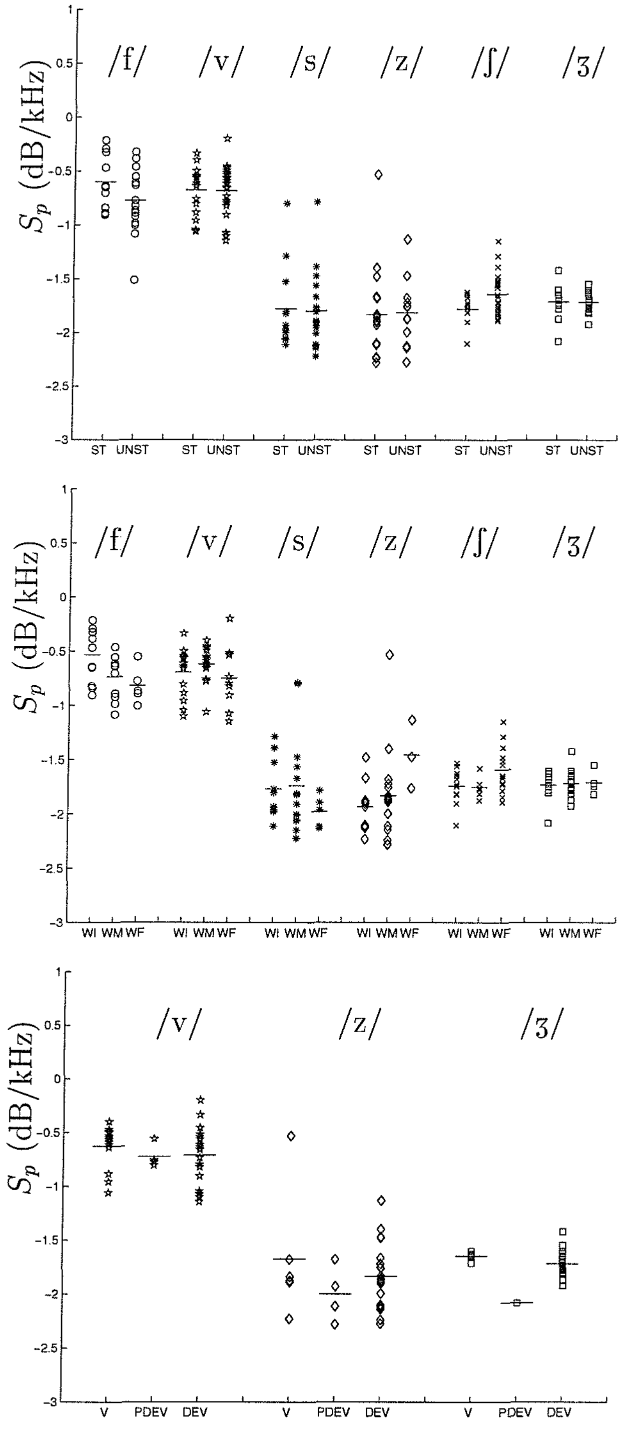

Figure 6.34: Spectral slope for: fricatives in stressed (ST) and unstressed (UNST) syllables; word-initial (WI), word-medial (WM) and word-final (WF) fricatives; voiced (V), partially devoiced (PDEV) and devoiced (DEV) fricatives. Corpus 3 (Speaker LMTJ).

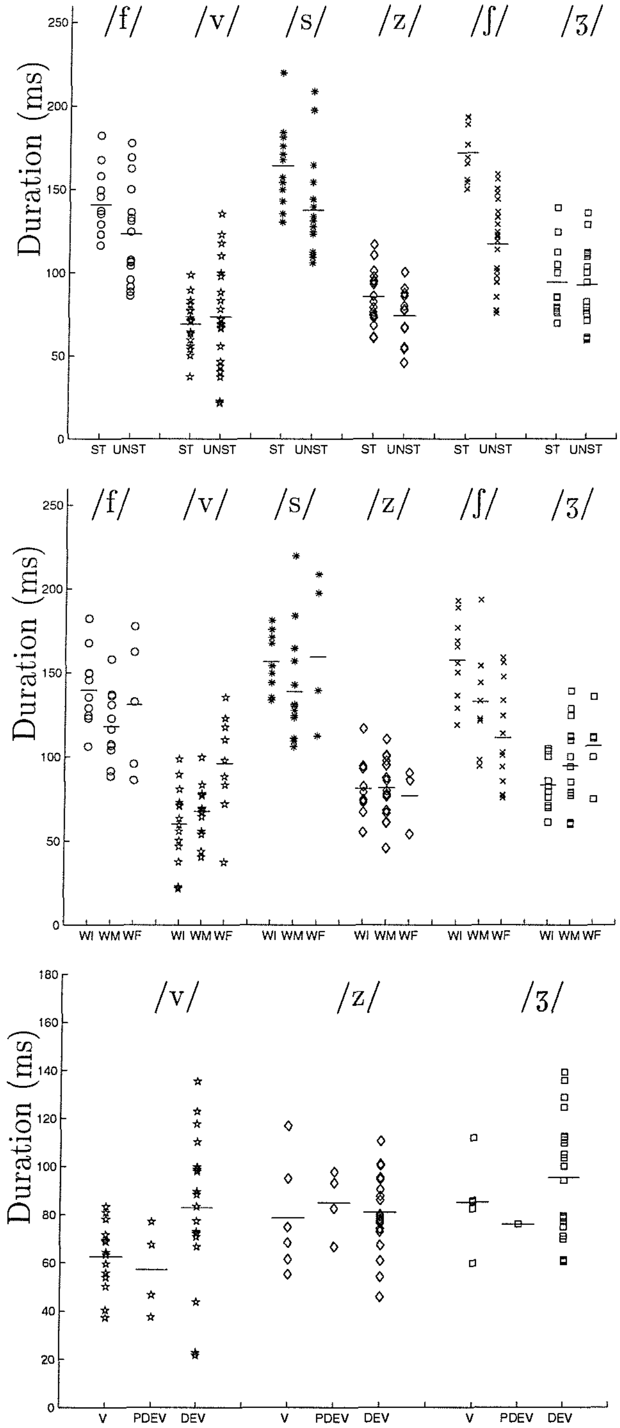

Figure 6.35: Duration for: fricatives in stressed (ST) and unstressed (UNST) syllables; word-initial (WI), word-medial (WM) and word-final (WF) fricatives; voiced (V), partially devoiced (PDEV) and devoiced (DEV) fricatives. Corpus 3 (Speaker LMTJ).

6.5 Summary

In this chapter, Portuguese fricatives were analysed in ways designed to enhance our description of the language and to increase our understanding of the production of fricatives. The parameters spectral slope, frequency of maximum amplitude, and dynamic amplitude, were developed to characterize fricative spectra, and applied to corpora. The parameters behaved as predicted for changes in effort level, voicing, and sampling time within the fricative. Some combinations were also useful for separating the fricatives by place or by sibilance.

The parameters capture source-related changes for the most part *as predicted*; for the sustained fricatives, they also separate fricatives by place. However, for the nonsense words of corpus 2, comparisons of stressed and unstressed fricatives indicate little or no change in A_d and S_p, which is *not as predicted.* Since this pattern occurs also in real words of Corpus 3, this may be a characteristic of Portuguese; the effect on a fricative of syllable stress is not the same as that of increased effort level.

A combination of parameters A_d and S_p was also useful for separating the fricatives by sibilance, and a combination of parameters S_p and S'_p separated the fricatives both by place and sibilance.

Results from the four subjects seem for the most part to be consistent. Therefore it is possible that these parameters are capturing aspects of Portuguese that differ from English or French fricatives. The quantified spectral characteristics of Portuguese fricatives can be related to specific properties of the transfer function and source spectrum during the production of these sounds, although using only the far-field acoustic signal will always present a limitation to source-filter separation.

Chapter 7

A Case Study of Bilinguality

7.1 Introduction

The main aim of this chapter was to compare the Portuguese results to previously published results for English fricatives. Corpora developed at the University of Southampton for American English did not include such a rich variety of real words as used in the Portuguese study. Therefore we designed a new British English corpus, which included some of the sentences used in a EU study by Shadle et al. (Shadle 1992; Shadle and Carter 1993), and collected, in separate recording sessions, both the Portuguese and English data, as produced by a male bilingual speaker, PS, and a female bilingual speaker, RS. It was then possible to compare the various acoustic characteristics previously examined for the fricatives of four European Portuguese speakers, with a similar set of English fricatives. We also wanted to eliminate one of the main production variation factors: the across-speaker differences.

According to Watson (1991, p. 27), when a child is learning the phonology and phonetics of a single language, he or she must:

1. learn to recognise distinct, but non-invariant acoustic patterns;
2. deduce the set of oppositions which constitute the phonological structure of the language;
3. associate the acoustic patterns with the phonological system,

despite the non-invariance of the former;

4. master the correct articulatory routines to produce acoustic patterns which satisfy other native speakers as being adequate realisations of different phonemes.

Although some bilinguals seem to attain monolingual-like speech production in both languages, it is very likely that bilinguals choose different strategies from monolinguals, which "reduce the difficulties created by their need to use two systems, without thereby sounding in any way abnormal in either" (Watson 1991, p. 37). Therefore we must consider the particulars of bilingual speech when interpreting our cross-language results, and begin our study by establishing the type of bilinguality exhibited by our subjects.

7.1.1 Type of Bilinguality

We used one of the measures of bilinguality proposed by Hamers and Blanc (2000, p. 40), which involved the collection of language biographies, self-evaluation, and judgements of bilingual production by monolingual speakers of Portuguese and of English. There are several ways of classifying bilinguals in terms of their fluency and language dominance (Hughes 1989; Bachman and Palmer 1996), so we used a previously tested procedure. Subjects filled in the questionnaire shown in Appendix F, which was originally designed by Hazan and Boulakia (1993).

The subjects used in this study were two adult bilingual siblings, with no reported history of hearing or speech disorders. Subject PS was a 22-year-old male and Speaker RS was an 18-year-old female. The siblings' mother is a European Portuguese speaker and the father a British English speaker; they reside in Cascais, Portugal. They have interacted with their parents since infancy in their mother tongues: in Portuguese with their mother and in English with their father. The age and context of acquisition of both languages, their past and present use, and the degree of literacy was found, after questioning, to be as follows:

Speaker PS – age 22.

1. Age and context of acquisition of Portuguese: acquired in the home and with friends from infancy; received some education in Portuguese.

2. Past and present use of Portuguese: used frequently at home and with friends.
3. Age and context of acquisition of English: acquired in the home from infancy and with friends from six years of age; received education mostly in English (English School in Portugal, Secondary Boarding School in England, and University in England).
4. Past and present use of English: used frequently at home, with friends and at the University.
5. Degree of literacy: undergraduate degree.

Speaker RS – age 18.

1. Age and context of acquisition of Portuguese: acquired in the home and with friends from infancy; received some education in Portuguese.
2. Past and present use of Portuguese: used frequently at home and with friends.
3. Age and context of acquisition of English: acquired in the home from infancy and with friends from six years of age; received education mostly in English (English School in Portugal, Secondary Boarding School in England, and University in England).
4. Past and present use of English: used frequently at home, with friends and at the University.
5. Degree of literacy: secondary school.

The level of bilingual competence was evaluated informally by two speech researchers, where the naturalness of the recorded Portuguese and English sentence corpora was judged to be close to "native-like" for both speakers and for both languages. Considering all of this information it is most probable that our subjects have developed a balanced and compound bilinguality (Hamers and Blanc 2000, pp. 25-30, 40, 129, 368, 369).

7.1.2 Previous Studies of Bilinguality

The subject of acoustic phonetics is such a complex area of research where a multitude of analysis and modelling methods is used, that it has always been difficult to find a conceptual framework to investigate bilinguality. Therefore, studies of bilingual speech have been mainly focused on categorical perception of plosives. Spanish and English bilinguals and monolinguals were analysed by Abramson and Lisker (1973), Williams (1977), Bond et al. (1980), and Konefal and Fokes (1981). Voice onset times (VOTs) and voicing perception of Spanish and English were different. The perception and production of plosives were also studied for French - English bilinguals, monolingual French speakers and monolingual English speakers by Caramazza et al. (1973). Results showed that French and English monolingual speakers have different VOTs, and that bilinguals use an "intermediate" voicing contrast. Watson (1990) also studied the acquisition of plosive voicing contrast of French and English monolinguals and bilinguals. Two cues of voicing were observed, with only marginal differences between monolinguals and bilinguals: overall duration of voicing of French VCVs and length of English vowels. Hazan and Boulakia (1993) also showed that French - English bilinguals did not always produce monolingual - like VOTs.

7.1.3 Previous Cross - Language Studies of Fricatives

In large cross - language studies of fricatives, different sets of acoustic parameters have been used. Ladefoged and Maddieson's (1996) book describes in great detail all fricatives in the IPA chart with examples from world languages, provides a good literature review and presents ideas to improve the acoustical description of fricatives. Maddieson (1984) examined the fricative inventories of many world languages, and described their frequency of occurrence and the structure of systems of fricatives.

Nartey (1982) studied the fricatives of twelve different languages, reporting the spectral peaks previously shown in Tables 5.1 and 5.2 for /f, v, θ, ð, s, z, ʃ, ʒ/ in three different nonsense - word vowel contexts /i, a, u/. The peaks reported for each fricative vary over a range which is common to most languages and the effect of lip rounding (the frequencies of all peaks are lowered) can be observed for all languages. The fricative repertoire of each language is quite different, which might account for some of the spectral differences.

Shinn's (1985) analysis of Mandarin, Czech and German CV syllables, produced by three native speakers, consisted of an investigation of voice onset time (VOT), noise duration (ND), the time interval from noise onset at the beginning of the syllable to the sample with the highest amplitude (RT), fall time (FT=VOT - RT), the energy (in dB) at noise onset divided by the energy of the background noise (EN), and the energy at the consonant onset divided by the average energy in the interval 20 to 70 ms after consonant onset (AT). Results were compared with a study by Howell and Rosen (1983), who used a similar methodology. Average results of the study by Shinn (1985) are shown in Table 7.1 for fricatives in /a/ and /u/ vowel contexts (Mandarin /s, ʃ/, Czech /s, z, ʃ, ʒ/ and German /f, v, z/).

Table 7.1: Results from the study of Shinn (1985).

	Mandarin	Czech	German
RT (ms)	101	116	90
EN	1.2	1.5	1.2
AT	1.7	1.9	1.6
ND (ms) – /f/	-	-	130
ND (ms) – /s/	-	221	-
ND (ms) – /z/	-	154	150
ND (ms) – /ʃ/	-	201	-
ND (ms) – /ʒ/	-	162	-

Parameters were averaged for the fricatives of each language, and the author presented a detailed discussion of which parameters were the most useful for place classification and language distinction. RT was very stable across languages, and VOT quite variable. Average NDs were not significantly different across languages, and the values of EN were very similar for the three languages. There was no significant difference between the Mandarin, Czech and German AT values. There was little information about the different production strategies used by Mandarin, Czech and German speakers.

Flege et al. (1988) studied the linguapalatal contact patterns for /s/ and /t/ produced normally and with a bite block used to fix the jaw. The subjects were two American English speakers and three Saudi Arabia Qassimi Arabic speakers. Electropalatography analysis of normal speech showed that the

Arabic subjects had more narrow and more anterior /s/ grooves than English subjects. English subjects compensated for the bite block more completely than Arabic subjects.

Cross-language studies of Japanese and Russian by Funatsu (1995), and of English, Bengali and Dutch by Evers et al. (1998), have been previously discussed in Section 6.2.

7.2 Corpora Design and Recording

The Portuguese corpora had a very similar design to the corpora described in Chapter 2. The English corpora was designed to provide valid data for cross-language comparisons with the Portuguese corpora. It also included sustained fricatives (Corpus 1a and 1b), a set of nonsense words (Corpus 2), words (Corpus 3) and sentences (Corpus 4), as listed in Appendix G. Previously used English corpora (Shadle 1992; Shadle and Carter 1993) were augmented to match the Portuguese corpora.

Each speaker was recorded in two separate sessions (Portuguese and English sessions), where the subjects counted and talked in the language of the current session, and the order of corpora recording was one of decreasing naturalness: we started by recording the sentence corpus (Corpus 4), followed by the real word corpus (Corpus 3), nonsense word corpus (Corpus 2), and finally the sustained fricative corpora (Corpora 1a and 1b). Technical aspects of the recording method were the same as described in Section 2.3.

7.3 Results

The segmentation techniques, temporal and spectral analysis methods, and parameterisation used in this Portuguese and English cross-language study, were the same as in the study of Portuguese. The value of $\overline{F}$ used for English dental fricatives was the same as used for Portuguese labiodental fricatives, that is, $\overline{F}_{/\text{f, v, }\theta\text{, ð}/} = 5\,\text{kHz}$. This resulted from an analysis of Corpus 1a and Corpus 1b, that consisted of a comparison of the spectra of fricatives /θ, ð/ with those of fricatives /f, v/. The overall amplitude and spectral peaks of /θ, ð/ did not differ substantially from /f, v/, so an $\overline{F} = 5\,\text{kHz}$ was considered adequate for both places of articulation.

7.3.1 Duration

The minimum, maximum and median [1] durations of Portuguese and English fricatives from Corpus 3 are shown in Table 7.2, and the median durations are graphed in Figure 7.1. The median duration of the unvoiced fricatives is always greater than the median duration of the voiced fricatives, which agrees with results for the English language (Hogan and Rozsypal 1980; Crystal and House 1988; Stevens et al. 1992; Pirello et al. 1997). However, there is no significant difference by place of articulation or between Portuguese and English in the results presented in Figure 7.1.

[1]Median - the 50th percentile of a sample. The median is a robust estimate of the centre of a sample of data, since outliers have little effect on it.

Table 7.2: Duration of fricatives in Corpus 3.

