ESTIMATION OF STOPS' SPECTRAL PLACE CUES USING MULTITAPER TECHNIQUES

(Estimação das características espectrais relacionadas com o ponto de articulação de oclusivas utilizando a técnica multitaper)

Marisa Lobo LOUSADA *(ESSUA, University of Aveiro)*[*]
Luis M. T. JESUS *(IEETA e ESSUA, University of Aveiro)*[**]
Daniel PAPE *(IEETA, University of Aveiro)*[***]

Abstract: *This study focuses on the spectral characteristics of the European Portuguese stops* /p, b, t, d, k, g/ *produced by six native speakers. We analysed the spectral peaks and troughs by means of multitaper spectra and performed a parameterisation of the stop spectra using slope and moment measures. In comparison to traditional spectral estimations, multitaper is more exact and, more importantly, not limited to a stationary signal length necessary for the analysis window. Therefore, it is well-suited for the rather short duration of the burst of a stop. Results show that the burst characteristics vary with place of articulation. While the global spectral frequencies match the data in classical literature, it is shown that other spectral measures in our data do not follow the typical classical spectral patterns. It is discussed whether these differences are due to the use of different methodology, or substantial cross-linguistic differences in the spectral characteristics.*

Key-words: *Multitaper Analysis; Peaks and troughs; Slopes; Moments.*

Resumo: *Este estudo analisa as caraterísticas espetrais das oclusivas* /p, b, t, d, k, g/ *do Português Europeu produzidas por seis informantes nativos. Procedeu-se à análise dos picos e dos vales espetrais (espetros multitaper) e à parametrização das caraterísticas espetrais através da análise dos declives dos espetros e do cálculo dos momentos de distribuição. Comparativamente às análises espetrais tradicionais, a análise multitaper permite um maior controlo da variância associada a este tipo de sinais fazendo uso de apenas uma curta*

---

[*] School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal.

[**] Corresponding author. Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal and School of Health Sciences (ESSUA), University of Aveiro, 3810-193 - Aveiro, Portugal.

[***] Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal.

D.E.L.T.A., 28:1, 2012 (1-26)

janela, o que se adequa às caraterísticas do burst das oclusivas. Os resultados mostram que as caraterísticas do burst variam de acordo com o ponto de articulação. As frequências espetrais obtidas correspondem aos resultados publicados na literatura clássica. No entanto, verifica-se que outras medidas espetrais, no presente estudo, não estão de acordo com os padrões espetrais clássicos. É discutido se estas diferenças se devem ao uso de diferentes metodologias, ou a diferenças substanciais nas caraterísticas espetrais das diferentes línguas.

Palavras-chave: *Análise Multitaper; Picos e vales; Declives; Momentos.*

List of abbreviations used:

EP - European Portuguese
VOT - voice onset time
CV - consonant-vowel
PSD - power spectral density
SFS - Speech Filing System
IR - beginning of the release
$\overline{F}$ - average frequency

m1 - low-frequency slope
m2 - high frequency slope
M1 - mean
M2 - variance
M3 - skewness
M4 - kurtosis

1. INTRODUCTION

The spectral characteristics of stops have been studied in several languages (Blumstein & Stevens, 1979; Halle, Hughes, & Radley, 1957; Lahiri, Gewirth, & Blumstein, 1984; Sussman, Hoemeke, & Ahmed, 1993; Sussman, Hoemeke, & McCaffrey, 1992). However, the few previous studies (Andrade, 1980; Veloso, 1995; Viana, 1984) of stops in European Portuguese (EP), presented measures of voice onset time (VOT) and stop duration, but none examined the spectral characteristics of stops. Thus, the different spectral analysis methods used in this work, together with the novel examination of other stop characteristics, could help in the description of spectral characteristics of stops for any given language, and especially for EP.

It has long been known (Halle et al., 1957) that the spectrum of a stop burst varies with the place of articulation since the short noise burst is shaped by the given vocal tract properties defined by a particular articulatory configuration at burst time. The results reported by Halle et al. (1957) indicate that bilabial stops have a concentration of energy in the low frequency range (500-1500 Hz) whereas alveolar stops contain energy in the high frequency range (above 4000 Hz). In contrast, velar stops have the main energy distribution in the intermediate frequencies (between 1500 and 4000 Hz).

In the spectrum, peaks are caused by the poles of the vocal tract frequency response. Spectral troughs are originated by the zeros of the vocal tract frequency response. Peaks and troughs that can be seen in the spectra are related to underlying poles and zeros, resulting in an extremely complex transfer function. Displacement in any pole or zero frequency influences all peaks and troughs, sometimes significantly (Stevens, 1998), resulting in a complex pattern of spectral properties.

These patterns of the different spectral properties of the stop release were further analysed by Blumstein and Stevens (1979), who could show three distinct spectral shapes corresponding to either bilabial, alveolar or velar stops. *Bilabial* stops have a number of diffuse spectral peaks. Their peak amplitude has either more energy in the low frequencies than in the high frequencies (diffuse-falling pattern) or an even distribution throughout the complete spectrum (diffuse-flat pattern). In turn, the *alveolar* stops have a similar diffuse spread of energy peaks, but in contrast to the bilabials their peak amplitude is more prominent for high frequencies (diffuse-rising pattern). In contrast to bilabial and alveolar, *velar* stops show one distinguishable spectral peak, which normally occurs in the mid frequencies and dominates the complete spectrum (compact pattern). Thus, bilabial and alveolar stops share the property of diffuseness and they are only distinguished by the shape of the spectral energy distribution between low and high frequencies, whereas the velars can be identified by their mid frequency energy dominance.

Other spectral characteristics previously studied are the burst amplitude (Ohde & Stevens, 1983) and relative spectral change from burst onset to voicing onset (Lahiri et al., 1984). Lahiri et al. (1984) attempted a classification of stops in Malayalam, French and English. They concluded that the static spectral properties could not distinguish labial and dental stops because both had a diffuse-flat spectrum. Nevertheless, these stops could be discriminated by the ratio of the relative change measured in the low frequency range (up to 1500 Hz) and the high frequency range (above 3500 Hz) over the time interval from release to voice onset.

Forrest, Weismer, Milenkovic and Dougall (1988) showed that the simultaneous evaluation of the spectral moments mean, skewness, and kurtosis of the burst spectra robustly differentiated the place of stop articulation: [p] and [t] differed consistently in skewness and mean but not in kurtosis, while [k] was similar to [p] in mean and skewness but differed from other places of articulation in kurtosis.