Speaker PS – Portuguese			
	Minimum (ms)	Median (ms)	Maximum (ms)
/f/	65	122	155
/v/	39	79	140
/s/	91	124	168
/z/	63	80	122
/ʃ/	77	123	166
/ʒ/	48	88	145
Speaker PS – English			
	Minimum (ms)	Median (ms)	Maximum (ms)
/f/	60	122	190
/v/	47	77	142
/θ/	89	99	153
/ð/	12	49	81
/s/	50	122	239
/z/	57	94	152
/ʃ/	94	120	204
/ʒ/	65	85	111
Speaker RS – Portuguese			
	Minimum (ms)	Median (ms)	Maximum (ms)
/f/	105	148	215
/v/	46	72	152
/s/	129	168	271
/z/	54	84	114
/ʃ/	118	153	228
/ʒ/	54	72	189
Speaker RS – English			
	Minimum (ms)	Median (ms)	Maximum (ms)
/f/	114	151	241
/v/	38	92	179
/θ/	95	138	192
/ð/	23	64	124
/s/	105	152	253
/z/	64	99	165
/ʃ/	120	173	256
/ʒ/	62	87	203

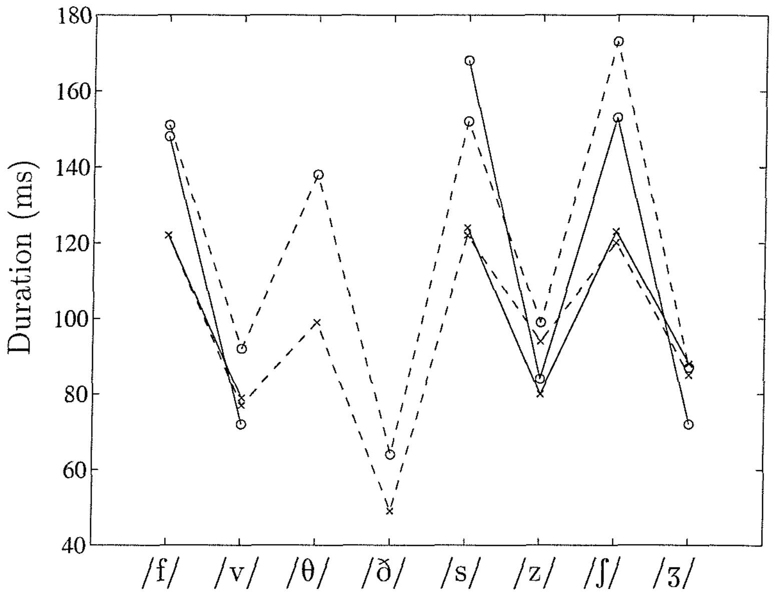

Figure 7.1: Median duration of fricatives /f, v, θ, ð, s, z, ʃ, ʒ/ in Corpus 3. Portuguese – solid line; English – dashed line; × – Speaker PS; o – Speaker RS. Number of tokens per point: /f/ (PS, P./E.) - 24/19; /f/ (RS, P./E.) - 24/19; /v/ (PS, P./E.) - 30/24; /v/ (RS, P./E.) - 29/24; /θ/ (PS) - 9; /θ/ (RS) - 9; /ð/ (PS) - 9; /ð/ (RS) - 9; /s/ (PS, P./E.) - 27/33; /s/ (RS, P./E.) - 27/33; /z/ (PS, P./E.) - 25/17; /z/ (RS, P./E.) - 25/16; /ʃ/ (PS, P./E.) - 32/20; /ʃ/ (RS, P./E.) - 32/20; /ʒ/ (PS, P./E.) - 25/11; /ʒ/ (RS, P./E.) - 25/11.

7.3.2 Devoicing

A complete inventory of devoiced, partially devoiced and voiced examples in Corpus 3 is presented for both speakers in Tables 7.3 to 7.6. For both Portuguese and English, most word-final fricative examples (48 out of 50) were totally devoiced. Overall results from the analysis of devoicing show that more than 50% of the fricatives devoice, except for /ð/ produced by PS, as shown in Figure 7.2. The percentage of devoicing is plotted by position in words in Figure 7.3. There is no significant difference between Portuguese and English in the results presented in Figures 7.2 and 7.3.

Table 7.3: Inventory of all cases of devoicing. Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3, Speaker PS – Portuguese.

Fricative	Word-Initial	Word-Medial	Word-Final	All Pos.	Voicing
/v/	10/11 (90.9%)	9/12 (75%)	7/7 (100%)	26/30 (86.7%)	Devoiced
	0	1/12 (8.3%)	0	1/30 (3.3%)	Partially Devoiced
	1/11 (9.1%)	2/12 (16.7%)	0	3/30 (10%)	Voiced
/z/	9/10 (90%)	10/12 (83.3%)	3/3 (100%)	22/25 (88%)	Devoiced
	0	2/12 (16.7%)	0	2/25 (8%)	Partially Devoiced
	1/10 (10%)	0	0	1/25 (4%)	Voiced
/ʒ/	9/10 (90%)	10/12 (83.3%)	3/3 (100%)	22/25 (88%)	Devoiced
	1/10 (10%)	1/12 (8.3%)	0	2/25 (8%)	Partially Devoiced
	0	1/12 (8.3%)	0	1/25 (4%)	Voiced
All Fric.	28/31 (90.3%)	29/36 (80.6%)	13/13 (100%)	70/80 (87.5%)	Devoiced
	1/31 (3.2%)	4/36 (11.1%)	0	5/80 (6.3%)	Partially Devoiced
	2/31 (6.5%)	3/36 (8.3%)	0	5/80 (6.3%)	Voiced

Table 7.4: Inventory of all cases of devoicing. Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3, Speaker PS – English.

Fricative	Word-Initial	Word-Medial	Word-Final	All Pos.	Voicing
/v/	7/7 (100%)	5/10 (50%)	6/7 (85.7%)	18/24 (75%)	Devoiced
	0	2/10 (20%)	0	2/24 (8.3%)	Partially Devoiced
	0	3/10 (30%)	1/7 (14.3%)	4/24 (16.7%)	Voiced
/ð/	1/1 (100%)	2/7 (28.6%)	0	3/9 (33.3%)	Devoiced
	0	1/7 (14.3%)	1/1 (100%)	2/9 (22.2%)	Partially Devoiced
	0	4/7 (57.1%)	0	4/9 (44.4%)	Voiced
/z/	3/5 (60%)	5/5 (100%)	7/7 (100%)	15/17 (88.2%)	Devoiced
	2/5 (40%)	0	0	2/17 (11.8%)	Partially Devoiced
	0	0	0	0	Voiced
/ʒ/	-	8/9 (88.9%)	2/2 (100%)	10/11 (90.9%)	Devoiced
	-	1/9 (11.1%)	0	1/11 (9.1%)	Partially Devoiced
	-	0	0	0	Voiced
All Fric.	11/13 (84.6%)	20/31 (64.5%)	15/17 (88.2%)	46/61 (75.4%)	Devoiced
	2/13 (15.4%)	4/31 (12.9%)	1/17 (5.9%)	7/61 (11.5%)	Partially Devoiced
	0	7/31 (22.6%)	1/17 (5.9%)	8/61 (13.1%)	Voiced

Table 7.5: Inventory of all cases of devoicing. Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3, Speaker RS – Portuguese.

Fricative	Word-Initial	Word-Medial	Word-Final	All Pos.	Voicing
/v/	8/11 (72.7%)	10/16 (62.5%)	3/3 (100%)	21/30 (70%)	Devoiced
	3/11 (27.3%)	3/16 (18.8%)	0	6/30 (20%)	Partially Devoiced
	0	3/16 (18.8%)	0	3/30 (10%)	Voiced
/z/	5/10 (50%)	8/15 (53.3%)	-	13/25 (52%)	Devoiced
	5/10 (50%)	6/15 (40%)	-	11/25 (44%)	Partially Devoiced
	0	1/15 (6.7%)	-	1/25 (4%)	Voiced
/ʒ/	7/10 (70%)	7/14 (50%)	1/1 (100%)	15/25 (60%)	Devoiced
	2/10 (20%)	3/14 (21.4%)	0	5/25 (20%)	Partially Devoiced
	1/10 (10%)	4/14 (28.6%)	0	5/25 (20%)	Voiced
All Fric.	20/31 (64.5%)	25/45 (55.6%)	4/4 (100%)	49/80 (61.3%)	Devoiced
	10/31 (32.3%)	12/45 (26.7%)	0	22/80 (27.5%)	Partially Devoiced
	1/31 (3.2%)	8/45 (17.8%)	0	9/80 (11.3%)	Voiced

Table 7.6: Inventory of all cases of devoicing. Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3, Speaker RS – English.

Fricative	Word-Initial	Word-Medial	Word-Final	All Pos.	Voicing
/v/	6/7 (85.7%)	8/10 (80%)	7/7 (100%)	21/24 (87.5%)	Devoiced
	0	2/10 (20%)	0	2/24 (8.3%)	Partially Devoiced
	1/7 (14.3%)	0	0	1/24 (4.2%)	Voiced
/ð/	1/1 (100%)	6/7 (85.7%)	1/1 (100%)	8/9 (88.9%)	Devoiced
	0	0	0	0	Partially Devoiced
	0	1/7 (14.3%)	0	1/9 (11.1%)	Voiced
/z/	2/5 (40%)	4/5 (80%)	6/6 (100%)	12/16 (75%)	Devoiced
	2/5 (40%)	0	0	2/16 (12.5%)	Partially Devoiced
	1/5 (20%)	1/5 (20%)	0	2/16 (12.5%)	Voiced
/ʒ/	-	7/9 (77.8%)	2/2 (100%)	9/11 (81.8%)	Devoiced
	-	2/9 (22.2%)	0	2/11 (18.2%)	Partially Devoiced
	-	0	0	0	Voiced
All Fric.	9/13 (69.2%)	25/31 (80.7%)	16/16 (100%)	50/60 (83.3%)	Devoiced
	2/13 (15.4%)	4/31 (12.9%)	0	6/60 (10%)	Partially Devoiced
	2/13 (15.4%)	2/31 (6.5%)	0	4/60 (6.7%)	Voiced

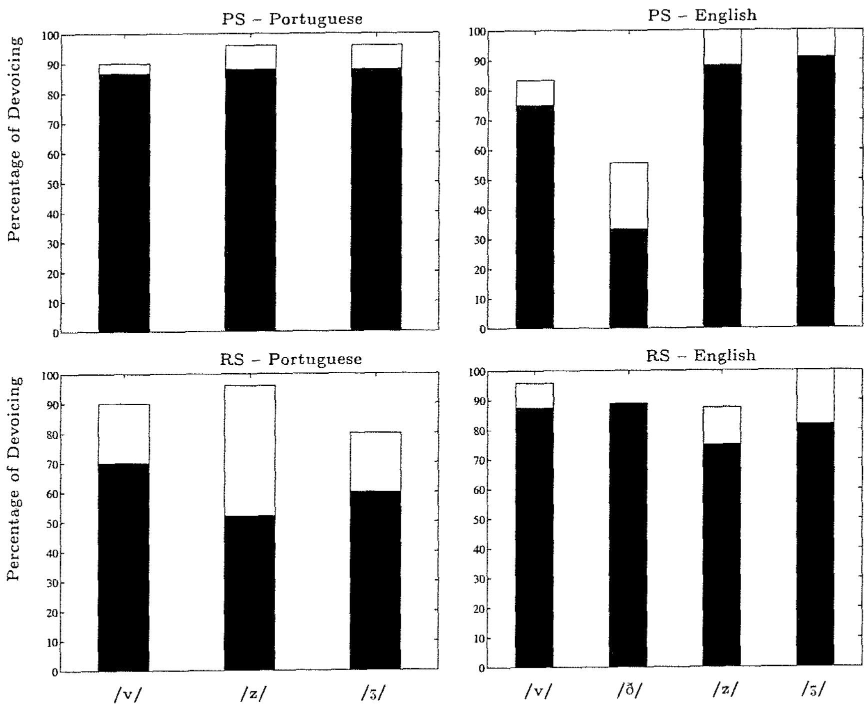

Figure 7.2: Percentage of complete and partial devoicing (Corpus 3), for each subject, each language. The black portion of each bar in the graph corresponds to the percentage of complete devoicing and the white portion to the percentage of partial devoicing.

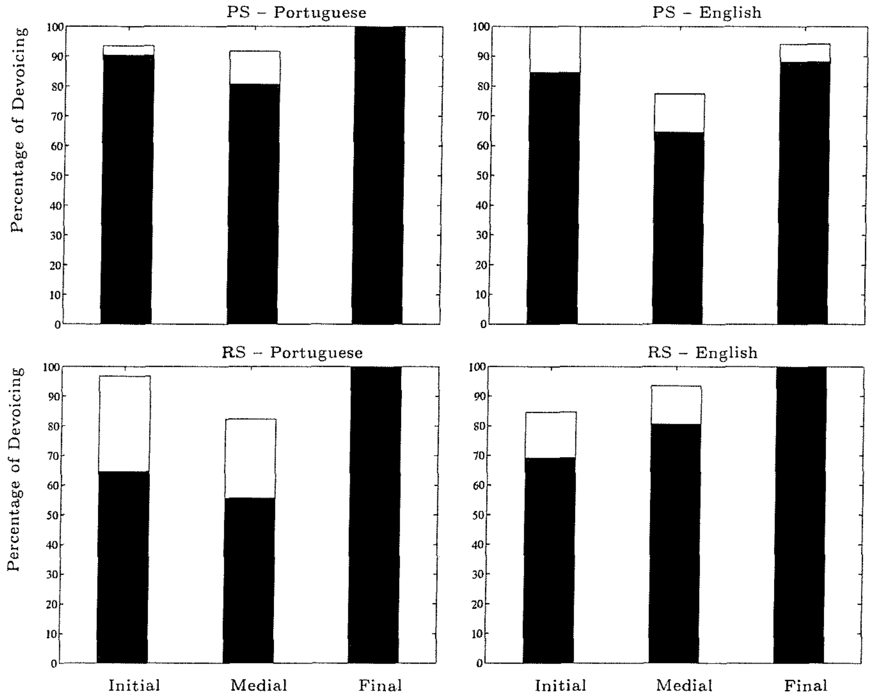

Figure 7.3: Percentage of complete and partial devoicing by position in word (Corpus 3). The black portion of each bar in the graph corresponds to the percentage of complete devoicing and the white portion to the percentage of partial devoicing.

7.3.3 Parameterisation of Spectra

We only examined Corpus 3 fricatives because these reflected language-specific characteristics. In plots of A_d and S_p by fricative, as shown in Figures 7.4 and 7.5, /s, z, ʃ, ʒ/ have higher A_d and lower S_p than /f, v, θ, ð/ for Speaker RS. The values of A_d and S_p for /θ, ð/ produced by Speaker PS seem to fall in between the values for /f, v/ and /s, z, ʃ, ʒ/. On A_d vs. S_p and S'_p vs. S_p plots, there are separate clusters of sibilants and /f, v, θ, ð/ for Speaker RS. However, the cluster for /θ, ð/ produced by Speaker PS seems to fall in between the /f, v/ cluster and the /s, z, ʃ, ʒ/ cluster. Results from both subjects seem, apart from this, to be consistent, and the same for Portuguese and English fricatives.

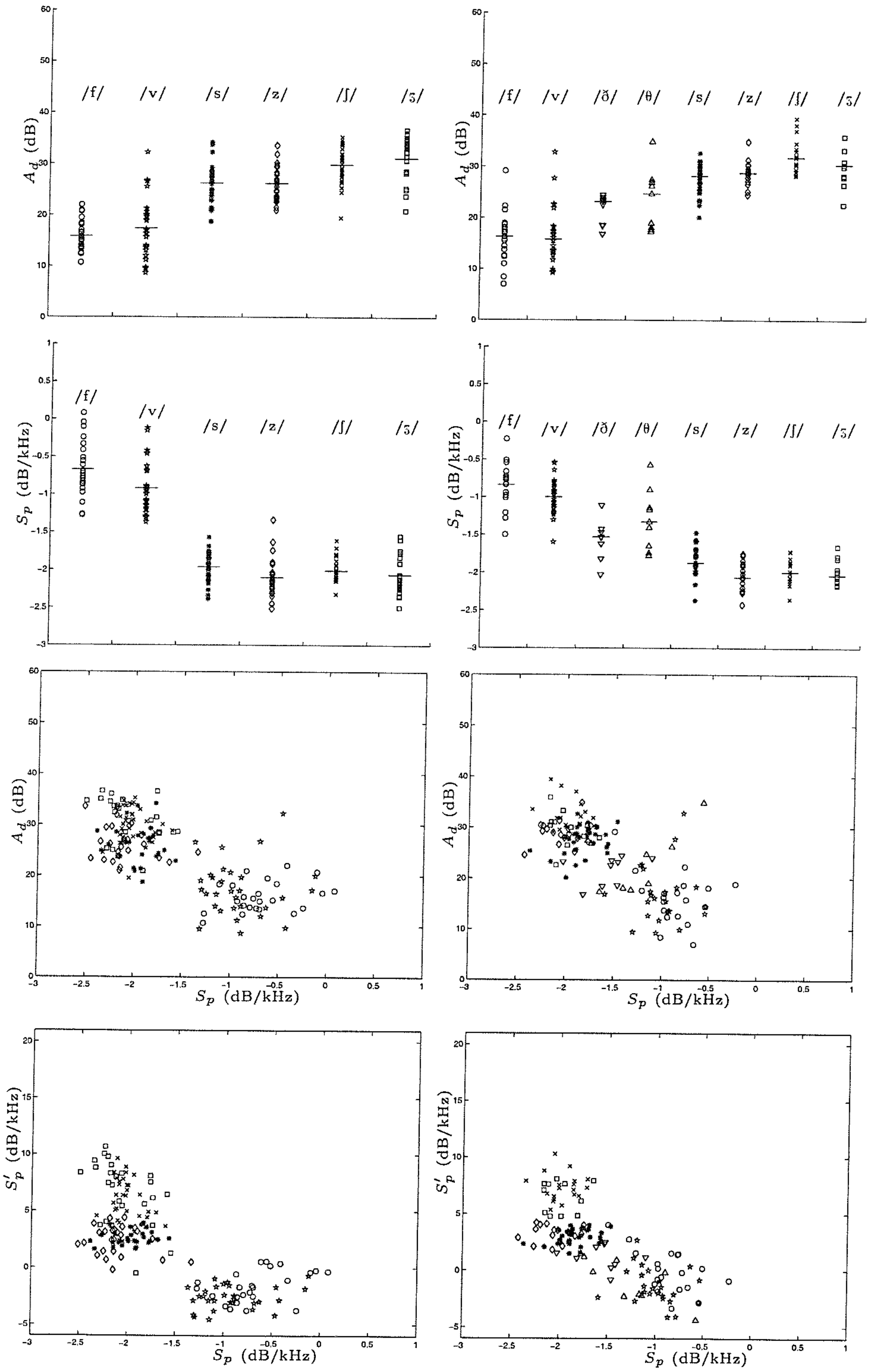

Figure 7.4: Corpus 3 (Speaker PS): Portuguese fricatives – left column; English fricatives – right column. $\circ$ – /f/, $\star$ – /v/, $\bigtriangledown$ – /θ/, $\triangle$ – /ð/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