Another theory widely used to differentiate the place of articulation is based on the locus equations. They describe the relation between the frequency of F2 of a vowel and the F2 measured at the onset of a consonant-vowel (CV) transition. Sussman, McCaffrey and Matthews (1991) used this method for identification of the place of articulation for stops. The authors determined locus equations for children learning American English (Sussman et al., 1992), adult speakers of Thai, Cairene Arabic, and Urdu (Sussman et al., 1993), adult speakers of American English (Sussman, 1994), adult speakers whose jaw was fixed by a bite block condition (Sussman, Fruchter, & Cable, 1995) and children with severe articulatory problems (Sussman, Fruchter, Hilbert, & Sirosh, 1998).

Hoelterhoff and Reetz (2007) examined bilabial and alveolar voiceless stops in German to obtain robust cues from the acoustic speech signal. They presented evidence that a logarithmic distance measure (log 10 (F2/F1)) and relative amplitude analysis of discrete frequency bands could robustly identify and distinguish the different places of articulation. Further, the results of the relative amplitude calculation showed that in initial and medial word position the bilabial stops had significant smaller relative amplitude in the frequency bands between 2-3 kHz and 7-8 kHz when compared to alveolar stops.

In the present study we aim to examine the spectral properties of /p, b, t, d, k, g/ with respect to place of articulation, produced in initial, medial and final word position. The principal objective of this work is to explore recently developed methods like multitaper techniques (Blacklock, 2004) for phonetic research and to use these new methods to investigate a Romance language. The broader research aim is to examine the properties of EP stops and fricatives and compare them to other languages like the well-studied English language, or languages with additional stop contrasts like Korean or Hindi.

We focus on the spectral properties of the burst since recent research (Mooshammer, Hoole, & Geumann, 2006) shows very promising articulatory and perceptual cues in the burst. Mooshammer et al. (2006) found a very stable difference in jaw coordination between different stops and sibilants. These differences could result from the need to produce a stable and salient burst to contrast voiced from voiceless consonants. These would then be encoded in the spectral properties of the burst.

From a perceptual perspective, Francis, Baldwin and Nussbaum (2000) showed that listeners could be trained to give more perceptual weight to a burst spectrum cue than to formant transitions. Thus, examining the burst properties with new methods could give a new insight into both articulatory coordination and perceptual cue weighting.

2. METHOD

We recorded a corpus of 54 European Portuguese real words containing the six stops /p, b, t, d, k, g/ using a Philips SBC ME 400 unidirectional condenser microphone located 0.2 m in front of the subject's mouth. The acoustic signal was recorded by means of a Sony PCM-R300 DAT recorder with 16 bits and a sampling frequency of 48 kHz.

The corpus contained an equal number of 18 words with stops in the three positions: initial position, followed by the vowels /a/, /i/ and /u/; medial position, preceded by the vowels /a/, /i/ and /u/ and followed by the vowel /ɐ/; final position, preceded by the vowels /ɔ/ and /a/. The words were produced in isolation (Corpus 1) and within the frame sentence "Diga... por favor." (Corpus 2). All words of Corpus 1 and 2 are listed in Appendix. The recorded subjects were six native speakers of EP: three men LJ, HR and PA (aged 25 to 34) and three women ML, IM and SC (aged 24 to 42), all without any speech, language or hearing disorders. LJ, HR, ML and IM lived in Aveiro, PA and SC lived in Porto. Thus, we collected a total number of six stops in three word positions, for three vowel contexts (two vowel contexts in word final position), and produced by six speakers both in isolation and in a frame sentence.

First, we analysed the words produced in isolation by speakers ML and LJ (Corpus 1). Then, after manually segmenting Corpus 2, we analysed the words produced by all six speakers in the frame sentence. All items were analysed using the *Speech Filing System* (*SFS*) (Huckvale et al., 1987). The beginning of the release (IR) was defined by a sudden peak in the waveform and the corresponding vertical bar in the spectrogram. When multiple bursts were found we chose the one with the highest amplitude (Fuchs, 2005) (see Figure 1).

The existing large variance in noise parts of the speech signal normally requires a massive variance reduction, but desirably without losing the distinctive properties of the examined phone. However, most variance reduction methods require certain assumptions to be met on the audio signal, mostly that the signal is regarded stationary and ergodic (i.e., the statistical properties of the process are independent of sample sequence). However, noise parts of the speech signal are neither stationary nor ergodic (Percival & Walden, 1993), which invalidates the majority of the data variance methods: Ensemble averaging techniques can only be used if the signal is ergodic; time-averaging over the phone duration assumes stationarity and thus the signal not changing from the onset until the offset, which is clearly not the case of speech signals (articulators are constantly on the move from one phone to the other).

Therefore, both time-averaging and ensemble averaging techniques cannot be used for short (stationary) periods of a given fricative or stop signal. One could also use the standard frequency smoothing techniques to reduce the variability. However, these algorithms use standard windowing techniques to smooth the spectra. The windows used invariably lose important data towards the edges (due to the necessary sidelobe suppression), which in turn leads to an undesirable increase in the variance of the spectral estimate.

This problem of losing data at the window edges is solved by the multitaper technique (Slepian, 1978; Thomson, 2000), which applies consecutive orthogonal windows (Slepian tapers: prolate speroidal windows): the second window is set orthogonally to the first window and so on. This technique guarantees that data is not lost towards the ends of each window. At the end of the process, the resulting smoothed spectrum is obtained by summing all orthogonal estimates. The resulting spectrum exhibit reduced variance, which is no longer dependent on the underlying spectral magnitude. Thus, in contrast to standard spectral estimation techniques, multitaper analysis provides an optimal way to reduce the bias of the spectral estimates when calculated over only short intervals of the data, and is thus highly suited to examine stochastic parts of the speech signal (see Blacklock (2004) for further explanations). The noise part of the burst is a stochastic signal like fricatives or frication parts of other phonemes, thus the main advantage of the multitaper (the accurate analysis of a single short time window) is thus well-suited for the characteristics of the very short stop burst.

Considering these advantages, we calculated multitaper spectra with 11ms windows (512 point Hamming window) aligned to the left-side to the stop release (see Figure 1). We used the power spectral density (PSD) estimate via the Thomson multitaper method (linear combination with unity weights of individual spectral estimates and the default FFT length) available in the Mathworks Signal Processing Toolbox Version 6.2. (MathWorks, 2007: 470-475).