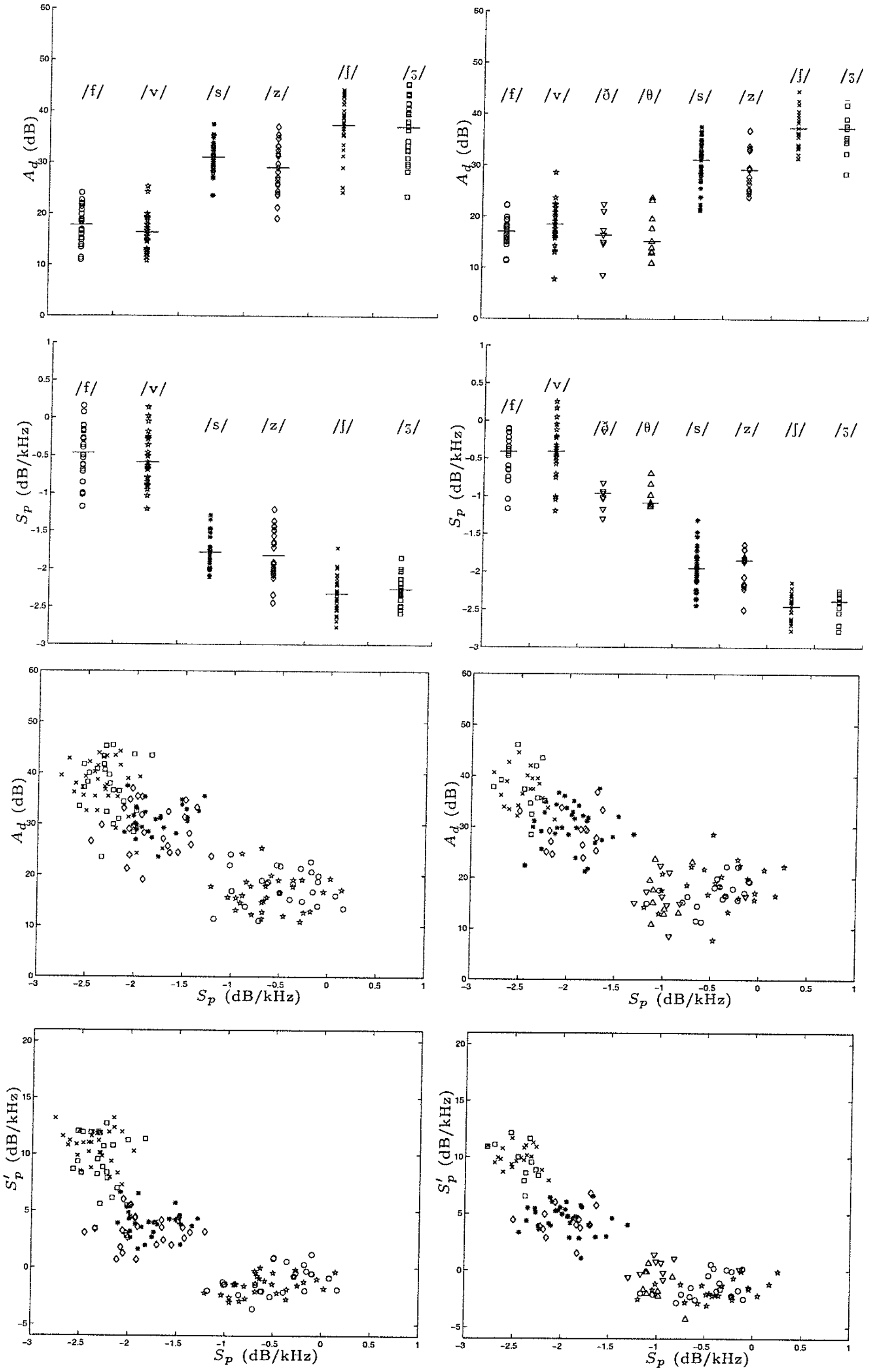

Figure 7.5: Corpus 3 (Speaker RS): Portuguese fricatives – left column; English fricatives – right column. $\circ$ – /f/, $\star$ – /v/, $\bigtriangledown$ – /θ/, $\triangle$ – /ð/, $*$ – /s/, $\diamond$ – /z/, $\times$ – /ʃ/ and $\square$ – /ʒ/.

7.4 Summary

In this study, we designed fricative corpora ranging from sustained fricatives to real Portuguese and English words, and recorded and analysed two bilingual speakers. Our goal was to test some conclusions from our study of four monolingual Portuguese subjects, and to compare our results to those of previous studies of English fricatives.

Our principal findings are as follows. Devoicing occurs more often in word-final than word-initial position, both for Portuguese and English fricatives. The percentage of totally devoiced Portuguese examples produced by the four monolingual subjects was higher than for English examples produced by the two bilingual subjects, but Portuguese and English bilingual results were very similar.

The parameters spectral slope, frequency of maximum amplitude, and dynamic amplitude, were applied to the bilingual corpora. A combination of parameters A_d and S_p and of parameters S_p and S'_p was useful for separating the fricatives by sibilance. Results for Portuguese and English fricatives seem to be very similar. The parameters A_d, S_p and S'_p are either capturing aspects of Portuguese that do not differ from English, or the subjects produce Portuguese and English fricatives the same way.

The parameters in this cross-language study might not be capturing subtle differences, and the time and frequency characteristics analysed for Portuguese and English fricatives appear to be quite invariant. A possible explanation for these results was given by Watson (1991, p. 40-44):

> A compromise seems to be reached by the bilinguals between two needs to sound sufficiently like a native speaker to conform to two different language communities and to reduce the processing load of having to master two different phonetic repertoires.
>
> ...
>
> Bilinguals modify the same variables in the production of both their languages as monolinguals, within approximately the same limits, but within these limits the details of their use of these variables may differ.

It is possible that speakers PS and RS used different production strategies from monolinguals, without this being perceptible, but resulting in an attenuation of language acoustical contrasts. Therefore our British English corpus should be used to collect data on monolingual subjects in the future.

Chapter 8

Conclusions and Future Work

8.1 Introduction

In this thesis, the design of a corpus of European Portuguese fricative consonants, the recording of four native Portuguese (monolingual) subjects, the acoustic properties of fricatives and the parameterisation of their spectra, have been described. This provided new clues to the production mechanisms of fricatives, i.e., how the fricative sound source and filter dynamic behaviour (deduced from the temporal and spectral analysis) are affected by different effort levels, vowel contexts, stress and position in word.

Corpora were designed including Portuguese words, nonsense words following Portuguese phonology, and sustained fricatives at three different effort levels; these were recorded for four speakers. The speech corpus reflects the variety of phonetic contexts in which fricatives occur.

The results from the temporal analysis, including durations of the fricatives, and of the VF and FV transitions and a study of devoicing, were discussed. An automatic measure of devoicing was compared with a manual one, and the correlation between devoicing and duration investigated.

The broad spectral envelope was also analysed, and a description of significant peaks and troughs was presented in great detail. The parameters spectral slope and dynamic amplitude were developed to characterize fricative spectra, and applied to corpora.

The various acoustic characteristics examined for the fricatives of four European Portuguese speakers, were also compared with a similar set of English fricatives. Both Portuguese and English data, as produced by a male bilingual speaker and a female bilingual speaker, were collected in separate recording sessions. The Portuguese corpus had a very similar design to the one described in Chapter 2. The English corpus was designed to provide valid data for cross-language comparisons with the Portuguese corpus.

8.2 Summary of Results for Labiodental, Alveolar and Postalveolar Fricatives

This detailed study increases our knowledge of the acoustic phonetics of Portuguese fricatives. Previous linguistic descriptions were based on very limited temporal and spectral information on fricatives, and did not identify devoicing as an intrinsic phenomenon of the Portuguese language. The reduction of vowel /ɨ/ (Andrade 1994) and the reduction of vowel /u/ in final word position have been shown always to occur in real words, far more often than previously reported. The language-specific phonological rule presented in Mateus and Andrade (2000, p. 11), stating that Portuguese only allows postalveolar fricatives word-finally, should be revised. Mateus and Andrade (2000, p. 12) also state "that phonetically any consonant may be found in word final position", which agrees with the results of the present study and constitutes a strong argument towards the revision of specific linguistic rules.

The sustained fricative corpora were better controlled and easier to analyse than Corpus 2, 3 and 4. They proved to be an important source of information of relevant acoustic cues; their use was validated by the fact that very similar spectral results were obtained for sustained fricatives and fricatives in real words. There were naturally some differences, such as the high frequency broad peaks not always being visible in real words or the "unique" devoicing pattern of Speaker ACC in Corpus 2, but these seem to result from the fact that the "naturalness" of the speech samples increases as we progress from Corpus 1a to Corpus 4, and that different "unnatural" production strategies were used by each speaker in Corpus 1a, 1b and 2.

The mean duration of the unvoiced fricatives is always greater than the mean duration of the voiced fricatives, which agrees with previous results for the English language, and the mean duration of the fricative is greater than the mean duration of the VF and FV transitions. These temporal characteristics could be observed in both Corpus 3 and Corpus 4.

Voiced fricatives devoice in over one-half of the cases in both nonsense and real words, the only exception being Speaker ACC who voiced most of the Corpus 2 tokens. A possible explanation for such high percentages of devoicing could be that, due to the structure of the language and its vocabulary, Portuguese speakers are very seldom faced with confusions between voiced and devoiced examples. Devoicing increases from word-initial, through word-medial to word-final positions, but there seems to be no particular vowel context that causes devoicing. In Corpus 3, the percentage of devoicing seems to increase as the place of articulation moves posteriorly, but in Corpus 4 /ʒ/ doesn't devoice as much as in Corpus 3. This could be because Corpus 4 is a sentence corpus, or because the place of articulation of /ʒ/ is less posterior. The percentage of examples which were classified in the same category (devoiced, partially voiced or voiced) using the manual and automatic devoicing criterion is quite high, which shows great potential for the use of the automatic technique in future work.

Spectral analysis showed that peak and trough locations are specific to each place of articulation, and quite similar to some peak frequencies reported previously for other languages. Only round back vowel context affects some of the peak and trough locations in the spectra. Fricatives in the same vowel context, within a word and across word boundaries, have similar temporal and spectral characteristics. The high effort level of fricatives in Corpus 1b does not correspond to any of the fricatives in words. The level used to produce Portuguese fricatives seems to correspond to something between soft and medium effort levels for sustained fricatives. The spectral amplitude has a fairly similar falloff at all effort levels. The differences between the three effort levels are smallest at low frequencies, and the amount of amplitude difference at high frequencies varies with the fricatives and tends to be smaller for the voiced fricatives. The overall amplitude of voiced fricatives is either the same or lower than their unvoiced counterparts, the difference varying between 0 and 20 dB. For postalveolar fricatives, there is a 20 - 30 dB drop of amplitude from the first broad peak (2.3 - 4.7 kHz) to 20 kHz. The effect of effort level on the spectral peaks and troughs of each fricative varies among speakers, and there is no correlation between stress of the syllable containing the fricative in Corpus 2 and word position in Corpus 3.

Each place has a different "family" of nearly-parallel average regression fit 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. /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. Slope generally increases with increased effort level, though this pattern is much more consistent for unvoiced fricatives. When A_d and S_p are plotted vs. location of the analysis window within the fricative for Corpus 2, 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.

A combination of parameters A_d and S_p was useful for separating the fricatives by sibilance, and a combination of parameters F, A_d and S_p separated the fricatives both by place and sibilance. On a $\overline{F}$ vs. A_d or S'_p vs. S_p graph the fricatives cluster by place. If we use place knowledge, i.e. use F, to plot A_d vs. S_p at the beginning, middle and end, the results are inconclusive.

Preliminary comparisons of stressed and unstressed fricatives indicate little or no change in A_d and S_p, for Corpus 2 fricatives, not as predicted. The overall amplitude of stressed and unstressed fricatives is the same, so the parameters seem to capture the main spectral features. The only significant difference seems to be the amplitude of the fundamental frequency component, which is 10 - 15 dB higher for stressed than for unstressed fricatives.

There seems to be no consistent effect of rounding in the values of A_d and S_p. There is also no correlation between duration and devoicing, and the values of parameters A_d and S_p. We also studied the correlation between the values of A_d, S_p and duration, and various other contextual factors (fricatives in stressed and unstressed syllables; word - initial, word - medial and word - final fricatives; voiced, partially devoiced and devoiced fricatives), without much success.

8.3 Summary of Results for Uvular Fricatives and Voiceless Tapped Alveolar Fricatives

The uvular fricative /χ/ seems to be produced on a regular basis only by speakers CFGA and ISSS, which is probably related to their particular production strategies, and its voiced counterpart /ʁ/ is used very seldom. Although the corpora we had available were very limited, for the study of uvular fricatives, it allows us to propose /χ/ as a phone of standard European Portuguese, but more data are needed to confirm this hypothesis. Frequency location of [χ] peaks (1.2 - 1.8 kHz, 2.4 - 3 kHz and 3.4 - 4 kHz) clearly indicate a back place of articulation, with median duration of 69 ms.

The median duration of [ɾ̥] was 22 ms, very similar to the closure duration of 20 - 30 ms for alveolar taps reported by Recasens (1991) for one Catalan speaker, the author, who also mentions that, for some vowel contexts, there is an incomplete closure at the central alveolar area. This probably means that some of his data included voiceless tapped alveolar fricatives (the author does not discuss in much detail the characteristics of the acoustic signal, and mostly shows electropalatography analysis results). The overall amplitude of [ɾ̥] was quite low, which perhaps suggests a different classification of this speech sound, as an allophone of both /t/ and /d/, as previously suggested by Kent and Read (1992, p. 141 - 142) for [ɾ]. The short duration of [ɾ̥] suggests a stop - like manner of articulation, but it has fricative turbulence noise characteristics, different from the transient burst noise of plosives. [ɾ̥] seems to be quite common in European Portuguese, and definitely should be considered in future fricative and plosive studies of this language.

8.4 Conclusions

The main contributions of the work described in this thesis were as follows. A novel methodology of corpus design, a systematic and coherent temporal analysis, including quantitative measures of devoicing, and spectral analysis (time, frequency and ensemble averaging of power spectra). The methodology is independent of the language, and so other researchers could use it. A set of relevant acoustic properties of fricatives and the parameterisation of their spectra, could be useful for future work on Portuguese phonetics and the synthesis of Portuguese fricatives.

Much data has been collected reflecting the variety of phonetic contexts in which fricatives occur, and there is a rich description of various dynamic behaviours (e.g. VF and FV transition durations), which could be used for improved speech synthesis. Durational data of both the fricatives and transitions, information about devoicing and the spectral parameters, should be of interest to other researchers. The peak frequencies, spectral amplitude characteristics, and temporal information could be useful for formant synthesis (Klatt 1980; Klatt and Klatt 1990; Holmes 1983; Holmes et al. 1990) and the parameterisation of the spectra allows researchers to deduce the behaviour of sources for articulatory synthesis models such as the one proposed by Narayanan and Alwan (2000).

Devoicing rate is generally very high, especially when compared with studies of other languages, and devoicing occurs more often in word-final than word-initial position. It is thought that these are important characteristic of European Portuguese, but comparisons with results from other languages is very limited because of different methodologies used in different studies. Therefore, the results presented in this thesis might be challenged by future work of other researchers, and cross-languages comparisons constitute a whole new area still to be explored further.

8.5 Further Work

A refined set of distinctive features for the Portuguese fricatives, that will be useful for synthesizing more natural fricatives and for comparing the characteristics of fricative consonants in Portuguese with other languages, could be produced from this study. In a preliminary experiment involving two bilingual siblings (one male and one female speakers of Portuguese and English) the two subjects were asked to read, in separate recording sessions, the Portuguese corpora, and English corpora very similar to that used in various studies by Shadle et al. (Shadle 1992; Shadle and Carter 1993). These corpora should be further analysed and extended to include monolingual English speakers, where the focus should be on finding specific differences between fricatives of Portuguese and English (devoicing, place and intensity), independent of inter-speaker variations.

The large annotated acoustic database of Portuguese should be extended to more speakers, therefore making it more representative of European Portuguese characteristics, and some of the corpora could also be extended to include aerodynamic measurements (Rothenberg mask and intraoral pressure) and acquisition of articulatory data (electropalatography).

It is very important to record different speakers in the future, because besides studying specific fricative attributes (Corpus 1a, Corpus 1b and Corpus 2) we have also been investigating a large number of words (Corpus 3 and Corpus 4) where the fricatives occur in a natural context. Some of the conclusions are language-specific and so to validate the results we need a larger number of speakers.

Electropalatography (EPG) could also be useful in the future, to investigate, for example, if there is any place shift for the different effort levels of sustained fricatives. If there are no place changes, it makes it more likely that the only difference between the three effort levels is source intensity. However, currently available articulatory measurement techniques have their limitations, e.g., /ʒ/'s possible place change might be hard to see with EPG (partly off palate) and MRI (it is not possible to obtain complete 3D MRI measurements of real words in sentences).

It could also be interesting to conduct a parallel study of the characteristics of fricative consonants in Portuguese and English, using both the data we have recently collected for the Portuguese language and the vast set of English data available at the Department of Electronics and Computer Science, University of Southampton. Magnetic resonance imaging (MRI) data collected by Shadle et al. (1996) have been processed by Holtrup (1998), and there is now detailed information about the dimensions of the vocal tract that can be used in a future speech production model. Acoustic and MRI data analysed in recent Ph.D. theses (Mohammad 1999; Jackson 2000), could also be a useful source of information for the English language.