Using this method, we first analysed the spectra considering the observed troughs, peaks and broad-peaks as shown in Figure 1. We manually analysed each spectral peak and trough for all words produced by speakers ML and LJ (Corpus 1).

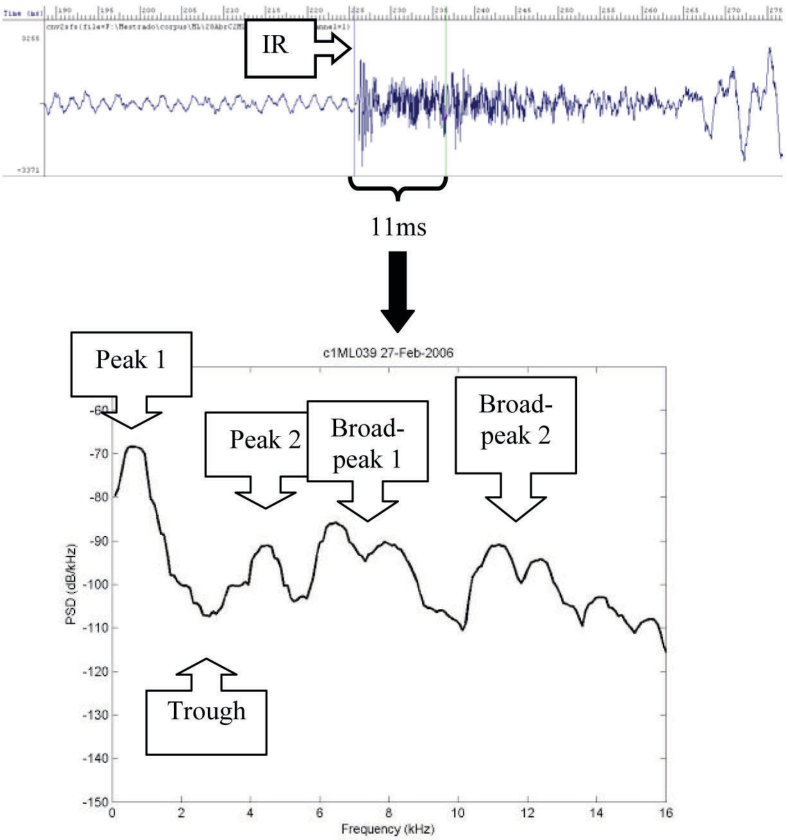


Figure 1: Waveform and spectrum of stop [k] produced by speaker ML. The beginning of the release (IR) is shown in the waveform and the corresponding peaks and troughs are shown in the spectrum.

For each spectrum we computed the average frequency ($\overline{F}$) at which the spectral amplitude had a maximum for the frequency range from 0Hz to 16000Hz (Jesus & Shadle, 2002: 447). This measure provides an endpoint and start point for two line fits (linear regression using *Matlab's polyfit* function) used to calculate the slopes m1 and m2 of the corresponding spectrum (see Figure 2).

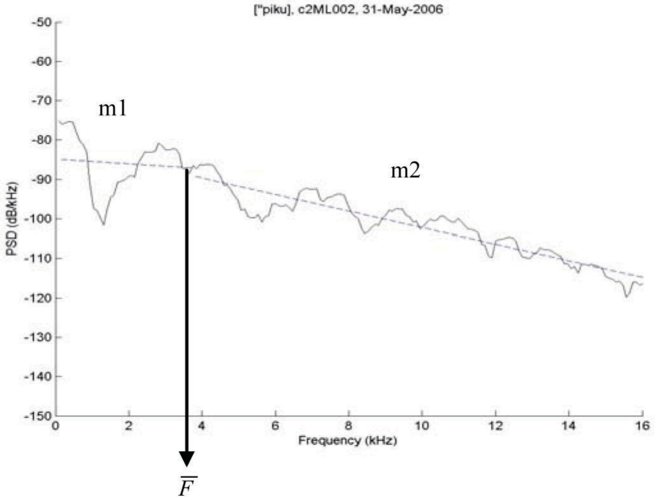
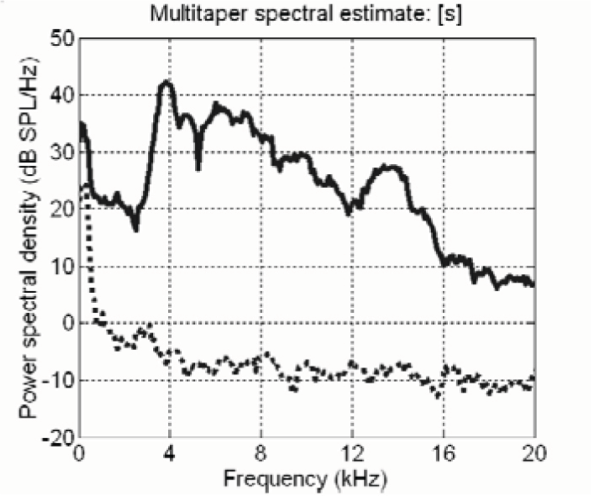


Figure 2: Multitaper spectra and regression lines used to calculate low-frequency slope (m1) and high frequency slope (m2), for [p] produced in initial position by speaker ML.

This analysis was motivated by Blumstein and Stevens's (1979) study and based in Jesus and Shadle's (2002) methodology. Jesus and Shadle (2002) studied EP fricative consonants and showed that the slope (m2) is related to the noise source strength. In the present study, we aim to adapt this technique to investigate the spectral patterns suggested by Blumstein and Stevens (1979) for different places of stop articulation.

Further, we normalised the PSD of each stop and calculated the moments that describe the underlying distribution, which can be reconstructed via the Gram-Charlier expansion (Blacklock, 2004: 44-46), as shown in Figure 3.



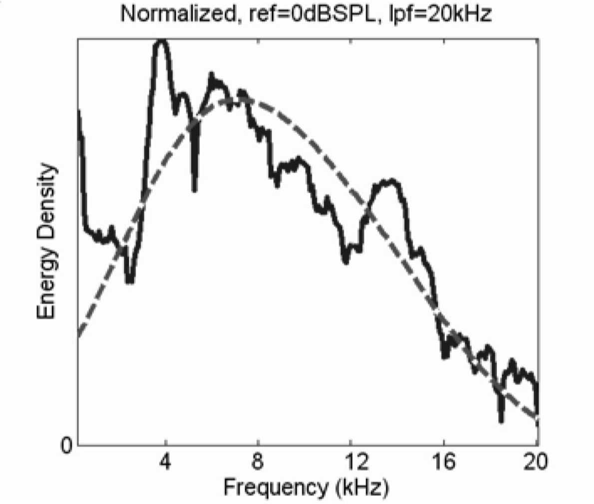


Figure 3: Multitaper (upper) and normalised spectra (lower) in solid lines. The dashed line represents the Gram-Charlier distribution. The Gram-Charlier distribution describes how accurately the computed spectral moments match the spectra from which they were derived. From Blacklock (2004).

Although the spectra shown in Figure 3 correspond to fricative turbulence noise, the same spectral analysis techniques can be applied to the phase of the transient burst noise of stops. Thus, on the basis of the normalised PSD we calculated the following four spectral moments according to Forrest et al. (1988): M1 (mean), M2 (variance), M3 (skewness) and M4 (kurtosis). M1 indicates the central point of the distribution, M2 indicates the spreading around the expected value, M3 shows if the distribution is skewed towards onset or offset, whereas M4 can be seen as a measure of the "peakedness" of the spectral distribution.

3. RESULTS

3.1. General distribution of peaks and troughs

Our results showed (see Table 1) that the type and the number of the spectral peaks and troughs are similar for stops with the same place of articulation (Figure 4 shows the method we used to register the different frequencies of troughs and peaks). The stops [p, b] had one trough and one broad-peak ($\overline{F} = 3.7$ kHz). Most [t, d] samples had one trough, two peaks and one or two broad-peaks ($\overline{F} = 3.9$ kHz). Most [k, g] samples had one trough, two peaks and two broad-peaks ($\overline{F} = 4.6$ kHz).

Table 1: Results of spectral peak and trough frequencies of the examined stops

	Trough (kHz)	Peak 1 (kHz)	Peak 2 (kHz)	Broad-peak 1 (kHz)	Broad-peak 2 (kHz)
[p]	0.8-4.6			1.4-5.6	
[b]	0.7-5.0			1.5-5.6	
[t]	1.7-7.0	0.3-3.7	2.8-4.6	6.0-10.4	
[d]	1.5-5.5	0.3-1.7	2.4-5.6	5.2-9.9	11.0-12.8
[k]	2.9-6.3	0.6-3.8	3.9-5.4	7.1-9.4	10.0-13.2
[g]	0.8-7.4	1.0-2.6	3.9-4.9	6.8-9.1	12.0-13.6

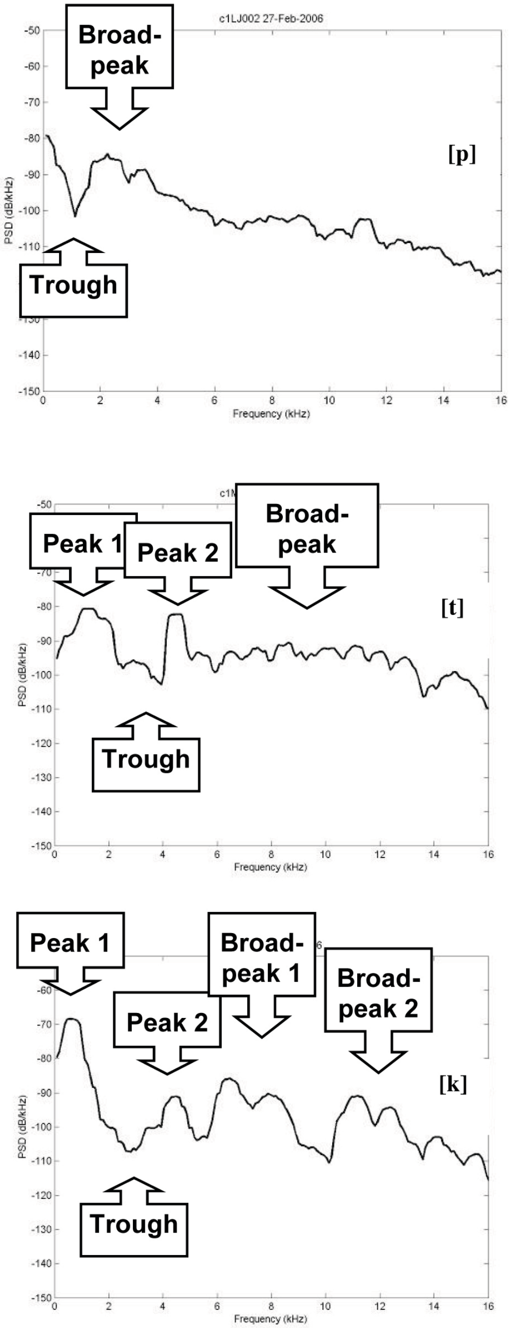


Figure 4: Method used to register the peaks and troughs of the spectra.

Further, also the frequency range of the spectral peaks and troughs was similar for stops with the same place of articulation. It can be seen in Figure 5 that our data corresponds to the measures of Table 1. Figure 5 shows all spectra of Corpus 1 with the overlaid mean for the speakers LJ and ML.

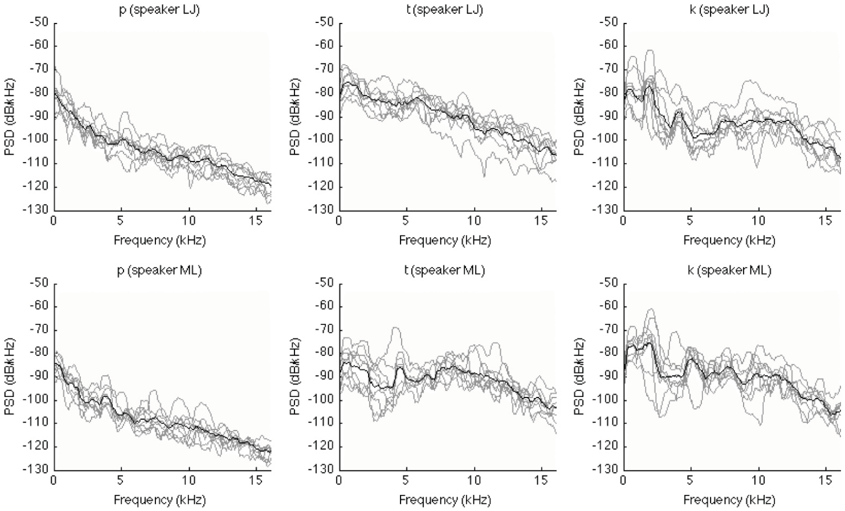


Figure 5: Plots of all spectra of [p, t, k] in Corpus1. The spectra of all examined stops are shown in grey. The mean of the spectra is overlaid in black.