As far as the definition of a speech production model for fricative consonants is concerned, an innovative hybrid of acoustic, aerodynamic and articulatory models, currently used for speech synthesis, seems to be the most promising way forward. A new fricative production model could be incorporated in an articulatory synthesizer based on models developed by Scully (1990), Davies et al. (1993), and Narayanan and Alwan (2000).

Appendix A

Listings of Corpora 3 and 4

A.1 Corpus 3: Real Words

Portuguese words with fricatives /f, v, s, z, ʃ, ʒ/ in initial and medial position (nearly minimal pairs):

fofa – /ˈfofɐ/ – English definition: soft.

viver – /viˈveɾ/ – to live.

viva – /ˈvivɐ/ – hurrah!, live (imperative), bless you.

cessa – /ˈsɛsɐ/ – ceases (verb form).

Zézé – /ˈzɛzɛ/ – (diminutive of Joseph = Joe).

chocha – /ˈʃoʃɐ/ – spineless, insipid, empty.

bochecha – /buˈʃeʃɐ/ – cheek.

Gigi – /ʒiˈʒi/ – (diminutive of a woman's first name).

Portuguese words with fricatives in initial position:

figo – /ˈfigu/ – fig.

ferir – /fɨ'ɾiɾ/ – to hurt, to injure.

febra – /'febɾɐ/ – a joint of pork.

ferro – /'fɛʀu/ – iron.

falir – /fɐ'liɾ/ – to go bankrupt.

fala – /'falɐ/ – speech.

foco – /'fɔku/ – torch.

fogo – /'fogu/ – fire.

furar – /fu'ɾaɾ/ – to drill.

vila – /'vilɐ/ – small town.

vermelho – /vɨɾ'mɐʎu/ – red .

ver – /veɾ/ – to see.

véu – /'vɛw/ – veil.

veia – /'vɐjɐ/ – vein.

vaca – /'vakɐ/ – cow.

volta – /'vɔltɐ/ – go/come back! (imperative), turn.

voo – /vow/ – flight.

vogar – /'vugaɾ/ – to row, to float.

sítio – /'sitju/ – place.

secar – /sɨ'kaɾ/ – to dry.

sede – /'sedɨ/ – thirst.

seta – /'sɛtɐ/ – arrow.

saber – /sɐ'beɾ/ – knowledge, to know.

sala – /'salɐ/ – room, suite.

só – /sɔ/ – alone, lonely, only.

sopa – /'sopɐ/ – soup.

subir – /su'biɾ/ – to climb, to mount, to rise.

Zita – /'zitɐ/ – (woman's name, diminutive of a woman's name).

zelar – /'zɨlaɾ/ – to watch over, to pay great attention to.

zelo – /'zelu/ – zeal.

Zé – /zɛ/ – (diminutive of José = Joseph).

zarpar – /zɐɾ'paɾ/ – to escape, to run away, to lift anchor.

Zaire – /'zajɾɨ/ – (proper noun).

Zópiro – /'zɔpiɾu/ – (proper noun).

zona – /'zonɐ/ – zone.

zurrar – /zu'ʀaɾ/ – to bray.

chicote – /ʃi'kɔtɨ/ – whip.

chegar – /ʃɨ'gaɾ/ – to arrive.

cheta – /'ʃetɐ/ – "não ter cheta" → "to be penniless".

cheque – /'ʃɛkɨ/ – cheque.

chamar – /ʃɐ'maɾ/ – to call.

chá – /ʃa/ – tea.

choca – /'ʃɔkɐ/ – brooding (feminine), shocks (verb form).

choco – /'ʃoku/ – brooding (masculine), cuttle fish.

chorar – /ʃu'ɾaɾ/ – to cry.

girar – /ʒi'ɾaɾ/ – to spin.

gelado – /ʒɨ'ladu/ – ice cream.

gelo – /'ʒelu/ – ice.

germe – /'ʒɛɾmɨ/ – germ.

jaqueta – /ʒɐ'ketɐ/ – short jacket.

jacto – /'ʒatu/ – jet.

jóia – /ˈʒɔjɐ/ – jewel.

jogo – /ˈʒogu/ – game.

judeu – /ʒuˈdew/ – Jew.

Portuguese words with fricatives in medial position:

efectuar – /iˈfɛtwaɾ/ – to accomplish.

benefício – /bɨnɨˈfisju/ – benefit.

trefo – /ˈtɾefu/ – cunning, astute.

benéfico – /biˈnɛfiku/ – beneficial.

afiar – /ɐfiˈaɾ/ – to sharpen.

café – /kɐˈfɛ/ – coffee .

garrafa – /gɐˈʀafɐ/ – bottle.

bafo – /ˈbafu/ – breath.

galhofa – /gɐˈʎɔfɐ/ – amusement, frolic.

mofo – /ˈmofu/ – mould.

bufa – /ˈbufɐ/ – to blow (verb form).

altivo – /alˈtivu/ – haughty, arrogant.

dever – /dɨˈveɾ/ – duty, to owe.

levar – /lɨˈvaɾ/ – to take.

relevo – /ʀɨˈlevu/ – relief.

leva – /ˈlɛvɐ/ – takes (verb form).

avó – /ɐˈvɔ/ – grandmother.

cava – /ˈkavɐ/ – digs (verb form).

bravo – /ˈbɾavu/ – brave, wild.

nova – /ˈnɔvɐ/ – new.

ovelha – /oˈvɐʎɐ/ – sheep.

mover – /muˈveɾ/ – to move.

uva – /ˈuvɐ/ – grape.

iça – /ˈisɐ/ – lifts (verb form).

ressaca – /ʀɨˈsakɐ/ – hangover.

condessa – /kõˈdesɐ/ – countess.

pêssego – /ˈpesɨgu/ – peach.

aquecer – /ɐkɛˈseɾ/ – to heat.

passear – /pɐˈsjaɾ/ – to walk, to go for a walk.

assar – /ɐˈsaɾ/ – to roast.

caça – /ˈkasɐ/ – hunting (verb form), game.

possa – /ˈpɔsɐ/ – can (verb form).

moça – /ˈmosɐ/ – girl.

possível – /puˈsivɛl/ – possible.

exacto – /iˈzatu/ – exact.

mesinha – /mɨˈziɲɐ/ – small table.

beleza – /bɨˈlezɐ/ – beauty.

peso – /ˈpezu/ – weight.

mezinha – /mɛˈziɲɐ/ – traditional medicine.

Brasil – /bɾɐˈzil/ – (proper noun).

azar – /ɐˈzaɾ] – bad luck.

azul – /ɐˈzul/ – blue.

mazinha – /maˈziɲɐ/ – pest, bad girl/woman.

asa – /ˈazɐ/ – wing.

rosa – /ˈɾɔzɐ/ – rose.

amoroso – /ɐmu'ɾozu/ – amorous, sweet.

acusar – /ɐku'zaɾ/ – to accuse.

bicha – /'biʃɐ/ – queue.

bexiga – /bɨ'ʃigɐ/ – bladder.

este – /'eʃtɨ/ – this one.

meche – /'mɛʃɨ/ – touches (verb form).

achar – /ɐ'ʃaɾ/ – to find, to think.

bolacha – /bu'laʃɐ/ – biscuit.

tacho – /'taʃu/ – pot, pan.

tocha – /'tɔʃɐ/ – torch.

mocho – /'moʃu/ – owl.

capucho – /kɐ'puʃu/ – hood.

originar – /oɾiʒi'naɾ/ – to originate, to generate.

tijolo – /ti'ʒolu/ – brick.

arejar – /ɐɾɨ'ʒaɾ/ – to ventilate.

pejo – /'peʒu/ – modesty.

Beja – /'bɛʒɐ/ – (proper noun).

agir – /ɐ'ʒiɾ/ – to act.

cajado – /kɐ'ʒadu/ – crook.

ajudar – /ɐ'ʒudaɾ/ – to help.

haja – /'aʒɐ/ – there is (verb form).

aloja – /ɐ'lɔʒɐ/ – lodges (verb form).

tojo – /'toʒu/ – gorse.

tugir – /tu'ʒiɾ/ – to speak low.

Portuguese words with fricative /ʃ/ in final position:

diz – /diʃ/ – says (verb form), tell me (imperative).

mares – /ˈmaɾɨʃ/ – seas.

mês – /ˈmeʃ/ – month.

pés – /pɛʃ/ – feet.

perdas – /ˈpɛɾdɐʃ/ – losses.

capaz – /kɐˈpaʃ/ – capable.

pós – /pɔʃ/ – powders.

pôs – /poʃ/ – put (verb form).

dos – /ˈduʃ/ – of (the).

Portuguese words with fricatives /f, v, s, z, ʒ/ in "simulated" final position.

chefe – /ˈʃɛfɨ/ – chief.

Fafe – /ˈfafɨ/ – (proper noun).

teve – /ˈtevɨ/ – had (verb form).

leve – /ˈlɛvɨ/ – light.

ave – /ˈavɨ/ – bird.

move – /ˈmɔvɨ/ – moves (verb form).

partisse – /paɾˈtisɨ/ – left (verb form).

batesse – /bɐˈtesɨ/ – hit (verb form).

asse – /ˈasɨ/ – roast (verb form).

posse – /ˈpɔsɨ/ – possession.

doce – /ˈdosɨ/ – sweet.

doze – /ˈdozɨ/ – twelve.

age – /ˈaʒɨ/ – acts (verb form).

hoje – /ˈoʒɨ/ – today.

A.2 Corpus 4: Real Words in Connected Speech

Corpus 4 sentences, listed in this appendix, were constructed using about half of the words from Corpus 3, with sentences 1 to 10 making some sense in Portuguese. Sentences 11 and 12 were devised to reproduce some of the vocalic contexts used for Corpus 3 across word boundaries (this is signaled in the phonetic transcription by boxes), but they make no sense.

1. "A Gigi é uma chocha e age em benefício da avó doce."
   /ɐ ʒiˈʒi ɛ ˈumɐ ˈʃoʃɐ i ˈaʒɨ ɐ̃j̃ bɨniˈfisju dɐ ɐˈvɔ ˈdosɨ/
   (Gigi is spineless and acts for the benefit of the sweet grandmother.)
2. "A vaca foge do gelo na zona."
   /ɐ ˈvakɐ ˈfɔʒɨ du ˈʒelu nɐ ˈzonɐ/
   (The cow runs from the ice in the zone.)
3. "A ave, no voo a subir, move a asa para zarpar da seta."
   /ɐ ˈavɨ nu vow ɐ suˈbiɾ ˈmɔvɨ ɐ ˈazɐ ˈpɐɾɐ zɐɾˈpar dɐ ˈsɛtɐ/
   (The bird, in a rising flight, moves his wing to escape from the arrow.)
4. "O chefe altivo fala à rosa de sede de beleza."
   /u ˈʃɛfɨ alˈtivu ˈfalɐ a ˈrɔzɐ dɨ ˈsedɨ dɨ bɨˈlezɐ/
   (The haughty chief speaks to the rose about thirst and beauty.)
5. "Quero chorar hoje, arejar o sítio e vogar."
   /ˈkɛɾu ʃuˈɾaɾ ˈoʒɨ ɐɾɨˈʒaɾ u ˈsitju i ˈvugaɾ/
   (I want to cry today, air the place out and float.)
6. "O bravo do Zé quer ajudar o judeu só."
   /u ˈbɾavu du zɛ kɛɾ ɐˈʒudaɾ u ʒuˈdew sɔ/
   (Brave Zé wants to help the lonely Jew.)
7. "O café cura a ressaca ao chegar da caça."
   /u kɐˈfɛ kuɾɐ ɐ ʀiˈsakɐ aw ʃiˈgaɾ dɐ ˈkasɐ/
   (Coffee cures a hangover when you arrive from hunting.)
8. "Ver a mesinha nova de volta à sala, é benéfico para o modo de viver dos doze."
   /veɾ ɐ miˈziɲɐ ˈnɔvɐ dɨ ˈvɔltɐ a ˈsalɐ ɛ bɨˈnɛfiku ˈpɐɾɐ u ˈmɔdu dɨ viˈveɾ duʃ ˈdozɨ/
   (It is beneficial for the way of living of the twelve, to see the new table back in the room.)

9. “Furar uma jóia choca a condessa Zita.”
/fuˈɾaɾ umɐ ˈʒɔjɐ ˈʃɔkɐ ɐ kõˈdesɐ ˈzitɐ/
(To drill a jewel shocks countess Zita.)

10. “O vermelho do fogo e o azul dos mares do Brasil.”
/u viɾˈmɐʎu du ˈfogu i u ɐˈzul duʃ ˈmaɾiʃ du bɾɐˈzil/
(The red of fire and the blue of the sea of Brasil.)

11. “A chá no sítio é possível achar.”
/ɐ ʃa n u ˈsi tju ɛ p uˈsi vɛl ɐˈʃa ɾ/
(Tea in the place is possible to find.)

12. “Cava sala meche ver o dever de assar.”
/ˈkav ɐ ˈsa lɐ ˈmɛʃ ɨ veɾ u d ɨˈveɾ dɨ ɐˈsa ɾ/
(Dig room touches see the duty to roast.)

Appendix B

Listings of Corpora as Presented to All Four Speakers

The listings of corpora in the following sections, include the instructions given to subjects in italics.

B.1 Corpus 1a

“as[a]r” → /ˈaz[ɐ]/	“l[e]var” → /l[ɨ]ˈvar/	“[Ch]á” → /[ʃ]a/	“[j]óia” → /ˈ[ʒ]ɔjɐ/

Sustain fricative for 5 s.

1. /uvvvv ... u/
2. /ɐʒʒʒʒ ... ɐ/
3. /ɐvvvv ... ɐ/
4. /uʃʃʃʃ ... u/
5. /ɨvvvv ... ɨ/
6. /ɐʃʃʃʃ... ɐ/
7. /uffff ... u/
8. /ussss ... u/
9. /ɐzzzz ... ɐ/
10. /ɨʒʒʒʒ ... ɨ/
11. /ɨssss ... ɨ/
12. /uzzzz ... u/
13. /ɐffff ... ɐ/
14. /ɨʃʃʃʃ ... ɨ/
15. /ɐssss ... ɐ/
16. /ɨzzzz ... ɨ/
17. /uʒʒʒʒ ... u/
18. /ɨffff ... ɨ/

B.2 Corpus 1b

“Chá” → /ʃa/	“jóia” → /ˈʒɔjɐ/

Sustain for 5s.

Medium, soft and loud for each fricative.

Repeat corpus twice.

1. /ʃ/
   baixo
   alto
2. /s/
   baixo
   alto
3. /v/
   baixo
   alto
4. /ʒ/
   baixo
   alto
5. /f/
   baixo
   alto
6. /z/
   baixo
   alto
7. /s/
   baixo
   alto
8. /v/
   baixo
   alto
9. /f/
   baixo
   alto
10. /ʒ/
    baixo
    alto
11. /z/
    baixo
    alto
12. /ʃ/
    baixo
    alto

B.3 Corpus 2

“asa” → /ˈazɐ/	“levar” → /lɨˈvar/	“Chá” → /ʃa/	“jóia” → /ˈʒɔjɐ/

Do about 12 repetitions in one breath.

1. /pɨˈʃu/
2. /ˈpɐfɨ/
3. /puˈʒu/
4. /pɨfɨ/
5. /pɐˈvu/
6. /pɨˈzɐ/

7. /pɨʒɐ/
8. /pɨʃɨ/
9. /pɨˈzu/
10. /ˈpɐvɨ/
11. /ˈpɐʒɨ/
12. /puvu/
13. /pɐzɐ/
14. /pɨˈsɐ/
15. /ˈpuvɨ/
16. /pɨsɨ/
17. /pɐˈʃɐ/
18. /pɐˈʒu/
19. /pɨˈvu/
20. /pɐˈʃu/
21. /pɐfɐ/
22. /ˈpuʒɨ/
23. /pɐˈzu/
24. /ˈpɐʃɨ/
25. /pɨˈfu/
26. /pɐvɐ/
27. /ˈpusɨ/
28. /puˈfɐ/
29. /ˈpɐsɨ/
30. /pɐʒɐ/
31. /pɨˈʃɐ/
32. /pusu/
33. /pɐˈfu/
34. /pɨvɨ/
35. /puzu/
36. /pɨzɨ/
37. /puˈvɐ/
38. /pɨˈfɐ/
39. /puʃu/
40. /ˈpufɨ/
41. /pɨˈʒu/
42. /ˈpuzɨ/
43. /pɨʒɨ/
44. /pufu/
45. /pɨˈvɐ/
46. /puˈʒɐ/
47. /pɨˈsu/
48. /ˈpuʃɐ/
49. /pɐsɐ/
50. /ˈpuʃɨ/
51. /ˈpɐzɨ/
52. /puˈsɐ/
53. /pɐsu/
54. /puˈzɐ/

B.4 Corpus 3

The carrier sentence is repeated in the following listings, because they were based on the material that the speaker read in the recording session.

1. "Diga originar, por favor."
2. "Diga fala, por favor."
3. "Diga jogo, por favor."
4. "Diga ferro, por favor."
5. "Diga só, por favor."
6. "Diga leve, por favor."
7. "Diga vogar, por favor."
8. "Diga amoroso, por favor."

9. “Diga tojo, por favor.”
10. “Diga Zópiro, por favor.”
11. “Diga chefe, por favor.”
12. “Diga haja, por favor.”
13. “Diga sala, por favor.”
14. “Diga rosa, por favor.”
15. “Diga bolacha, por favor.”
16. “Diga zurrar, por favor.”
17. “Diga voo, por favor.”
18. “Diga ferir, por favor.”
19. “Diga asa, por favor.”
20. “Diga jaqueta, por favor.”
21. “Diga posse, por favor.”
22. “Diga dever, por favor.”
23. “Diga fofa, por favor.”
24. “Diga cava, por favor.”
25. “Diga pejo, por favor.”
26. “Diga bafo, por favor.”
27. “Diga mezinha, por favor.”
28. “Diga chamar, por favor.”
29. “Diga perdas, por favor.”
30. “Diga cheta, por favor.”
31. “Diga café, por favor.”
32. “Diga judeu, por favor.”
33. “Diga azul, por favor.”
34. “Diga seta, por favor.”
35. “Diga caça, por favor.”
36. “Diga cheque, por favor.”
37. “Diga mazinha, por favor.”
38. “Diga hoje, por favor.”
39. “Diga saber, por favor.”
40. “Diga uva, por favor.”
41. “Diga figo, por favor.”
42. “Diga teve, por favor.”
43. “Diga exacto, por favor.”
44. “Diga furar, por favor.”
45. “Diga ave, por favor.”
46. “Diga arejar, por favor.”
47. “Diga falir, por favor.”
48. “Diga girar, por favor.”
49. “Diga benefício, por favor.”
50. “Diga mesinha, por favor.”
51. “Diga chá, por favor.”
52. “Diga Beja, por favor.”
53. “Diga afiar, por favor.”
54. “Diga este, por favor.”
55. “Diga mofo, por favor.”
56. “Diga doce, por favor.”
57. “Diga mover, por favor.”