3.2. Spectral slopes m1 and m2

Figure 6 shows scatterplots for the numerical values of the slopes m1 and m2, and Table 2 shows the means and standard deviations when comparing all examined stops. It can be seen that the values of m1 and m2 do not differentiate voiced from voiceless stops. Nevertheless, the distribution of values for the different stops in the scatterplot suggests that place of articulation strongly influences the values for the spectral slopes. First, the slope m1 generally is steeper for [p] than for [t], while m2 is approximately the same for both stops. Second, it can be observed that for m2 stop [k] differs from both [p] and [t]. Third, m2 is lower for [g] than for [b], while m1 has the same values. Finally, stop [d] shows values for both slopes similar to [b] and [g].

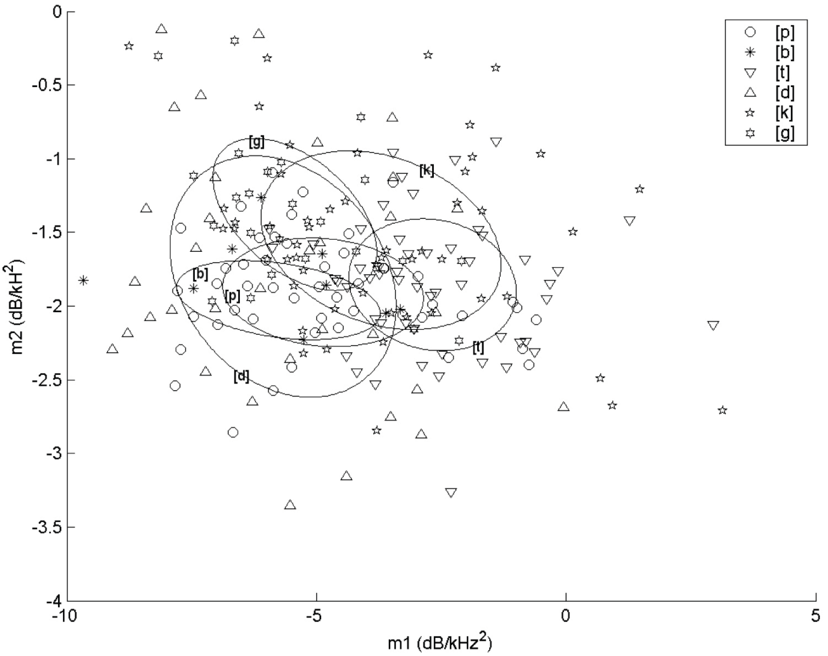


Figure 6: Relation between slopes (m1 versus m2) of all stops produced by different speakers. Variance ellipses (centred in the mean) are also shown for each stop.

When looking at the divergence ellipses (m1 versus m2) we noticed that they were too variant and without any apparent separation between the different stops. So, without further reduction techniques, it was not possible to investigate this divergence.

Table 2: The mean and standard deviations (std) over all speakers are given for m1 and m2, separately for all stops (voiced and unvoiced)

	m1		m2	
	mean	std	mean	std
[p]	-4.87	2.01	-1.91	0.37
[b]	-5.77	2.04	-1.96	0.27
[t]	-2.68	1.68	-1.86	0.45
[d]	-5.67	2.26	-1.80	0.82
[k]	-3.71	2.40	-1.55	0.61
[g]	-5.43	1.64	-1.38	0.52

With the aim to classify the different shapes suggested by Blumstein and Stevens (1979), we analysed the voiceless stops in all word positions from Corpus 2. This analysis was performed by examining the spectral patterns, which were superimposed by the line fits m1 and m2. However, in our data we did not succeed to observe the general reported patterns, since our spectral patterns differed with adjacent vowel context.

When followed by low vowel [a], the voiceless bilabial stops in initial position had spectra with steeper negative slopes than velars for high frequencies (see Figure 7).

When followed by the high vowel [i], bilabials were mostly flat, whereas velars had positive slopes for low frequencies. In contrast, when followed by the vowel [u], bilabials and velars generally showed a negative slope for low frequencies (see Figure 7). Dentals showed positive, sometimes flat slopes (for low frequencies) when followed by the high back vowel [u].

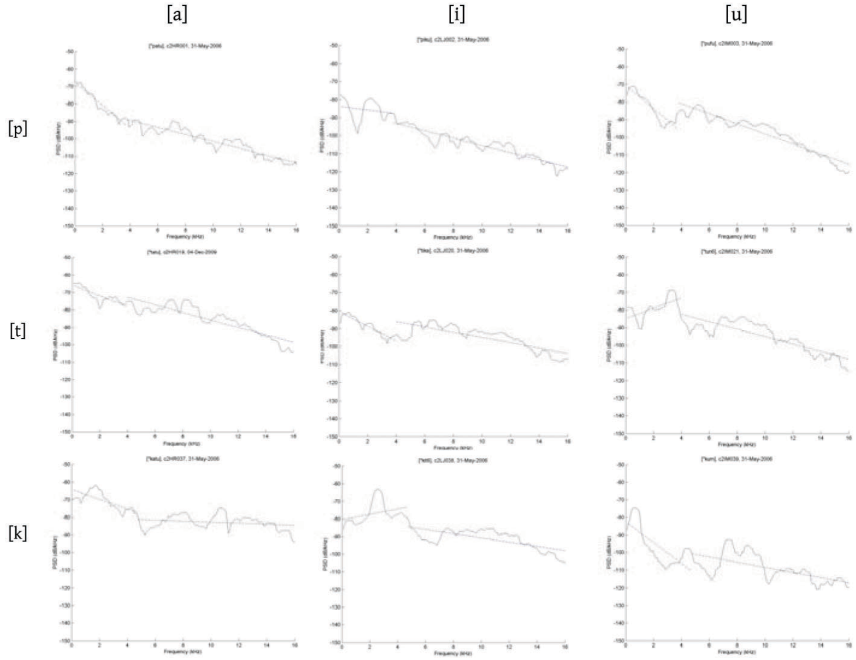


Figure 7: Multitaper spectra and regression line fits for bilabial, dental and velar stops in initial position.

3.3. Spectral moments

3.3.1. Voiceless stops in initial position and [a] vowel context: The analysis of our spectral moment data did not allow us to distinguish the three voiced stops [b, d, g]. Further, for most of our data the voiced stop [b] was produced without a release (resulting in missing data for spectral moments), so there was no sufficient data for the analysis. For these reasons, in Figure 8 we show the spectral moments for all voiceless stops in initial position for all speakers in the vowel context [a] to eliminate the added variability related to vowel context. Figure 8 shows that [p] and [t] exhibited different skewness values with positive values for [p] and negative values for [t] (except for speakers PA and SC). The stops [p] and [t] had different M1 (mean) values, except for speakers PA and SC.

Stops [k] and [p] differed in the values for skewness for all speakers. Stop [k] had negative values of skewness except for speakers PA and IM and [p] had positive values of skewness. Figure 8 also shows that [k], [p] and [t] did not differ in kurtosis and [p] and [t] in M2 (variance) except for speaker IM.

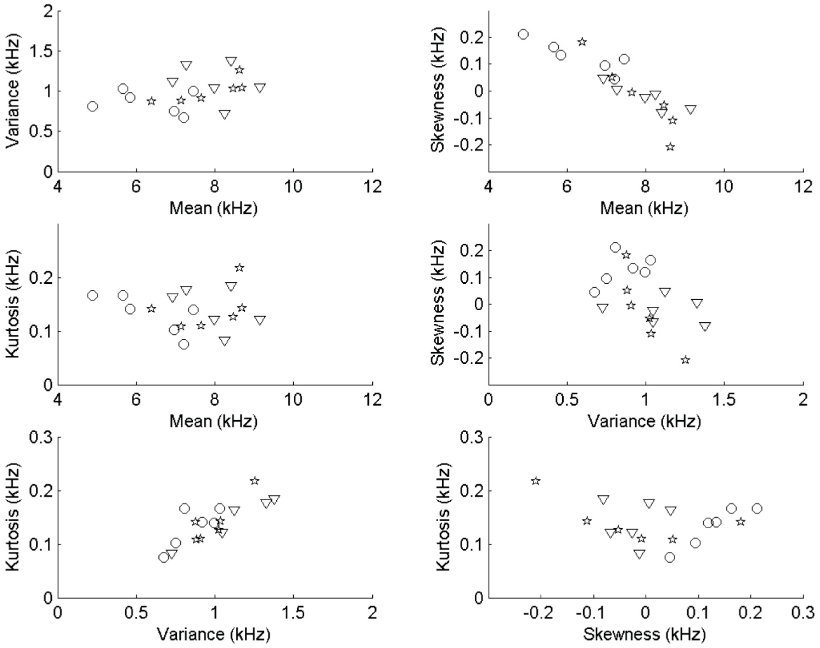


Figure 8: Moments of voiceless stops in initial position followed by vowel [a] for all speakers. [p] – ○; [t] – ▽; [k] – ☆.

3.3.2. Generalisation for all vowel contexts and word positions: In the following, we aim to generalise the above results for all vowel contexts for the examined moments.

Skewness: Stops [p] and [t] had different values of skewness for all speakers. Stop [p] presented values of skewness close to zero (almost symmetric distribution), or positive values, which suggested positive asymmetric distribution, so the highest spectral amplitude values were left-concentrated (see Figure 9). In contrast, [t] usually exhibited skewness values close to zero (almost symmetric distribution), or negative values (negative asymmetric distribution) indicating high spectral amplitude values concentrated to the right (see Figure 9).

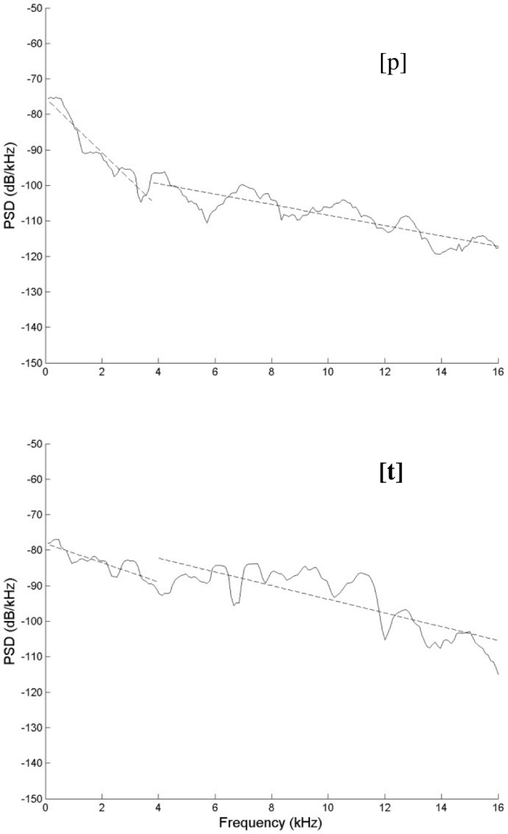


Figure 9: Spectrum of bilabial (top) and dental (bottom) voiceless stops produced by speaker IM to show spectral differences (left-concentrated versus right-concentrated).

The values of [k] and [p] were approximately the same for the speakers SC (see Figure 10 for an example), PA and IM (however, plots for all speakers in all contexts will not be presented due to the large amount of data examined).

Mean, kurtosis and variance: For the moment mean (M1), the stops [p] and [t] had different values except for speaker PA. Stop [p] showed lower means than [t] which suggests that the central point of the distribution was lower for [p] than for [t].

The Kurtosis values show the spread of energy and are therefore related to stop strength. In our data, the results showed that all voiceless stops [p], [t] and [k] had approximately the same kurtosis. This means that there was no apparent difference of stop strength for the data we analysed.

The moment variance (M2) for stops [p] and [t] had approximately the same values for all speakers.

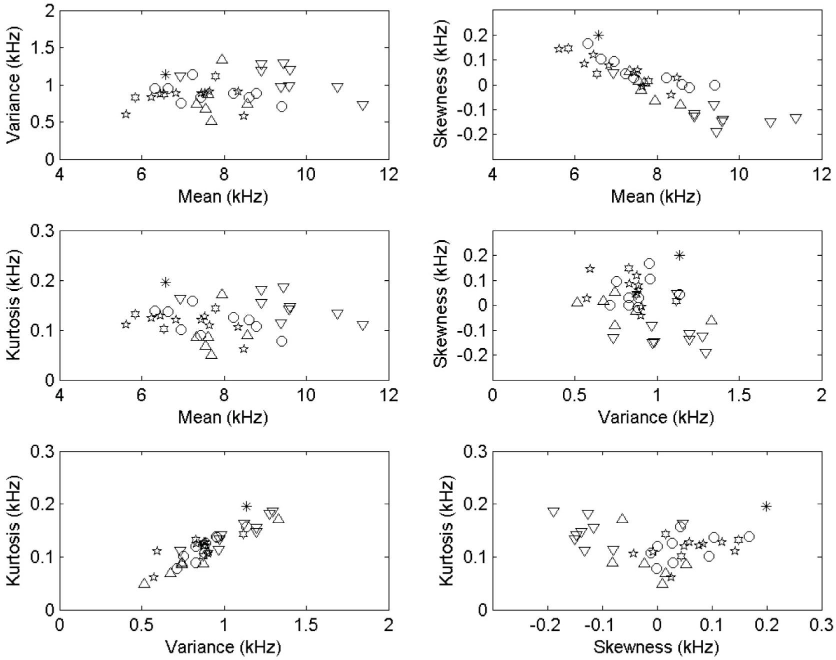


Figure 10: Moments of stops produced by speaker SC.
[p] – ○; [b] – ✳; [t] – ▽; [d] – △; [k] – ☆; [g] – ✡.