58. "Diga Brasil, por favor."
59. "Diga vila, por favor."
60. "Diga peso, por favor."
61. "Diga febra, por favor."
62. "Diga pés, por favor."
63. "Diga cajado, por favor."
64. "Diga cessa, por favor."
65. "Diga chegar, por favor."
66. "Diga ajudar, por favor."
67. "Diga move, por favor."
68. "Diga Zita, por favor."
69. "Diga batesse, por favor."
70. "Diga pós, por favor."
71. "Diga garrafa, por favor."
72. "Diga germe, por favor."
73. "Diga bufa, por favor."
74. "Diga capucho, por favor."
75. "Diga sede, por favor."
76. "Diga choco, por favor."
77. "Diga sítio, por favor."
78. "Diga chorar, por favor."
79. "Diga age, por favor."
80. "Diga chocha, por favor."
81. "Diga volta, por favor."
82. "Diga ressaca, por favor."
83. "Diga sopa, por favor."
84. "Diga zona, por favor."
85. "Diga Zé, por favor."
86. "Diga vermelho, por favor."
87. "Diga pêssego, por favor."
88. "Diga avó, por favor."
89. "Diga benéfico, por favor."
90. "Diga foco, por favor."
91. "Diga capaz, por favor."
92. "Diga vaca, por favor."
93. "Diga secar, por favor."
94. "Diga relevo, por favor."
95. "Diga possível, por favor."
96. "Diga diz, por favor."
97. "Diga azar, por favor."
98. "Diga efectuar, por favor."
99. "Diga Zaire, por favor."
100. "Diga bravo, por favor."
101. "Diga fogo, por favor."
102. "Diga ver, por favor."
103. "Diga passear, por favor."
104. "Diga véu, por favor."
105. "Diga levar, por favor."
106. "Diga bexiga, por favor."

107. "Diga trefo, por favor."
108. "Diga agir, por favor."
109. "Diga mocho, por favor."
110. "Diga tijolo, por favor."
111. "Diga veia, por favor."
112. "Diga mês, por favor."
113. "Diga zarpar, por favor."
114. "Diga achar, por favor."
115. "Diga zelo, por favor."
116. "Diga choca, por favor."
117. "Diga partisse, por favor."
118. "Diga asse, por favor."
119. "Diga gelo, por favor."
120. "Diga tugir, por favor."
121. "Diga tocha, por favor."
122. "Diga viver, por favor."
123. "Diga jacto, por favor."
124. "Diga altivo, por favor."
125. "Diga Fafe, por favor."
126. "Diga zelar, por favor."
127. "Diga gelado, por favor."
128. "Diga galhofa, por favor."
129. "Diga Zézé, por favor."
130. "Diga mares, por favor."
131. "Diga aquecer, por favor."
132. "Diga doze, por favor."
133. "Diga tacho, por favor."
134. "Diga chicote, por favor."
135. "Diga beleza, por favor."
136. "Diga ovelha, por favor."
137. "Diga assar, por favor."
138. "Diga dos, por favor."
139. "Diga subir, por favor."
140. "Diga viva, por favor."
141. "Diga meche, por favor."
142. "Diga bicha, por favor."
143. "Diga jóia, por favor."
144. "Diga pôs, por favor."
145. "Diga Gigi, por favor."
146. "Diga acusar, por favor."
147. "Diga nova, por favor."
148. "Diga iça, por favor."
149. "Diga leva, por favor."
150. "Diga possa, por favor."
151. "Diga aloja, por favor."
152. "Diga moça, por favor."
153. "Diga condessa, por favor."
154. "Diga bochecha, por favor."

B.5 Corpus 4

Read twice.

1. "A Gigi é uma chocha e age em benefício da avó doce."
2. "A vaca foge do gelo na zona."
3. "A ave, no voo a subir, move a asa para zarpar da seta."
4. "O chefe altivo fala à rosa de sede de beleza."
5. "Quero chorar hoje, arejar o sítio e vogar."
6. "O bravo do Zé quer ajudar o judeu só."
7. "O café cura a ressaca ao chegar da caça."
8. "Ver a mesinha nova de volta à sala, é benéfico para o modo de viver dos doze."
9. "Furar uma jóia choca a condessa Zita."
10. "O vermelho do fogo e o azul dos mares do Brasil."
11. "A chá no sítio é possível achar."
12. "Cava sala meche ver o dever de assar."

Appendix C

Calibration Method

A 94 dB, 1000 Hz calibration tone produced by a Bruel & Kjaer 4620 calibrator was recorded on the same tape on which speech was recorded, with the amplification varied by a known amount (see Table C.1).

Table C.1: Recordings' settings.

Date	Speaker	DAT Rec. Level	Input Gain		Output Gain	
			Speech	Tone	Speech	Tone
6/11/1998	LMTJ	6	20	10	20	10
25/1/1999	LMTJ	Not registered	20	10	20	20
22/6/1999	ACC	6	20	10	20	10
19/11/1999	CFGA	3.5	20	10	30	10
19/11/1999	ISSS	5	20	10	20	10

To obtain an absolute spectral amplitude we will start by calculating a factor A_1 which, when added to the internal arbitrary amplitude of the recorded calibration tone, makes the sum equal to the known amplitude of the calibration tone:

$$A_1 = 94.1 - 20\log(Y_{arb}(1000)) \quad \text{(dB)} \tag{C.1}$$

where $Y_{arb}(1000)$ is the arbitrary internal amplitude of the Fourier transform at 1 kHz of the calibration tone. We will also have to calculate a second A_2 that will be equal to the difference in amplification for the tone and speech:

$$A_2 = G_{cal} - G_{sp} \quad \text{(dB)} \tag{C.2}$$

where G_{cal} is the gain applied when the calibration signal was recorded, and

G_{sp} is the gain applied when the speech signal was recorded. Therefore the absolute spectral amplitude of the speech signal $X_{arb}(1000)$ is given by

$$X_{abs} = 20\log(X_{arb}(1000)) + A_1 + A_2 \quad \text{(dB)} \tag{C.3}$$

The spectra shown in this thesis do not present an absolute amplitude. We are currently working on a method that uses the calibration signal to calculate an absolute spectral amplitude that will be referred to a 1 Hz interval and will thus allow comparison regardless of window lengths and averaging techniques.

The power spectrum (energy) of the speech signal is defined as:

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \tag{C.4}$$

If we increase the number of points in $x(t)$ (i.e. the size of the window) the value of the integral (area delimited by the function) also increases. Therefore, the window length used to calculate the power spectra affects the overall amplitude. All else being equal, the larger the size of the window the higher is the overall amplitude.

We used the same window size to calculate the power spectra of ambient noise, sustained fricatives, fricatives in nonsense words and real words. We used a larger number of windows to calculate the averaged power spectrum of a longer segment of signal (ambient noise and sustained fricatives). This allowed us to compare spectral amplitudes of Corpus 1a, 1b, 2, 3 and 4, for a given recording session.

Appendix D

Results of Devoicing Analysis

Table D.1: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3 (Speaker LMTJ).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	6/14 (42.9%)	4/14 (28.6%)	8/9 (88.9%)	18/37 (48.7%)	Devoiced
	2/14 (14.3%)	2/14 (14.3%)	0	4/37 (10.8%)	Partially Devoiced
	6/14 (42.9%)	8/14 (57.1%)	1/9 (11.1%)	15/37 (40.5%)	Voiced
/z/	5/10 (50%)	12/17 (70.6%)	3/3 (100%)	20/30 (66.7%)	Devoiced
	2/10 (20%)	2/17 (11.8%)	0	4/30 (13.3%)	Partially Devoiced
	3/10 (30%)	3/17 (17.7%)	0	6/30 (20%)	Voiced
/ʒ/	7/10 (70%)	13/15 (86.7%)	4/5 (80%)	24/30 (80%)	Devoiced
	1/10 (10%)	0	0	1/30 (3.3%)	Partially Devoiced
	2/10 (20%)	2/15 (13.3%)	1/5 (20%)	5/30 (16.7%)	Voiced
All Fric.	18/34 (52.9%)	29/46 (63%)	15/17 (88.2%)	62/97 (63.9%)	Devoiced
	5/34 (14.7%)	4/46 (8.7%)	0	9/97 (9.3%)	Partially Devoiced
	11/34 (32.4%)	13/46 (28.3%)	2/17 (11.8%)	26/97 (26.8%)	Voiced

Table D.2: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3 (Speaker CFGA).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	7/11 (63.6%)	8/12 (66.7%)	7/7 (100%)	22/30 (73.3%)	Devoiced
	0	2/12 (16.7%)	0	2/30 (6.7%)	Partially Devoiced
	4/11 (36.4%)	2/12 (16.7%)	0	6/30 (20%)	Voiced
/z/	8/10 (80%)	12/14 (85.7%)	2/3 (66.7%)	22/27 (81.5%)	Devoiced
	1/10 (10%)	2/14 (14.3%)	0	3/27 (11.1%)	Partially Devoiced
	1/10 (10%)	0	1/3 (33.3%)	2/27 (7.4%)	Voiced
/ʒ/	8/10 (80%)	12/13 (92.3%)	4/4 (100%)	24/27 (88.9%)	Devoiced
	1/10 (10%)	0	0	1/27 (3.7%)	Partially Devoiced
	1/10 (10%)	1/13 (7.7%)	0	2/27 (7.4%)	Voiced
All Fric.	23/31 (74.2%)	32/39 (82.1%)	13/14 (92.9%)	68/84 (81%)	Devoiced
	2/31 (6.5%)	4/39 (10.3%)	0	6/84 (7.1%)	Partially Devoiced
	6/31 (19.4%)	3/39 (7.7%)	1/14 (7.1%)	10/84 (11.9%)	Voiced

Table D.3: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3 (Speaker ACC).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	4/11 (36.4%)	6/12 (50%)	7/7 (100%)	17/30 (56.7%)	Devoiced
	2/11 (18.2%)	2/12 (16.7%)	0	4/30 (13.3%)	Partially Devoiced
	5/11 (45.5%)	4/12 (33.3%)	0	9/30 (30%)	Voiced
/z/	4/10 (40%)	8/12 (66.7%)	3/3 (100%)	15/25 (60%)	Devoiced
	4/10 (40%)	3/12 (25%)	0	7/25 (28%)	Partially Devoiced
	2/10 (20%)	1/12 (8.3%)	0	3/25 (12%)	Voiced
/ʒ/	9/10 (90%)	11/11 (100%)	3/4 (75%)	23/25 (92%)	Devoiced
	1/10 (10%)	0	1/4 (25%)	2/25 (8%)	Partially Devoiced
	0	0	0	0	Voiced
All Fric.	17/31 (54.8%)	25/35 (71.4%)	13/14 (92.9%)	55/80 (68.8%)	Devoiced
	7/31 (22.6%)	5/35 (14.3%)	1/14 (7.1%)	13/80 (16.3%)	Partially Devoiced
	7/31 (22.6%)	5/35 (14.3%)	0	12/80 (15%)	Voiced

Table D.4: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 3 (Speaker ISSS).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	1/11 (9.1%)	5/12 (41.7%)	7/7 (100%)	13/30 (43.3%)	Devoiced
	4/11 (36.4%)	0	0	4/30 (13.3%)	Partially Devoiced
	6/11 (54.6%)	7/12 (58.3%)	0	13/30 (43.3%)	Voiced
/z/	8/10 (80%)	11/12 (91.7%)	3/3 (100%)	22/25 (88%)	Devoiced
	0	0	0	0	Partially Devoiced
	2/10 (20%)	1/12 (8.3%)	0	3/25 (12%)	Voiced
/ʒ/	8/10 (80%)	9/11 (81.8%)	4/4 (100%)	21/25 (84%)	Devoiced
	2/10 (20%)	1/11 (9.1%)	0	3/25 (12%)	Partially Devoiced
	0	1/11 (9.1%)	0	1/25 (4%)	Voiced
All Fric.	17/31 (54.8%)	25/35 (71.4%)	14/14 (100%)	56/80 (70%)	Devoiced
	6/31 (19.4%)	1/35 (2.9%)	0	7/80 (8.8%)	Partially Devoiced
	8/31 (25.8%)	9/35 (25.7%)	0	17/80 (21.3%)	Voiced

Table D.5: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 4 (Speaker LMTJ).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	5/14 (35.7%)	7/18 (38.9%)	1/2 (50%)	13/34 (38.2%)	Devoiced
	1/14 (7.1%)	3/18 (16.7%)	0	4/34 (11.8%)	Partially Devoiced
	8/14 (57.1%)	8/18 (44.4%)	1/2 (50%)	17/34 (50%)	Voiced
/z/	5/8 (62.5%)	7/10 (70%)	4/4 (100%)	16/22 (72.7%)	Devoiced
	1/8 (12.5%)	2/10 (20%)	0	3/22 (13.6%)	Partially Devoiced
	2/8 (25%)	1/10 (10%)	0	3/22 (13.6%)	Voiced
/ʒ/	7/8 (87.5%)	5/9 (55.6%)	3/5 (60.0%)	15/22 (68.2%)	Devoiced
	1/8 (12.5%)	3/9 (33.3%)	0	4/22 (18.2%)	Partially Devoiced
	0	1/9 (11.1%)	2/5 (40%)	3/22 (13.6%)	Voiced
All Fric.	17/30 (56.7%)	19/37 (51.4%)	8/11 (72.7%)	44/78 (56.4%)	Devoiced
	3/30 (10%)	8/37 (21.6%)	0	11/78 (14.1%)	Partially Devoiced
	10/30 (33.3%)	10/37 (27%)	3/11 (27.3%)	23/78 (29.5%)	Voiced

Table D.6: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 4 (Speaker CFGA).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	11/24 (45.8%)	9/24 (37.5%)	2/6 (33.3%)	22/54 (40.7%)	Devoiced
	4/24 (16.7%)	6/24 (25%)	3/6 (50%)	13/54 (24.1%)	Partially Devoiced
	9/24 (37.5%)	9/24 (37.5%)	1/6 (16.7%)	19/54 (35.2%)	Voiced
/z/	10/12 (83.3%)	12/16 (75%)	5/5 (100%)	27/33 (81.8%)	Devoiced
	1/12 (8.3%)	2/16 (12.5%)	0	3/33 (9.1%)	Partially Devoiced
	1/12 (8.3%)	2/16 (12.5%)	0	3/33 (9.1%)	Voiced
/ʒ/	9/12 (75%)	10/15 (66.7%)	11/12 (91.7%)	30/39 (76.9%)	Devoiced
	1/12 (8.3%)	2/15 (13.3%)	1/12 (8.3%)	4/39 (10.3%)	Partially Devoiced
	2/12 (16.7%)	3/15 (20%)	0	5/39 (12.8%)	Voiced
All Fric.	30/48 (62.5%)	31/55 (56.4%)	18/23 (78.3%)	79/126 (62.7%)	Devoiced
	6/48 (12.5%)	10/55 (18.2%)	4/23 (17.4%)	20/126 (15.9%)	Partially Devoiced
	12/48 (25%)	14/55 (25.5%)	1/23 (4.4%)	27/126 (21.4%)	Voiced

Table D.7: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 4 (Speaker ACC).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	14/23 (60.9%)	12/26 (46.2%)	3/4 (75%)	29/53 (54.7%)	Devoiced
	1/23 (4.4%)	2/26 (7.7%)	0	3/53 (5.7%)	Partially Devoiced
	8/23 (34.8%)	12/26 (46.2%)	1/4 (25%)	21/53 (39.6%)	Voiced
/z/	10/12 (83.3%)	16/19 (84.2%)	2/2 (100%)	28/33 (84.9%)	Devoiced
	2/12 (16.7%)	0	0	2/33 (6.1%)	Partially Devoiced
	0	3/19 (15.8%)	0	3/33 (9.1%)	Voiced
/ʒ/	9/12 (75%)	13/18 (72.2%)	3/9 (33.3%)	25/39 (64.1%)	Devoiced
	3/12 (25%)	3/18 (16.7%)	3/9 (33.3%)	9/39 (23.1%)	Partially Devoiced
	0	2/18 (11.1%)	3/9 (33.3%)	5/39 (12.8%)	Voiced
All Fric.	33/47 (70.2%)	41/63 (65.1%)	8/15 (53.3%)	82/125 (65.6%)	Devoiced
	6/47 (12.8%)	15/63 (7.9%)	3/15 (20%)	14/125 (11.2%)	Partially Devoiced
	8/47 (17%)	17/63 (27%)	4/15 (26.7%)	29/125 (23.2%)	Voiced

Table D.8: Inventory of all cases of devoicing (using the manual criterion). Values given are in the form x/y, where x = number of devoiced, partially devoiced or voiced examples, and y = total number of examples. Corpus 4 (Speaker ISSS).

	Word-Initial	Word-Medial	Word-Final	All Pos.	
/v/	6/16 (37.5%)	5/16 (31.3%)	2/4 (50%)	13/36 (36.1%)	Devoiced
	2/16 (12.5%)	1/16 (6.3%)	0	3/36 (8.3%)	Partially Devoiced
	8/16 (50%)	10/16 (62.5%)	2/4 (50%)	20/36 (55.6%)	Voiced
/z/	6/8 (75%)	7/12 (58.3%)	2/2 (100%)	15/22 (68.2%)	Devoiced
	1/8 (12.5%)	3/12 (25%)	0	4/22 (18.2%)	Partially Devoiced
	1/8 (12.5%)	2/12 (16.7%)	0	3/22 (13.6%)	Voiced
/ʒ/	5/8 (62.5%)	8/12 (66.7%)	6/6 (100%)	19/26 (73.1%)	Devoiced
	3/8 (37.5%)	3/12 (25%)	0	6/26 (23.1%)	Partially Devoiced
	0	1/12 (8.3%)	0	1/26 (3.9%)	Voiced
All Fric.	17/32 (53.1%)	20/40 (50%)	10/12 (83.3%)	47/84 (56%)	Devoiced
	6/32 (18.8%)	7/40 (17.5%)	0	13/84 (15.5%)	Partially Devoiced
	9/32 (28.1%)	13/40 (32.5%)	2/12 (16.7%)	24/84 (28.6%)	Voiced

Appendix E

Listings of Uvular Fricatives and Voiceless Tapped Alveolar Fricatives

This appendix lists the results of the time analysis of Corpus 3 and 4 uvular fricatives and voiceless tapped alveolar fricatives, and their VF and FV transitions. A broad phonetic transcription is also included. The data presented include: the VF transition duration, the fricative duration F, and the FV transition duration.

The file numbering of words from Corpus 4 has two parts: a number that refers to the sentence where the words occur and a word number which is the same as the one used in Corpus 3.

When the phonetic transcription, of words with fricatives at the beginning or end has an additional initial or final phoneme, separated from the transcription of the word we are analysing by a white space, this means that there is coarticulation between the fricative and the final or initial phoneme of the previous or following word in the sentence.

Table E.1: Voiceless uvular fricative /χ/. The data in the table is grouped by speaker, the words are separated by fricative word position (initial, medial and final), and ordered according to vowel context /i, ɨ, e, ɛ, ɐ, a, ɔ, o, u/.

Corpus	Speaker	Example	IPA	File N.	VF (ms)	F (ms)	FV (ms)
3	CFGA	relevo	[ɐ χɨ'lev]	94	27	82	43
3	CFGA	ressaca	[ɐ χ'sakɐ]	82	15	84	-
4	CFGA	ressaca	[ɐ χ'sakɐ]	7_82	28	55	-
4	CFGA	ressaca	[ɐ χ'sakɐ]	7r1_82	41	102	-
4	CFGA	ressaca	[ɐ χ'sakɐ]	7r2_82	25	75	-
3	CFGA	rosa	[ɐ 'χɔzɐ]	14	36	55	33
4	CFGA	rosa	[ɐ 'χɔzɐ]	4_14	25	69	40
4	CFGA	rosa	[ɐ 'χɔzɐ]	4r2_14	24	54	33
3	CFGA	garrafa	[gɐ'χafɐ]	71	18	64	24
3	CFGA	zurrar	[zu'χar]	16	38	47	36
3	ACC	ferro	[ɐ 'fɛχ]	4	17	23	-
3	ISSS	relevo	[ɐ χɨ'lev]	94	19	102	33
3	ISSS	ressaca	[ɐ χ'sakɐ]	82	18	66	-
4	ISSS	ressaca	[ɐ χ'sakɐ]	7_82	25	68	-
4	ISSS	ressaca	[ɐ χ'sakɐ]	7r_82	25	55	-
3	ISSS	rosa	[ɐ 'χɔzɐ]	14	20	100	25
4	ISSS	rosa	[ɐ 'χɔzɐ]	4_14	26	106	25
4	ISSS	rosa	[ɐ 'χɔzɐ]	4r_14	23	61	31
3	ISSS	garrafa	[gɐ'χafɐ]	71	30	72	22
3	ISSS	zurrar	[zu'χaɾ]	16	21	117	31
3	ISSS	ferro	['fɛχ]	4	23	100	-

Table E.2: Voiced uvular fricative /ʁ/.

Corpus	Speaker	Example	IPA	File N.	VF (ms)	F (ms)	FV (ms)
4	CFGA	rosa	[ɐ 'ʁɔzɐ]	4r1_14	28	28	63
3	ACC	garrafa	[gɐ'ʁafɐ]	71	32	42	39

Table E.3: Voiceless tapped alveolar fricative /ɾ̥/ (Speaker LMTJ).

Corpus	Example	IPA	File N.	VF (ms)	F (ms)	FV (ms)
4	ressaca	[ɾ̥ɨ'sakɐ]	7_82	-	12	13
3	rosa	[ɨ'ɾ̥ɔzɐ]	14	-	15	39
4	rosa	[a͡'ɾ̥ɔzɐ]	4_14	34	43	19
4	rosa	[a 'ɾ̥ɔzɐ]	4_14r	35	52	16
4	ressaca	[ɾ̥'sakɐ]	7_82r	-	28	-
3	girar	[ʒi'ɾ̥aɾ]	48	33	23	38
3	ferir	[fi'ɾ̥iɾ]	18	36	15	21
3	vermelho	[viɾ̥'mɐʎ]	86_1r	41	21	-
3	vermelho	[viɾ̥'mɐʎ]	86_2r	32	24	-
3	Zaire	['zajɾ̥ɨ]	99	32	22	16
3	arejar	[ɐɾ̥ɨ'ʒa]	46	43	16	14
4	arejar	[ɐɾ̥ɨ'ʒaɾ]	5_46r	33	27	-
3	garrafa	[gɐ͡'ɾ̥afɐ]	71	34	12	58
4	zarpar	[zɐɾ̥ɾ'paɾ]	3_113	25	23	-
4	zarpar	[zɐɾ̥͡ɾ'paɾ]	3_113r	17	27	-
3	amoroso	[ɐmu͡'ɾ̥oz]	8	31	27	30
3	chorar	[ʃu'ɾ̥aɾ]	78	40	20	39
3	falir	[fɐ'liɾ̥]	47	29	18	-
3	subir	[su'biɾ̥]	139	33	32	-
4	subir	[su'biɾ̥]	3_139r	40	42	-
3	viver	[vi'veɾ̥]	122	32	17	-
3	ver	[veɾ̥ɨ]	102	24	23	-
4	ver	[veɾ̥͡u]	12_102r	30	24	19
3	dever	[di'veɾ̥ɨ]	22	32	20	-
3	mover	[mu'veɾ̥͡ɨ]	57	16	19	-
3	saber	[sɐ'beɾ̥͡]	39	29	25	-
3	efectuar	[i'fɛtwaɾ̥]	98	37	12	-
3	vogar	[ɐ 'vugaɾ̥]	7	21	21	-
3	assar	[ɐ'saɾ̥ɨ]	137	28	21	-
4	assar	[ɐ'saɾ̥͡]	12_137r	103	36	-
3	zarpar	[zɐɾ'paɾ̥ɨ]	113	24	37	-
3	zurrar	[zu'Raɾ̥͡]	16	24	20	-
3	azar	[ɐ'zaɾ̥ɨ]	97	34	16	-
3	acusar	[ɐku'a͡ɾ̥]	146	30	24	-
3	chegar	[ʃɨ'gaɾ̥]	65	21	20	-
3	chamar	[ʃɐ'maɾ̥ɨ]	28	22	24	-
3	achar	[ɐ'ʃaɾ̥͡]	114	25	18	-

Table E.4: Voiceless tapped alveolar fricative /ɾ̥/ (Speaker ACC).

Corpus	Example	IPA	File N.	VF (ms)	F (ms)	FV (ms)
3	rosa	['ɾ̥ɔzɐ]	14	-	34	40
3	ressaca	[ɾ̥'sakɐ]	82	-	16	-
3	Zópiro	['zɔpiɾ̥u]	10	36	23	19
3	germe	['ʒɛɾ̥m]	72	34	18	-
4	zarpar	[zɐɾ̥'paɾ]	3r1_113	32	22	-
4	zarpar	[zɐɾ̥'paɾ]	3r2_113	22	21	-
3	zurrar	[zu'ɾ̥aɾ̥ɨ]	16	21	31	34
3	zurrar	[zu'ɾ̥aɾ̥ɨ]	16	27	13	26
4	furar	[fu'ɾ̥aɾ]	9r1_44	35	11	13
4	Brasil	[bɾ̥ɐ'zil]	10r1_58	-	15	19
3	ferir	[fi'ɾiɾ̥]	18	23	28	-
3	falir	[fɐ'liɾ̥]	47	44	15	-
3	subir	[su'biɾ̥]	139	30	24	-
3	agir	[ɐ'ʒiɾ̥]	108	30	20	-
3	Zaire	['zajɾ̥ɨ]	99	30	19	-
3	viver	[vi'veɾ̥]	122	14	33	-
3	ver	[veɾ̥]	102	30	23	-
4	ver	[veɾ̥]	12r1_102	51	24	-
3	dever	[dɨ'veɾ̥]	22	27	17	-
4	dever	[dɨ'veɾ̥ɨ]	12r1_22	26	18	-
3	aquecer	[ɐkɛ'seɾ̥]	131	40	25	-
3	efectuar	[i'fɛtwaɾ̥]	98	24	23	-
3	levar	[li'vaɾ̥ɨ]	105	19	14	-
3	assar	[ɐ'saɾ̥]	137	17	25	-
4	assar	[ɐ'saɾ̥]	12_137	30	33	-
4	assar	[ɐ'saɾ̥]	12r2_137	53	16	-
3	zelar	['zɨlaɾ̥]	126	27	20	-
3	azar	[ɐ'zaɾ̥]	97	25	13	-
3	girar	[ʒi'ɾaɾ̥ɨ]	48	19	15	-
3	originar	[oriʒi'naɾ̥ɨ]	1	27	19	-
4	furar	[fu'ɾaɾ̥]	9r2_44	31	15	-
3	tugir	[tu'ʒiɾɾ̥]	120	-	15	-

Table E.5: Voiceless tapped alveolar fricative /ɾ̥/ (Speaker CFGA).

Corpus	Example	IPA	File N.	VF (ms)	F (ms)	FV (ms)
4	ver	[veɾ̥ u]	12r2_102	65	36	21
4	assar	[ɐ'saɾ̥]	12r1_137	37	85	-

Table E.6: Voiceless tapped alveolar fricative /ɾ̥/ (Speaker ISSS).

Corpus	Example	IPA	File N.	VF (ms)	F (ms)	FV (ms)
3	vermelho	[viɾ̥'meʎu]	86	29	14	-
4	arejar	[ɐɾ̥i'ʒaɾ]	5_46	27	12	25
4	arejar	[ɐɾ̥'ʒaɾ]	5r_46	28	26	-
3	mares	['maɾ̥ʃ]	130	12	22	-
3	ferir	[fɨ'ɾiɾ̥]	18	37	32	-
3	subir	[su'biɾ̥]	139	40	22	-
4	subir	[su'biɾ̥]	3r_139	40	42	-
3	agir	[ɐ'ʒiɾ̥]	108	43	36	-
3	Zópiro	['zɔpiɾ̥]	10	35	24	-
3	chamar	[ʃɐ'maɾ̥]	28	29	20	-
4	assar	[ɐ'saɾ̥]	12_137	27	54	-
4	assar	[ɐ'saɾ̥]	12r_137	17	66	-
4	achar	[ɐ'ʃaɾ̥]	11r_114	46	41	-
3	Zaire	['zajɾ̥]	99	37	21	-

Appendix F

Bilingual Questionnaire

F.1 Speaker PS

1. Did you begin to learn both languages under five years of age? *Yes.*
2. If not, when did you become fluent in your second language?
3. Which language do you consider to be your primary language? *Both.*
4. Do you speak any other languages? Which one(s)? *None.*
5. What language do you speak with:
   (a) parents? *Portuguese with mother and English with father.*
   (b) brothers and sisters? *Mainly Portuguese.*
   (c) partner? *Portuguese.*
   (d) friends? *Portuguese and English.*
6. What language do you use:
   (a) for general use? *Portuguese and English.*
   (b) in the home? *Portuguese and English.*
   (c) at work? *English.*
   (d) on holiday? *Depends on place.*
7. When you think to yourself, do you think in:

(a) Portuguese?
(b) English?
(c) both? $\surd$

8. Which language did you use:
   (a) at school? *English.*
   (b) at university? *English.*
9. How many years have you lived in:
   (a) Portugal? 16
   (b) England? 6
   (c) other countries? 0
10. When you read books, do you mainly read in Portuguese or English? *English.*
11. Which language do you:
   (a) count in? *Usually English.*
   (b) swear in? *Both, usually Portuguese.*
   (c) dream in? *Portuguese and English.*

F.2 Speaker RS

1. Did you begin to learn both languages under five years of age? *Yes.*
2. If not, when did you become fluent in your second language?
3. Which language do you consider to be your primary language? *Both.*
4. Do you speak any other languages? Which one(s)? *French.*
5. What language do you speak with:
   (a) parents? *Portuguese with mother and English with father.*
   (b) brothers and sisters? *Portuguese.*
   (c) partner?
   (d) friends? *Portuguese and English.*

6. What language do you use:
    (a) for general use? *Portuguese and English.*
    (b) in the home? *Portuguese and English.*
    (c) at work? *English.*
    (d) on holiday? *Portuguese.*
7. When you think to yourself, do you think in:
    (a) Portuguese?
    (b) English?
    (c) both? $\surd$
8. Which language did you use:
    (a) at school? *English.*
    (b) at university? *English.*
9. How many years have you lived in:
    (a) Portugal? 16
    (b) England? 2
    (c) other countries? 0
10. When you read books, do you mainly read in Portuguese or English? *English.*
11. Which language do you:
    (a) count in? *English.*
    (b) swear in? *Portuguese.*
    (c) dream in? *Portuguese and English.*

Appendix G

Listings of English Corpora

G.1 Corpus 1a

1. /uvvvv ... u/
2. /əʒʒʒʒ ... ə/
3. /uʃʃʃʃ ... u/
4. /əvvvv ... ə/
5. /əʃʃʃʃ... ə/
6. /uffff ... u/
7. /ussss ... u/
8. /əzzzz ... ə/
9. /əffff ... ə/
10. /uzzzz ... u/
11. /əssss ... ə/
12. /uʒʒʒʒ ... u/

G.2 Corpus 1b

1. /ʃ/
   soft
   loud
2. /s/
   soft
   loud
3. /v/
   soft
   loud
4. /ʒ/
   soft
   loud
5. /f/
   soft
   loud
6. /z/
   soft
   loud
7. /s/
   soft
   loud
8. /v/
   soft
   loud
9. /f/
   soft
   loud
10. /ʒ/
    soft
    loud
11. /z/
    soft
    loud
12. /ʃ/
    soft
    loud

G.3 Corpus 2

1. /puʒu/
2. /pəzə/
3. /puvu/
4. /pəʃə/
5. /pəfə/
6. /pusu/
7. /pəʒə/
8. /puzu/
9. /pəvə/
10. /puʃu/
11. /pəsə/
12. /pufu/

G.4 Corpus 3

A total of 118 different words containing 142 fricatives: /f/ – 19 (11 word-initial; 4 word-medial; 4 word-final); /v/ – 24 (7 word-initial; 10 word-medial; 7 word-final); /θ/ – 9 (4 word-initial; 1 word-medial; 4 word-final); /ð/ – 9 (1 word-initial; 7 word-medial; 1 word-final); /s/ – 33 (14 word-initial; 12 word-medial; 7 word-final); /z/ – 17 (5 word-initial; 5 word-medial; 7 word-final); /ʃ/ – 20 (6 word-initial; 8 word-medial; 6 word-final); /ʒ/ – 11 (9 word-medial; 2 word-final [1]).

Fifi – /ˈfifi/.

fear – /fir/.

fever – /ˈfivɝ/.

finny – /ˈfɪn̩i/.

fishing – /ˈfɪʃiŋ/.

father – /ˈfɑðɝ/.

for – /fɔr/.

follow – /ˈfɔlo͜ʊ/.

Flossie – /ˈflɔsi/.

floozy – /ˈfluzi/.

frothing – /ˈfrɔθiŋ/.

Fifi – /ˈfifi/.

leafy – /ˈlifi/.

safer – /ˈse͜ɪfɝ/.

coffee – /ˈkɔfi/.

thief – /ˈθif/.

knife – /naɪf/.

half – /hɑf/.

[1]No English word begins with /ʒ/ (Ladefoged 1993, pp. 29).

roof – /ruf/.

veal – /vil/.

vision – /ˈvɪʒn̩/.

vet – /vɛt/.

velvet – /ˈvɛlvət/.

vying – /ˈva͜ɪjɪŋ/.

void – /vɔɪd/.

vulgar – /ˈvʌlgɝ/.

leaving – /ˈliviŋ/.

TV – /tiˈvi/ .

Levis – /ˈliva͜ɪz/.

fever – /ˈfivɝ/.

saver – /ˈse͜ɪvɝ/.

ivy – /ˈa͜ɪvi/.

lover – /ˈlʌvɝ/.

hover – /ˈhʌvɝ/.

cover – /ˈkʌvɝ/.

velvet – /ˈvɛlvət/.

live – /lɪv/.

give – /gɪv/.

save – /se͜ɪv/.

of – /əv/.

dove – /dəv/.

move – /muv/.

prove – /pruv/.

theatre – /ˈθiət^{h}ɜ/.

thief – /ˈθif/.

think – /ˈθɪŋk/.

thistle – /ˈθɪsl̩/.

frothing – /ˈfrɔθiŋ/.

wreath – /ˈriθ/.

death – /ˈdɛθ/.

breath – /ˈbrɛθ/.

bath – /ˈbæθ/.

the – /ðə/.

dither – /ˈdiðɝ/.

bather – /ˈbe͜ɪðər/.

mother – /ˈməðɝ/.

father – /ˈfɑðɝ/.

brother – /ˈbrʌðɝ/.

clothing – /ˈkloʊðiŋ/.

soothing – /ˈsuðiŋ/.

breathe – /ˈbrið/ .

see – /si/.

seizure – /ˈsiʒɝ/.

sees – /ˈsiz/.

seizing – /ˈsiziŋ/.

safer – /ˈseɪfɝ/.

save – /seɪv/.

saver – /ˈseɪvɝ/.

saying – /ˈseɪiŋ/.

says – /ˈsɛz/.

sandbar – /ˈsændbar/.

sewing – /ˈsoʊiŋ/.

Suzie – /ˈsuzi/.

soothing – /ˈsuðɪŋ/.

splashes – /ˈsplæʃəz/.

greasy – /ˈgrisi/.

missing – /ˈmɪsiŋ/.

kissing – /ˈkɪsiŋ/.

icy – /ˈɑɪsi/.

thistle – /ˈθɪsl̩/.

precision – /prəˈsɪʒn̩/.

desk – /dɛsk/.