4. Discussion and conclusions

In this study, we examined spectral characteristics correlated with place of articulation, based on a corpus that included EP stops in different word positions. Our principal findings are as follows.

4.1. General distribution of peaks and troughs

The examination of spectral peak and trough frequencies showed that, as expected, the burst characteristics vary with place of articulation. The type, number and frequency range of spectral peaks and troughs were similar for stops with the same place of articulation. The spectral peak frequencies lie in the range proposed by Halle et al. (1957) for American English and correspond to the data of Blumstein and Stevens (1979). Unfortunately, to our knowledge there is no comparable EP data available, so we cannot provide comparisons with other studies.

4.2. Spectral moments

When analysing the spectral moments, we could show that bilabial and dental stops differ in terms of mean and skewness, which supports previous results for American English (Forrest et al., 1988). This indicates that mean and skewness in fact could differentiate bilabial from dental stops. In contrast, our analyses show that the kurtosis does not differ with varying place of articulation. This result differs from previous results for American English (Forrest et al., 1988), and could therefore be an important indicator for EP. The reason for this difference could be one of the following: cross-linguistic differences between American English and EP could be the underlying reason for the different result; due to methodological differences between Forrest et al. (1988) and our study (position of stops, vowel context and spectral averaging technique) different results are obtained.

4.3. Parameterisation of the stop spectra

Our approach to parameterise the stop spectra revealed that the slopes differ with varying place of articulation. Further, we could show that the characteristics of adjacent vowels strongly influence the spectral properties of the burst. However, in contrast to the classical observation of special spectral patterns for each place of articulation as reported by Blumstein and Stevens (1979), we were not able to observe these patterns in our data. The probable reasons for this could be twofold.

First, the differences could be accounted for by methodological differences, since the spectra suggested by Blumstein and Stevens (1979) were obtained with standard FFTs, as opposed to our multitaper approach, and with a limited frequency range up to 5 kHz whereas our spectra were computed with the full frequency range up to 16 kHz. Naturally, both the difference in algorithm and frequency cut-off will influence to a certain amount the spectral Gestalt and therefore lead to differences in the spectral shape and the corresponding linear fits. However, since the same frequency region (0Hz to 5000Hz) analysed by Blumstein and Stevens (1979) is included in our spectra, it is unlikely that the reported differences are only due to methodology.

We propose a complementary explanation. It is widely accepted that different context influences the articulatory target of *place of articulation*, i.e., *place of articulation* for stops and fricatives is highly dependent on coarticulatory effects. These effects have been found in Electro-Magnetic Midsagittal Articulometry (EMMA) studies for German (Geng, 2008) and French (Perrier, Payan, Zandipour, & Perkell, 2003) and are reported to be universal. It is logical to assume that a difference in *place of articulation* also triggers differences in the burst spectra. Thus, context in fact would influence the spectral properties of the stop burst, as we have found in our study. However, this effect of corarticulation is not very well-studied for acoustic cues in burst spectra (Li, Menon, & Allen, 2010; Smits, Bosch, & Collier, 1996), so further investigations for various languages have to be conducted.

A second reason for the difference of our patterns and the observed patterns of Blumstein and Stevens (1979) could be in fact the cross-linguistic differences in articulation. While in EP the coronal stops are dental (Cruz-Ferreira, 1999), they are alveolar in American English (Ladefoged & Johnson, 2010), thus this difference in *place of articulation* should, according to what we know about the acoustic theory of speech production (Stevens, 1998), affect the spectral characteristics of the stops and therefore result in the observed divergence (Lahiri et al., 1984).

Since this reason would explain only the differences for the alveolar stops, for bilabials and velars one would assume the same place of articulation for the two languages (American English and EP). However, since there is no articulatory data available for EP, up to this point we cannot comment on this conclusion.

As an alternative to the traditional spectral algorithms used throughout the available literature, we propose in this paper the multitaper technique for the spectral estimation of the stop burst. It is well suited to accurately display the spectral characteristics of the very short stop burst phase. A short 11ms window is sufficient to obtain the accurate spectral information, and the technique is ideal to represent the stochastic nature of the short burst, since it is neither stationary nor continuous, an assumption other spectral estimation algorithms require.

ACKNOWLEDGEMENTS

This work was developed during the M.Sc. in Speech and Hearing Sciences at the University of Aveiro, Portugal.

This work was partially funded by FEDER through the Operational Program Competitiveness Factors - COMPETE and by National Funds through FCT - Fo undation for Science and Technology in the context of the project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011) and the post-doctoral fellowship SFRH/BPD/48002/2008.

Recebido em novembro de 2010
Aprovado em novembro de 2011

E-mails: marisalousada@ua.pt
lmtj@ua.pt
danielpape@ua.pt

REFERENCES

ANDRADE, A. (1980). *Estudos experimentais aerodinâmicos, acústicos e palatográficos do vozeamento nas consoantes*. Lisboa, Portugal: Centro de Linguística da Universidade de Lisboa.

BLACKLOCK, O. 2004. *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods.* Ph.D. Thesis, University of Southampton, Southampton, UK.

BLUMSTEIN, S., & K. STEVENS. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America, 66*: 1001-1017.

CRUZ-FERREIRA, M. 1999. Portuguese (European). In IPA. Ed. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (pp. 126-130). Cambridge: Cambridge University Press.

FORREST, K., G. WEISMER, P. MILENKOVIC, & R. DOUGALL. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84*: 115-123.

FRANCIS, A., K. BALDWIN, & H. NUSSBAUM. 2000. Effects of training on attention to acoustic cues. *Perception and Psychophysics, 62*: 1668-1680.