Flossie – /ˈflɔsi/.

awesome – /ˈɔsʌm/.

Lucy – /ˈlusi/.

inducing – /ɪnˈduseŋ/.

monster – /ˈmɑnstɝ/.

across – /əˈkrɑs/.

hiss – /hɪs/.

mess – /mɛs/.

pass – /pas/.

moss – /mɔs/ .

blouse – /ˈblaʊs/.

cups – /ˈkʌps/.

zeal – /zil/ .

zero – /ˈzɪərəʊ/.

zip – /zɪp/.

zone – /zəʊn/.

zoo – /zu/.

seizing – /ˈsiziŋ/.

easy – /ˈizi/.

Suzie – /ˈsuzi/.

oozy – /ˈuzi/.

floozy – /ˈfluzi/.

sees – /ˈsiz/.

is – /ɪz/.

Levis – /ˈliva͜ɪz/.

was – /wəz/.

splashes – /ˈsplæʃəz/.

says – /ˈsɛz/.

has – /hæz/.

shocking – /ˈʃɔkɪŋ/.

shoo – /ʃuː/.

wishy – /ˈwɪʃi/.

fishing – /ˈfɪʃiŋ/.

machine – /məˈʃin̩/.

splashes – /ˈsplæʃəz/.

crashing – /ˈkræʃiŋ/.

washy – /ˈwɔʃi/.

washing – /ˈwɔʃiŋ/.

ocean – /ˈo͜ʊʃn̩/.

dish – /dɪʃ/.

mesh – /mɛʃ/.

rash – /ræʃ/.

hush – /hʌʃ/.

crush – /krʌʃ/.

push – /pʊʃ/.

she – /ʃɪ/.

shiny – /ˈʃa͜ɪn̩i/.

share – /ʃɛr/.

shark – /ʃark/.

seizure – /ˈsiʒɝ/.

Bijou – /bɪˈʒu/.

vision – /ˈvɪʒn̩/.

precision – /prəˈsɪʒn̩/.

leisure – /ˈlɛʒɝ/.

treasure – /ˈtrɛʒɝ/.

pleasure – /ˈplɛʒɝ/.

measure – /ˈmɛʒɝ/.

azure – /ˈæʒɝ/.

beige – /bɑɪʒ/ .

rouge – /ruʒ/.

G.5 Corpus 4

Only 62 of the 118 words in Corpus 3 were used in the sentences of Corpus 4. They contained 74 fricatives: /f/ – 9; /v/ – 7; /θ/ – 3; /ð/ – 5; /s/ – 22; /z/ – 11; /ʃ/ – 12; /ʒ/ – 5.

1. "I see Fifi vying for a share of icy ocean."
   /ɑ͜ɪ si ˈfifi ˈvɑ͜ɪjɪŋ fɔr ə ʃɛr əv ˈɑ͜ɪsi ˈo͜ʊʃn̩/

2. "She was in her leisure Levis and a shiny velvet blouse."
   /ʃɪ wəz ɪn hɝ ˈlɛʒɝ ˈlivɑ͜ɪz ænd ə ˈʃɑ͜ɪn̩i ˈvɛlvət ˈbla͜ʊs/

3. "She splashes in the ocean, across a sandbar, saying "I fear a shark may think a bather is a vision of an azure ocean". Shoo finny shark!"
   /ʃɪ ˈsplæʃəz ɪn ðə ˈo͜ʊʃn̩ əˈkrɑs ə ˈsændbar ˈse͜ɪiŋ ɑ͜ɪ fir ə ʃark me͜ɪ ˈθɪŋk ə ˈbe͜ɪðər ɪz ə ˈvɪʒn̩ əv ən̩ ˈæʒɝ ˈo͜ʊʃn̩ ʃu: ˈfɪn̩i ʃark/

4. "I see Fifi follow Suzie into the Bijou Theatre."
   /ɑ͜ɪ si ˈfifi ˈfɔlo͜ʊ ˈsuzi ˈintu ðə bɪˈʒu ˈθiət^hɜ/

5. "The oozy shocking murk inducing a seizure in the azure monster is awesome!"
   /ðə ˈuzi ˈʃɔkɪŋ mɝk ɪnˈduseŋ ə ˈsiʒɝ In̩ ðə ˈæʒɝ ˈmɑnstɝ ɪz ˈɔsʌm/

6. "He sees coffee cups on the desk."
   /hi ˈsiz ˈkɔfi ˈkʌps ɔn̩ ðə dɛsk/

7. "My mother has a washing machine and a lover."
   /mɑ͜ɪ ˈməðɝ hæz ə ˈwɔʃiŋ məˈʃin̩ ænd ə ˈlʌvɝ/

8. "He sees you and Lucy seizing Flossie."
   /hi ˈsiz j͜u ænd ˈlusi ˈsiziŋ ˈflɔsi/

9. "Wishy - washy, wishy - washy says the soothing frothing water, leaving easy sewing of the greasy clothing."
/ˈwɪʃi ˈwɔʃi ˈwɪʃi ˈwɔʃi ˈsɛz ðə ˈsuðiŋ ˈfrɔθiŋ ˈwɔtɝ ˈliviŋ ˈizi ˈso͜ʊiŋ əv ðə ˈgrisi ˈklo͜ʊðiŋ/

Bibliography

Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

Abramson, A. S. and L. Lisker (1973). Voice - timing perception in Spanish word - initial stops. *Journal of Phonetics 1*, 1–8.

Alwan, A. A. H. (1986). Acoustic and perceptual correlates of pharyngeal and uvular consonants. M.Sc. thesis, Massachusetts Institute of Technology, Cambridge, USA.

Andrade, A. (1982). Reduction of unstressed vowels in Portuguese. Seminar presented at the Department of Linguistics and Phonetics, University of Leeds, UK.

Andrade, A. (1994). Reflexões sobre o 'e mudo' em português europeu. In *Actas do Congresso Internacional Sobre o Português*, Volume 2, Lisboa, Portugal, pp. 303–344.

Andrade, A. (1995). Percepção de C ou CC oclusivas por ouvintes nativos de português europeu. In *Actas do XI Encontro Nacional da Associação Portuguesa de Linguística*, Volume 3, Lisboa, Portugal, pp. 153–186.

Andrade, A., I. Simas, and D. Sarroeira (1999). Efeito do arredondamento vocálico sobre o contexto consonântico: Notícia de uma investigação em curso. In *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*, Volume 2, Faro, Portugal, pp. 625–635. Presented at XIV Encontro Nacional da Associação Portuguesa de Linguística, Aveiro, Portugal, September 1998.

Bachman, L. F. and A. S. Palmer (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests.* Oxford: Oxford University Press.

Badin, P. (1989). Acoustics of voiceless fricatives: Production theory and data. Quarterly Progress and Status Report 3/1989, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden.

Badin, P. (1991). Fricative consonants: Acoustic and X-ray measurements. *Journal of Phonetics 19*(3-4), 397–408.

Badin, P. and C. G. M. Fant (1984). Notes on vocal tract computation. Quarterly Progress and Status Report 2-3/1984, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden.

Badin, P., C. H. Shadle, Y. P. T. Ngoc, J. N. Carter, W. S. C. Chiu, C. Scully, and K. Stromberg (1994). Frication and aspiration noise sources: Contribution of experimental data to articulatory synthesis. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 94)*, Volume 1, Yokohama, Japan, pp. 163–166.

Baum, S. R. and S. E. Blumstein (1987). Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *Journal of the Acoustical Society of America 82*(3), 1073–1077.

Beautemps, D., P. Badin, and R. Laboissière (1993). Recovery of vocal tract midsagittal and area functions from speech signal for vowels and fricative consonants. In *Proceedings of the 3rd European Conference on Speech Communication and Technology (EuroSpeech'93)*, Volume 1, Berlin, Germany, pp. 73–76.

Beautemps, D., P. Badin, and R. Laboissière (1995). Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data. *Speech Communication 16*, 27–47.

Behrens, S. J. and S. E. Blumstein (1988). Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics 16*, 295–298.

Bendat, J. S. and A. G. Piersol (2000). *Random Data: Analysis and Measurement Procedures* (Third ed.). New York: John Wiley.

Berkovits, R. (1993). Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech 36*(1), 89–98.

Bond, Z. S., J. E. Eddey, and J. J. Bermejo (1980). VOT del Español to English: Comparison of a language-disordered and normal child. *Journal of Phonetics 8*, 287–291.

Caramazza, A., G. H. Y. Komshian, E. B. Zurif, and E. Carbone (1973). The aquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America 54*(2), 421–428.

Chen, W. S. and A. A. H. Alwan (2000). Place of articulation cues for voiced and voiceless plosives and fricatives in syllable-initial position. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China.

Choo, W. (1999). The relationship between perceptual and physical space of fricatives. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 99)*, San Francisco, USA, pp. 163–166.

Choo, W. and M. A. Huckvale (1997). Spatial relationships in fricative perception. Speech Hearing and Language: Work in Progress 10, University College London, Department of Phonetics and Linguistics, London, UK. `http://www.phon.ucl.ac.uk/home/shl10/won/space.htm`.

Cole, R. A. and W. E. Cooper (1975). Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America 58*(6), 1280–1287.

Crystal, D. (1997). *A Dictionary of Linguistics and Phonetics* (Fourth ed.). Oxford: Blackwell.

Crystal, T. H. and A. S. House (1988). A note on the durations of fricatives in American English. *Journal of the Acoustical Society of America 84*(5), 1932–1935.

Cunha, C. and L. F. L. Cintra (1992). *Nova Gramática do Português Comtemporâneo* (Ninth ed.). Lisboa: Edições João Sá da Costa.

Davies, P. O. A. L., R. S. McGowan, and C. H. Shadle (1993). Practical flow duct acoustics applied to the vocal tract. In I. R. Titze (Ed.), *Vocal Fold Physiology: Frontiers in Basic Science*, pp. 93–142. San Diego: Singular.

Delattre, P. C. (1971). Pharyngeal features in the consonants of Arabic, German, Spanish, French, and American English. *Phonetica 23*, 129–155.

Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents.* Berlin: Foris Publications.

Engwall, O. and P. Badin (2000). An MRI study of Swedish fricatives: Coarticulatory effects. In *Proceedings of the 5th Seminar on Speech Production Models and Data*, Kloster Seeon, Bavaria, Germany, pp. 297–300.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association 26*(2), 65–88.

Evers, V., H. Reetz, and A. Lahiri (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics 26*, 345–370.

Field, A. (2000). *Discovering Statistics Using SPSS for Windows.* London: SAGE.

Flege, J. E., S. G. Fletcher, and A. Homiedan (1988). Compensating for a bite block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data. *Journal of the Acoustical Society of America 83*(1), 212–228.

Fletcher, S. G. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research 32*, 736–748.

Fletcher, S. G. and D. G. Newman (1991). [s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width. *Journal of the Acoustical Society of America 89*(2), 850–858.

Forrest, K., G. Weismer, P. Milenkovic, and R. N. Dougall (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America 84*(1), 115–123.

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception and Psychophysics 55*(6), 597–610.

Funatsu, S. (1995). Cross language study of perception of dental fricatives in Japanese and Russian. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 4, Stockholm, Sweden, pp. 124–127.

Goldstein, M. E. (1976). *Aeroacoustics.* New York: McGraw - Hill.

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics 6*, 95–102.

Hamers, J. F. and M. H. A. Blanc (2000). *Bilinguality and Bilingualism* (Second ed.). Cambridge: Cambridge University Press.

Hazan, V. L. and G. Boulakia (1993). Perception and production of a voicing contrast by French - English bilinguals. *Language and Speech 36*(1), 17–38.

Hixon, T. J. (1966). Turbulent noise sources for speech. *Folia Phoniatrica 18*(3), 168–182.

Hixon, T. J., F. D. Minifie, and C. A. Tait (1967). Correlates of turbulent noise production for speech. *Journal of Speech and Hearing Research 10*, 133–140.

Hogan, J. T. and A. J. Rozsypal (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word - final consonant. *Journal of the Acoustical Society of America 67*(5), 1764–1771.

Holmes, J. N. (1983). Research report – formant synthesizers: Cascade or parallel? *Speech Communication 2*(4), 251–273.

Holmes, W. J., J. N. Holmes, and M. W. Judd (1990). Extension of the bandwidth of the JSRU parallel - formant synthesizer for high quality synthesis of male and female speech. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 90)*, Volume 1, Albuquerque, USA, pp. 313–316.

Holtrup, G. (1998). From magnetic resonance imaging (MRI) data to simulation of speech sounds. Tripartite 5th year project report, Department of Electronics and Computer Science, University of Southampton, Southampton, UK.

Hoole, P., N. N. Trong, and W. J. Hardcastle (1993). A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. *Language and Speech 36*(2, 3), 235–260.

Hoole, P., W. Ziegler, E. Hartmann, and W. J. Hardcastle (1989). Parallel electropalatographic and acoustic measures of fricatives. *Clinical Linguistics and Phonetics 3*(1), 59–69.

Howell, P. and S. Rosen (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of the Acoustical Society of America 73*(3), 976–984.

Hughes, A. (1989). *Testing for Language Teachers.* Cambridge: Cambridge University Press.

Hughes, G. W. and M. Halle (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America 28*(3), 303–310.

IPA (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Jackson, P. J. B. (2000). *Characterisation of Plosive, Fricative and Aspiration Components in Speech Production.* Ph.D. Thesis, Department of Electronics and Computer Science, University of Southampton, Southampton, UK.

Jassem, W. (1967). Acoustical description of voiceless fricatives in terms of spectral parameters. In W. Jassem (Ed.), *Speech Analysis and Synthesis I*, pp. 189–206. Polish Academy of Sciences/Posnán.

Johns, C. M. (1972). Slips of the tongue in Portuguese. M.Litt. Thesis, University of Edinburgh, Edinburgh, UK.

Jongman, A., R. Wayland, and S. Wong (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America 108*(3), 1252–1263.

Kent, R. D. and C. Read (1992). *The Acoustic Analysis of Speech.* San Diego: Singular.

Klatt, D. H. (1971). On predicting the duration of the phonetic segment [s] in English. Quarterly Progress Report 103, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, USA.

Klatt, D. H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research 17*(1), 51–63.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics 3*(3), 129–140.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America 59*(5), 1208–1221.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America 67*(3), 971–995.

Klatt, D. H. and L. C. Klatt (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America 87*(2), 820–857.

Komshian, G. H. Y. and S. D. Soli (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative - vowel coarticulation. *Journal of the Acoustical Society of America 70*(4), 966–975.

Konefal, J. A. and J. Fokes (1981). Voice onset time: The development of Spanish / English distinction in normal and language disordered children. *Journal of Phonetics 9*, 437–444.

Krane, M. H. (1999). Fluid dynamic effects in speech. *Journal of the Acoustical Society of America 105*(2, Pt. 2), 1159.

Lacerda, A. and F. M. Rogers (1939). *Sons Dependentes da Fricativa Palatal Áfona, em Português.* Coimbra: Laboratório de Fonética Experimental da Faculdade de Letras da Universidade de Coimbra, Fundação do Instituto para a Alta Cultura, Portugal.

Lacerda, F. P. (1982). Acoustic perceptual study of the Portuguese voiceless fricatives. *Journal of Phonetics 10*, 11–22.

Ladefoged, P. (1993). *A Course in Phonetics* (Third ed.). Fort Worth: Harcourt Brace.

Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages.* Oxford: Blackwell.

LaRiviere, C., H. Winitz, and E. Herriman (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of the Acoustical Society of America 18*(4), 613–622.

Laver, J. (1994). *Principles of Phonetics.* Cambridge: Cambridge University Press.

Lindblad, P. (1980). *Svenskans SJE- Och TJE-LJU: I Ett Allmänfonetiskt Perspektiv (Some Swedish Sibilants).* Travaux de L'Institut de Linguistic de Lund XVI, CWK Gleerup.

Liu, M. and A. Lacroix (1997). Pole-zero modeling of vocal tract for fricative sounds. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 97)*, Volume 3, Munich, Germany, pp. 1659–1662.

Maddieson, I. (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Mair, S. J. and C. H. Shadle (1996). The voiced/voiceless distinction in fricatives: EPG, acoustic and aerodynamic data. In *Proceedings of the Institute of Acoustics Autumn Conference (Speech and Hearing 96), book 1*, Volume 18, part 9, Windermere, UK, pp. 163–169.

Manrique, A. M. B. and M. I. Massone (1981). Acoustic analysis and perception of Spanish fricative consonants. *Journal of the Acoustical Society of America 69*(4), 1145–1153.

Martins, M. R. D. (1975). Vogais e consoantes do português: Estatística de ocorrência, duração e intensidade. *Boletim de Filologia, Centro de Estudos Filológicos, Lisboa, Portugal 24*(1-4), 1–11.

Martins, M. R. D., B. Harmegnies, and D. Poch (1995). Changement phonétique en cours du portugais européen. In *Actas do XI Encontro Nacional da Associação Portuguesa de Linguística*, Volume 3, Lisboa, Portugal, pp. 249–259.

Mateus, M. H. M. (1996). Fonologia. In I. H. Faria, E. R. Pedro, I. Duarte, and C. A. M. Gouveia (Eds.), *Introdução à Linguística Geral e Portuguesa*, Chapter 4, pp. 171–199. Lisboa: Editorial Caminho.

Mateus, M. H. M. and E. Andrade (2000). *The Phonology of Portuguese.* Oxford: Oxford University Press.

Mohammad, M. A. S. (1999). *Dynamic Measurements of Speech Articulators Using Magnetic Resonance Imaging.* Ph.D., Department of Electronics and Computer Science, University of Southampton, Southampton, UK.

Motoki, K., P. Badin, X. Pelorson, and H. Matsuzaki (2000). A modal parametric method for computing acoustic characteristics of three-dimensional vocal tract models. In *Proceedings of the 5th Seminar on Speech Production Models and Data*, Kloster Seeon, Bavaria, Germany, pp. 325–328.

Motoki, K., X. Pelorson, P. Badin, and H. Matsuzaki (2000). Computation of 3-D vocal tract acoustics based on mode-matching technique. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China.

Narayanan, S. S. (1995). *Fricative Consonants: An Articulatory, Acoustic, and Systems Study.* Ph.D Thesis, Department of Electrical Engineering, University of California at Los Angeles (UCLA), Los Angeles, USA.

Narayanan, S. S. and A. A. H. Alwan (2000). Noise source models for fricative consonants. *IEEE Transactions on Speech and Audio Processing 8*(2), 328–344.

Narayanan, S. S., A. A. H. Alwan, and K. Haker (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America 98*(3), 1325–1347.

Nartey, J. N. A. (1982). On fricative phones and phonemes: Measuring the phonetic differences within and between languages. UCLA Working Papers in Phonetics 55, University of California, Los Angeles, USA.

Nelson, P. A. and C. L. Morfey (1981). Aerodynamic sound production in low speed flow ducts. *Journal of Sound and Vibration 79*(2), 263–289.

Niikawa, T., M. Matsumura, T. Tachimura, and T. Wada (2000). Modeling of a speech production system based on MRI measurement of three-dimensional vocal tract shapes during fricative consonant phonation. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China.

O'Shaughnessy, D. (1974). Consonant durations in clusters. *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-22*(4), 282–295.

Pirello, K., S. E. Blumstein, and K. Kurowski (1997). The characteristics of voicing in syllable-initial fricatives in American English. *Journal of the Acoustical Society of America 101*(6), 3754–3765.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America 51*(4), 1296–1303.

Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics 19*(3,4), 267–280.

Riegelsberger, E. L. (1997). *The Acoustic-to-Articulatory Mapping of Voiced and Fricated Speech.* Ph.D., Department of Electrical Engineering, The Ohio State University, Ohio, USA.

Schwartz, M. F. (1969). Influence of vowel environment upon the duration of /s/ and /ʃ/. *Journal of the Acoustical Society of America 46*(2), 480–481.

Scully, C. (1971). A comparison of /s/ and /z/ for an English speaker. *Language and Speech 14*(2), 187–200.

Scully, C. (1979). Model prediction and real speech: Fricative dynamics. In B. Lindblom and S. E. G. Öhman (Eds.), *Frontiers of Speech Communication Research*, pp. 35–48. London: Academic Press.

Scully, C. (1990). Articulatory synthesis. In W. J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling*, pp. 151–186. Dordrecht: Kluewer Academic.

Scully, C. (1992). Articulatory actions within a phonological systems and the resulting complexity of speech signals. *Phonetica 49*, 212–221.

Scully, C. and E. Allwood (1985). Production and perception of an articulatory continuum for fricatives of English. *Speech Communication 4*, 237–245.

Scully, C., E. Castelli, E. Brearley, and M. Shirt (1992). Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives. *Journal of Phonetics 20*(1), 39–51.

Scully, C., E. G. Georges, and E. Castelli (1992). Articulatory paths for some fricatives in connected speech. *Speech Communication 11*(4-5), 411–416.

Shadle, C. H. (1985). *The Acoustics of Fricative Consonants.* Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA. Released as Research Laboratory of Electronics Technical Report 506.

Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling*, pp. 187–209. Dordrecht: Kluwer Academic.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics 19*(3-4), 409–424.

Shadle, C. H. (1992). Progress reports 1990 - 92. In B. Guerin, editor, Mesure, Caractérisation et Modélisation des Sons Fricatifs, EC SCIENCE Project SCI*0147 - C(EDB).

Shadle, C. H. (1995). Modelling the noise source in voiced fricatives. In *Proceedings of the 15th International Congress on Acoustics (ICA 95)*, Trondheim, Norway, pp. 145–148.

Shadle, C. H., P. Badin, and A. Moulinier (1991). Towards the spectral characteristics of fricative consonants. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS 91)*, Volume 3, Aix-en-Provence, France, pp. 42–45.

Shadle, C. H. and J. N. Carter (1993). WP1: From speech signal to acoustic sources. In P. Badin, C. Abry and C. Scully, editors, Speech MAPS Year 1 Report, ESPRIT project 6975, v.2.

Shadle, C. H., C. U. Dobelke, and C. Scully (1992). Spectral analysis of fricatives in vowel context. *Journal de Physique 2*, 295–298.

Shadle, C. H. and S. J. Mair (1996). Quantifying spectral characteristics of fricatives. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 96)*, Philadelphia, USA, pp. 1517–1520.

Shadle, C. H., S. J. Mair, and J. N. Carter (1996). Acoustic characteristics of the front fricatives [f, v, θ, ð]. In *Proceedings of the 1st ESCA Tutorial and Research Workshop (ETRW) on Speech Production Modeling – 4th Speech Production Seminar*, Autrans, France, pp. 193–196.

Shadle, C. H., S. J. Mair, J. N. Carter, and N. Millner (1995). The effect of vowel context on acoustic characteristics of [ç, x]. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 1, Stockholm, Sweden, pp. 66–69.

Shadle, C. H., A. Moulinier, C. U. Dobelke, and C. Scully (1992). Ensemble averaging applied to the analysis of fricative consonants. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 92)*, Volume 1, Banff, pp. 53–56.

Shadle, C. H. and C. Scully (1995). An articulatory - acoustic - aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics 23*(1, 2), 53–66.

Shadle, C. H., M. Tiede, S. Masaki, Y. Shimada, and I. Fujimoto (1996). An MRI study of the effects of vowel context on fricatives. In *Pro-*

ceedings of the Institute of Acoustics Autumn Conference (Speech and Hearing 96), book 1, Volume 18, part 9, Windermere, UK, pp. 187–194.

Shinn, P. C. (1985). *A Cross-Language Investigation of Stop, Affricate and Fricative Manners of Articulation.* Ph.D. Thesis, Brown University, Providence, USA.

Slis, I. H. and A. Cohen (1969). On the complex regulating the voiced-voiceless distinction II. *Language and Speech 12*(3), 137–155.

Smith, C. L. (1995). Contextual influences on devoicing of /z/ in American English. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 1, Stockholm, Sweden, pp. 380–383.

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics 25*(4), 471–500.

Solé, M. J., J. J. Ohala, and G. Ying (1998). Aerodynamic characteristics of trills. In *Proceedings of the 16th International Congress on Acoustics (ICA 98) and 135th Meeting of the Acoustical Society of America*, Volume 4, Seattle, USA, pp. 2923–2924.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America 70*(4), 976–984.

Soli, S. D. (1982). Structure and duration of vowels together specify fricative voicing. *Journal of the Acoustical Society of America 72*(2), 366–378.

SPSS (1999a). *SPSS Advanced Models 10.0.* Chicago: SPSS Inc.

SPSS (1999b). *SPSS Base 10.0 User's Guide.* Chicago: SPSS Inc.

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America 50*(4), 1180–1192.

Stevens, K. N. (1987). Interaction between acoustic sources and vocal-tract configurations for consonants. In *Proceedings of the 11th International Congress of Phonetic Sciences (ICPhS 87)*, Volume 3, Tallinn, Estonia, USSR, pp. 385–389.

Stevens, K. N. (1991). Vocal-fold vibration for obstruent consonants. In J. Gauffin and B. Hammarberg (Eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, pp. 29–36. San Diego: Singular.

Stevens, K. N. (1997). Articulatory-acoustic-auditory relationships. In W. J. Hardcastle and J. Laver (Eds.), *The Handbook of Phonetic Sciences*, Chapter 15, pp. 462–506. Oxford: Blackwell.

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge: MIT Press.

Stevens, K. N., S. E. Blumstein, L. B. Glicksman, M. Burton, and K. Kurowski (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America 91*(5), 2979–3000.

Stone, M., A. Faber, L. J. Raphael, and T. H. Shawker (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics 20*(2), 253–270.

Stone, M. and A. Lundberg (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America 99*(6), 3728–3737.

Sussman, H. M. (1994). The phonological reality of locus equations across manner class distinctions: Preliminary observations. *Phonetica 51*, 119–131.

Trong, N. N. and P. Hoole (1993). Frequency variations of the lowest main spectral peak in sibilant clusters. In *Proceedings of the 3rd European Conference on Speech Communication and Technology (EuroSpeech'93)*, Volume 1, Berlin, Germany, pp. 81–84.

Trong, N. N., P. Hoole, and A. Marchal (1994). Regenerating the spectral shape of [s] and [ʃ] from a limited set of articulatory parameters. *Journal of the Acoustical Society of America 96*(1), 33–39.

Veatch, T. C. (1989). Word-final devoicing of fricatives in English. In *Linguistic Society of America Winter Meeting Handbook*, Washington, USA.

Vescovi, C. and E. Castelli (1995). Inversion of the voice source for some fricatives. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 1, Stockholm, Sweden, pp. 70–73.

Viana, M. C. (1984). *Etude de Deux Aspects du Consonantisme du Portugais: Fricatisation et Dévoisement.* Ph.D. Thesis (Doct. 3ème Cycle), Université des Sciences Humaines de Strasbourg, Strasbourg, France.

Watson, I. (1990). Aquiring the voicing contrast in French: A comparative study of monolingual and bilingual children. In J. N. Green and W. A. Bennett (Eds.), *Variation and Change in French: Essays Presented to Rebecca Posner on Occasion of her Sixtieth Birthday*, pp. 37–60. London: Routledge.

Watson, I. (1991). Phonological processing in two languages. In E. Bialystok (Ed.), *Language Processing in Bilingual Children*, Chapter 2, pp. 25–48. Cambridge: Cambridge University Press.

Wilde, L. F. (1993). Inferring articulatory movements from acoustic properties at fricative - vowel boundaries. In *The Acoustical Society of America (ASA) 126th Meeting*, Volume 5aSP11, Denver, USA.

Wilde, L. F. (1995a). *Analysis and Synthesis of Fricative Consonants.* Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA.

Wilde, L. F. (1995b). Quantifying time - varying spectra of English fricatives. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 4, Stockholm, Sweden, pp. 120–123.

Williams, L. (1977). The perception of stop consonant voicing by Spanish - English bilinguals. *Perception and Psychophysics 21* (4), 289–297.

Zagar, L. E. (1986). *The Fricative Sound Source Spectrum Derived from a Vocal Tract Analog.* Ph.D. Thesis, Agricultural and Mechanical College, Louisiana State University, Baton Rouge, USA.

Further Reading on Fricatives

Aeroacoustics of Fricative Production

Stromberg, K., C. Scully, P. Badin, and C. H. Shadle (1994). Aerodynamic patterns as indicators of articulation and acoustic sources for fricatives produced by different speakers. In *Proceedings of the Institute of Acoustics*, Volume 16, part 5, UK, pp. 325–333.

Badin, P., K. Mawass, and E. Castelli (1995). A model of frication noise source based on data from fricative consonants in vowel context. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 2, Stockholm, Sweden, pp. 202–205.

Nonlinear Dynamics

Narayanan, S. S. and A. A. H. Alwan (1993). Strange attractors and chaotic dynamics in the production of voiced and voiceless fricatives. In *Proceedings of the 3rd European Conference on Speech Communication and Technology (EuroSpeech'93)*, Volume 1, Berlin, Germany, pp. 77–80.

Narayanan, S. S. and A. A. H. Alwan (1995). A nonlinear dynamical systems analysis of fricative consonants. *Journal of the Acoustical Society of America 97*(4), 2511–2524.

Acoustic to Articulatory Mapping

Djéradi, A., B. Guérin, P. Badin, and P. Perrier (1991). Measurement of the acoustic transfer function of the vocal tract: A fast and accurate method. *Journal of Phonetics 19*(3/4), 387–395.

Sorokin, V. N. (1994). Inverse problem for fricatives. *Speech Communication 14*(3), 249–262.

Badin, P., D. Beautemps, R. Laboissière, and J. L. Schwartz (1995). Recovery of vocal tract geometry from formants for vowels and fricative consonants using a midsagittal-to-area function conversion model. *Journal of Phonetics 23*(1, 2), 221–229.

Direct and Indirect Measurements of Fricative Production

Hasegawa, A., J. M. Christensen, M. J. McCutcheon, and S. G. Fletcher (1979). Articulatory properties of /s/ in selected consonant clusters. In J. J. Wolf and D. H. Klatt (Eds.), *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, Cambridge, USA, pp. 115–118.

Hamlet, S. L., H. T. Bunnell, and B. Struntz (1986). Articulatory asymmetries. *Journal of the Acoustical Society of America 79*(4), 1165–1169.

Boussaffa, F., M. Jomaa, and R. Sock (1991). Les constraintes temporelles des types consonantiques sur le timing mandibulaire de la quantité en arabe tunisien. In *Proceedings of the 12th International Congress of Phonetic Sciences (ICPhS 91)*, Volume 3, Aix-en-Provence, France, pp. 306–309.

Badin, P., K. Motoki, N. Miki, D. Ritterhaus, and M.-T. Lallouache (1994). Some geometric and acoustic properties of the lip horn. *Journal of the Acoustical Society of Japan 15*(4), 243–253.

Badin, P., B. Gabioud, D. Beautemps, T. M. Lallouache, G. Bailly, S. Maeda, J. P. Zerling, and G. Brock (1995). Cineradiography of VCV sequences: Articulatory-acoustic data for a speech production model. In *Proceedings of the 15th International Congress on Acoustics (ICA 95)*, Trondheim, Norway, pp. 349–352.

Stone, M. (1991). Toward a model of three-dimensional tongue movement. *Journal of Phonetics 19*(3/4), 309–320.

Lindblad, P. and S. Lundqvist (1995). The groove production of Swedish sibilants - an EPG analysis. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 2, Stockholm, Sweden, pp. 458–461.

Chiu, W. S. C. and C. H. Shadle (1992). Use of palate shape data in an enhanced electropalatography system. In *Proceedings of the Institute of Acoustics, 1992 Autumn Conference (Speech and Hearing), Book 2*, Volume 14, part 6, pp. 415–422.

Perkell, J. S., M. L. Matthies, and M. Zandipour (1998). Motor equivalence in the production of /ʃ/. In *Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, Volume 4, Seattle, USA, pp. 2925–2926.

Source and Vocal Tract Modelling: Interactions Between Source Aeroacoustic Mechanisms and Vocal Tract Configurations

Badin, P. and C. G. M. Fant (1989). Fricative production modelling: Aerodynamic and acoustic data. In *Proceedings of the 2nd European Conference on Speech Communication and Technology (EuroSpeech'89)*, Paris, France.

Shadle, C. H. (1988). Experimental derivation of fricative source models. In *Proceedings of Speech'88 (7th FASE Symposium)*, Edinburgh, UK, pp. 399–406.

Jackson, P. J. B. and C. H. Shadle (2000). Frication noise modulated by voicing, as revealed by pitch - scaled decomposition. *Journal of the Acoustical Society of America 108*(4), 1421–1434.

Pastel, L. M. P. (1987). Turbulent noise sources in vocal tract models. M.Sc. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA.

Sinder, D. J. (1999). *Speech Synthesis Using an Aeroacoustic Fricative Model.* Ph.D. Thesis, Rutgers, The State University of New Jersey, New Brunswick, USA.

Perceptual Studies

Massaro, D. W. and M. M. Cohen (1976). The contribution of fundamental frequency and voice onset time to the /zi/–/si/ distinction. *Journal of the Acoustical Society of America 60*(3), 704–717.

McCasland, G. P. (1979). Noise intensity and spectrum cues for spoken fricatives. In J. J. Wolf and D. H. Klatt (Eds.), *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, Cambridge, USA, pp. 303–306.

Mann, V. A. and B. H. Repp (1980). Influence of vocalic context on perception of the [ʃ] - [s] distinction. *Perception and Psychophysics 28*(3), 213–228.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception and Psychophysics 30*(3), 217–227.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America 85*(4), 1718–1725.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s] - [š] boundary. *Journal of the Acoustical Society of America 69*(1), 275–282.

Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and Speech 26*(1), 91–100.

Whalen, D. H. (1991). Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America 90*(4), 1776–1785.

Faulkner, A., S. Rosen, A. M. Darling, and M. Huckvale (1995). Cue interaction in the perception of intervocalic and syllable-initial voiceless fricative/affricate contrasts. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 2, Stockholm, Sweden, pp. 502–505.

Nguyen, N. (1995). Contextual and lexical effects in the identification of fricatives. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 2, Stockholm, Sweden, pp. 530–533.

Hedrick, M. S. and R. N. Ohde (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America 94*(4), 2005–2026.

Hedrick, M. S. (1997). Effect of acoustic cues on labeling fricatives and affricates. *Journal of Speech and Hearing Research 40*(4), 925–938.

Development of Speech and Pathological Speech

McGowan, R. S. and S. Nittrouer (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and

/s/. *Journal of the Acoustical Society of America 83*(1), 229–236.

Dent, H., F. Gibbon, W. J. Hardcastle, and M. Wakumoto (1995). Articulatory/acoustic relationships in lateralised productions of sibilant fricatives. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95)*, Volume 2, Stockholm, Sweden, pp. 654–657.

Matthies, M. L., M. A. Svirsky, H. L. Lane, and J. S. Perkell (1994). A preliminary study of the effects of cochlear implants on the production of sibilants. *Journal of the Acoustical Society of America 96*(3), 1367–1373.

Matthies, M. L., M. A. Svirsky, J. S. Perkell, and H. L. Lane (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *Journal of Speech and Hearing Research 39*, 936–946.

Baum, S. R. (1996). Fricative production in aphasia: Effects of speaking rate. *Brain and Language 52*(2), 328–341.

Author's Relevant Publications

Jesus, L. M. T. (1999). *Analysis of Portuguese Fricative Consonants.* Mini Thesis, Department of Electronics and Computer Science, University of Southampton, Southampton, UK.

Jesus, L. M. T. and C. H. Shadle (1999). Acoustic analysis of a speech corpus of European Portuguese fricative consonants. In *Proceedings of the 6th European Conference on Speech Communication and Technology (EuroSpeech'99)*, Volume 1, Budapest, Hungary, pp. 431-434.

Jesus, L. M. T. and C. H. Shadle (2000). Parameterizing spectral characteristics of European Portuguese fricatives. In *Proceedings of the 5th Seminar on Speech Production Models and Data*, Kloster Seeon, Bavaria, Germany, pp. 301-304.

Jesus, L. M. T. and C. H. Shadle (2001). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics.* To be published in a Special Issue.