FUCHS, S. 2005. *Articulatory correlates of the voicing contrast in alveolar obstruent production in German.* Ph.D. Thesis. Queen Margaret University College, Edinburgh, UK.

GENG, C. 2008. *A cross-linguistic study on the phonetics of dorsal obstruents: Experimental investigations.* Suedwestdeutscher Verlag fuer Hochschulschriften.

HALLE, M., G. HUGHES, & J. RADLEY. 1957. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29*: 107-116.

HOELTERHOFF, J., & H. REETZ. 2007. Acoustic cues discriminating German obstruents in place and manner of articulation. *Journal of the Acoustical Society of America, 121*: 1142-1156.

HUCKVALE, M., D. M. BROOKES, L. DWORKIN, M. JOHNSON, D. PEARCE, & L. WHITAKER. (1987). *The SPAR Speech Filing System.* Paper presented at the European Conference on Speech Technology, Edinburgh.

JESUS, L. M. T., & C. H. SHADLE. 2002. A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics 30*(3): 437-464.

LADEFOGED, P., & K. JOHNSON. 2010. *A course in phonetics.* (6th ed.) Boston: Thomson Wadsworth.

LAHIRI, A., L. GEWIRTH, & S. BLUMSTEIN. 1984. A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study. *Journal of the Acoustical Society of America, 76*: 391-403.

Li, F., A. Menon, & J. Allen. 2010. A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *Journal of the Acoustical Society of America, 127*: 2599-2610.

MathWorks. 2007. Signal Processing Toolbox 6 User's Guide. Natick: MathWorks.

Mooshammer, C., P. Hoole, & A. Geumann. 2006. Interarticulator cohesion within coronal consonant production. *Journal of the Acoustical Society of America, 120*: 1028-1039.

Ohde, R., & K. Stevens. 1983. Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America, 74*: 706-714.

Percival, D., & A. Walden. 1993. *Spectral analysis for physical applications: Multitaper and conventional univariate techniques.* Cambridge: Cambridge University Press.

Perrier, P., Y. Payan, M. Zandipour, & J. Perkell. 2003. Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America, 114*: 1582-1599.

Slepian, D. 1978. Prolate spheroidal wave functions, fourier analysis, and uncertainty V: the discrete case. *Bell System Technical Journal, 43*: 3009-3057.

Smits, R., L. t. Bosch, & R. Collier. 1996. Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. II. Modeling and evaluation. *Journal of the Acoustical Society of America, 100*: 3865-3881.

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

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

Sussman, H., D. Fruchter, & A. Cable. 1995. Locus equations derived from compensatory articulation. *Journal of the Acoustical Society of America, 97*: 3112-3124.

Sussman, H., D. Fruchter, J. Hilbert, & J. Sirosh. 1998. Linear correlates in the speech signal: The orderly output constraint. *Behavioral and Brain Sciences, 21*: 241-299.

Sussman, H., K. Hoemeke, & F. Ahmed. 1993. A cross-linguistic investigation of locus equations as a relationally invariant descriptor for place of articulation. *Journal of the Acoustical Society of America, 94*: 1256-1268.

SUSSMAN, H., K. HOEMEKE, & H. MCCAFFREY. 1992. Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech and Hearing Research, 35*: 769-781.

SUSSMAN, H., H. MCCAFFREY, & S. MATTHEWS. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America, 90*: 1309-1325.

THOMSON, D. 2000. Multitaper analysis of nonstationary and nonlinear time series data. In W. Fitzgerald, R. Smith, A. Walden & P. Young. Eds. *Nonlinear and Nonstationary Signal Processing* (pp. 317-394). Cambridge: Cambridge University Press.

VELOSO, J. 1995. *Aspectos da percepção das "oclusivas fricatizadas" do Português: Contributo para a compreensão do processamento de contrastes alofónicos.* M.Sc. Thesis. Universidade do Porto, Porto, Portugal.

VIANA, M. C. 1984. *Étude de deux aspects du consonantisme du portugais: fricatisation et devoisement.* Ph.D. Thesis. U. Sciences Humaines de Strasbourg, Strasbourg, France.

Appendix

Table 3: Corpora used for the experiments. The International Phonetic Association (IPA) version was adopted from the illustration proposed by Cruz-Ferreira (1999) for European Portuguese.

Table 3a: Words of Corpus 1 and 2 with stops /p, b/ in initial, medial and final position.

Stop	Position	Word	IPA
/p/	Initial	pato	['patu]
		pico	['piku]
		pufo	['pufu]
	Medial	napa	['napɐ]
		ripa	['ripɐ]
		lupa	['lupɐ]
	Final	top	['tɔp]
		pape	['pap]
		tape	['tap]
/b/	Initial	bato	['batu]
		bico	['biku]
		bufo	['bufu]
	Medial	naba	['nabɐ]
		chiba	['ʃibɐ]
		juba	['ʒubɐ]
	Final	sobe	['sɔb]
		sabe	['sab]
		cabe	['kab]

Table 3b: Words of Corpus 1 and 2 with stops /t, d/ in initial, medial and final position.

Stop	Position	Word	IPA
/t/	Initial	tacto	['tatu]
		tica	['tikɐ]
		tuna	['tunɐ]
	Medial	nata	['natɐ]
		Rita	['ritɐ]
		luta	['lutɐ]
	Final	pote	['pɔt]
		bate	['bat]
		date	['dat]
/d/	Initial	dato	['datu]
		dica	['dikɐ]
		duna	['dunɐ]
	Medial	nada	['nadɐ]
		vida	['vidɐ]
		buda	['budɐ]
	Final	pode	['pɔd]
		nade	['nad]
		jade	['ʒad]

Table 3c: Words of Corpus 1 and 2 with stops /k, ɡ/ in initial, medial and final position.

Stop	Position	Word	IPA
/k/	Initial	cacto	['katu]
		quita	['kitɐ]
		cume	['kum]
	Medial	vaca	['vakɐ]
		pica	['pikɐ]
		nuca	['nukɐ]
	Final	Roque	['rɔk]
		saque	['sak]
		taque	['tak]
/ɡ/	Initial	gato	['ɡatu]
		guita	['ɡitɐ]
		gume	['ɡum]
	Medial	vaga	['vaɡɐ]
		viga	['viɡɐ]
		guga	['ɡuɡɐ]
	Final	rogue	['rɔɡ]
		pague	['paɡ]
		vague	['vaɡ]