ISCA Archive
http://www.isca-speech.org/archive

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

ACOUSTIC AND PERCEPTUAL CHARACTERISTICS OF THE SPANISH FRICATIVES

Sergio Feijóo, Santiago Fernández, Nieves Barros and Ramón Balsa
Departamento de Física Aplicada
Universidad de Santiago de Compostela
15706 Santiago de Compostela, Spain
fasergio@usc.es

ABSTRACT

This study deals with the relation between the spectral representation and the perceptual identification of the Spanish fricatives and affricates. Several spectral representations have been analyzed: FFT-derived linear cepstrum, mel cepstrum coefficients, LPC cepstral coefficients and the first four statistical moments. Quadratic discriminant analysis including the leave-one-out method have been carried out on a large database. For this particular classification procedure, both the order of every spectral parametrization and the order of the temporal trajectory of those parameters have been optimized. The results indicate that a low order representation performs satisfactorily and that a three order temporal trajectory is adequate to encode the dynamics of the fricatives. The best classification rates were obtained by the cepstral (79.5%) and linear cepstrum coefficients (75.2%). They attained a correlation coefficient with respect to the perceptual identification of 0.78 and 0.75, respectively.

Keywords: fricative, affricate, acoustic characterization, perceptual identification, spanish.

10.21437/Eurospeech.1999-382

1. INTRODUCTION

Fricatives are produced with a narrow constriction in the vocal tract. The flow of air through this constriction results into turbulent flow whose random pressure fluctuations generate the fricative sound. The size and shape of the cavity in front of that constriction determine the shape of the spectrum of the fricative noise. The spectral properties of the fricative noise are considered the primary cue for the identification of place of articulation in fricatives.

Due to their turbulent nature it is not easy to select a set of representative spectral characteristics of the fricative noise. A gross spectral representation like statistical spectral moments has been widely used to characterize fricatives [1, 2, 3]. Nevertheless, its success has been limited. Other more complete spectral representations could be appropriate for the characterization of fricative noise, namely FFT-derived linear cepstrum or cepstral coefficients. According to some authors [2] the inability of the LPC method to model zeros, could make it an inappropriate method for characterizing fricatives. Other spectral representations, based on global spectral shape could not adequately represent some detailed spectral characteristics. On the other hand, since the spectral peaks of fricatives are much broader than those of vowels, a low order spectral representation would seem more adequate.

The objective of this paper is to compare the performance of several spectral representations in the automatic classification of fricatives and affricates. The spectral representations considered included the FFT-derived linear cepstrum, mel cepstral coefficients, LPC cepstral coefficients and the first four statistical moments. The question of the appropriate order of the representation was also investigated.

2. MATERIALS

The tokens were bisyllabic natural Spanish words stressed in the first syllable. The syllables were formed by the combination of one of the Spanish voiceless fricatives /θ, f, s, ʃ, x/ or the affricate /tʃ/ with one of the five vowels /a, e, i, o, u/. The affricate /tʃ/ was included due to its turbulence nature and its spectral similarity with fricative /ʃ/. The words were pronounced by ten men and ten women. The total number of stimuli was 600 tokens = 6 fricatives × 5 vowels × 10 speakers × 2 sexes. All the tokens were recorded in a normal office at the Faculty of Physics with a Rion microphone (type UC-53A) and sampled at 20 kHz using a DT-2801-A card with 12 bits of precision. They were band pass filtered with cutoff frequencies of 100 Hz and 9.2 kHz and normalized with respect to their maximum amplitude value.

Fricative noises were isolated by means of auditory, visual and spectral inspection. In case of doubt, the end of the fricative was considered as the point previous to the rise of the second formant.

3. PERCEPTUAL EXPERIMENT

A perceptual experiment was carried out in order to study the identification of the place of articulation of the Spanish fricatives from the isolated fricative noise. Eleven native speakers of Spanish (Galician region) served as listeners. They were instructed in the nature of the experiment and became familiar with the sound of the isolated fricative noises. Tokens were presented randomly to the subjects through SONY MDR-CD570 headphones, one repetition being allowed. The whole process was controlled by a computer program

	/θ/	/f/	/s/	/ʃ/	/x/	/tʃ/	*other*
/θ/	**56.5**	38.2	2.4	0.3	1.6	0.7	0.3
/f/	19.8	**78.4**	0.2	0.0	1.5	0.0	0.1
/s/	1.7	0.4	**80.8**	15.6	0.0	1.4	0.0
/ʃ/	0.4	0.2	16.8	**78.4**	0.0	4.0	0.2
/x/	0.1	0.4	0.0	0.0	**99.5**	0.0	0.0
/tʃ/	0.3	0.0	0.7	1.7	0.0	**97.3**	0.0

Table 1. Confusion matrix for the fricative noise stimuli. Mean classification rate: 81.8 %.

developed at our laboratory. The possible options for the open test were: the fricatives /θ, f, s, ʃ, x/, the affricate /tʃ/ or *other*. This last option was included to reduce the amount of guessing.

Table 1 shows the confusion matrix for the perceptual experiment. Isolated fricative noises were satisfactorily identified except /θ/, the mean correct classification rate being 81.8 % (56.5 % for /θ/, 78.4 % for /f/, 80.8 % for /s/, 78.4 % for /ʃ/, 99.5 % for /x/ and 97.3 % for /tʃ/). Confusions were more important for the pairs /θ/-/f/ and /s/-/ʃ/, while /x/ and /tʃ/ are almost perfectly identified.

4. ACOUSTIC ANALYSIS

Fricative spectra vary a great deal from fricative onset until fricative offset [2]. Thus, for the acoustic analysis the fricative noises were represented by three windows of 25.6 ms: one initial, one in the middle and one at the end of the fricative. The characterization of the spectral properties of the fricatives were carried out computing various spectral representations for each window: 24 LPC-derived cepstral coefficients, 20 cepstrum coefficients [4], 20 mel-cepstrum coefficients [4] and 4 spectral moments [3]. Thus, the fricative noises were characterized by vectors of 3 windows × N coefficients, where N is the number of coefficients computed for a particular spectral representation.

In order to assess the usefulness of those parameters a classification procedure was applied to every spectral representation. The analysis selected was the Quadratic Discriminant including the Leave-One-Out method. It involves forming quadratic combinations of the independent variables to determine the classification group for each case. Therefore, it is assumed that the covariance matrices for each group are not equal. From the classification space it is posible to compute for each case *a posteriori* probabilities (APP) of membership in each group. The Leave-One-Out method implies that the classification space is computed with every case in the database except the case that is being classified. In

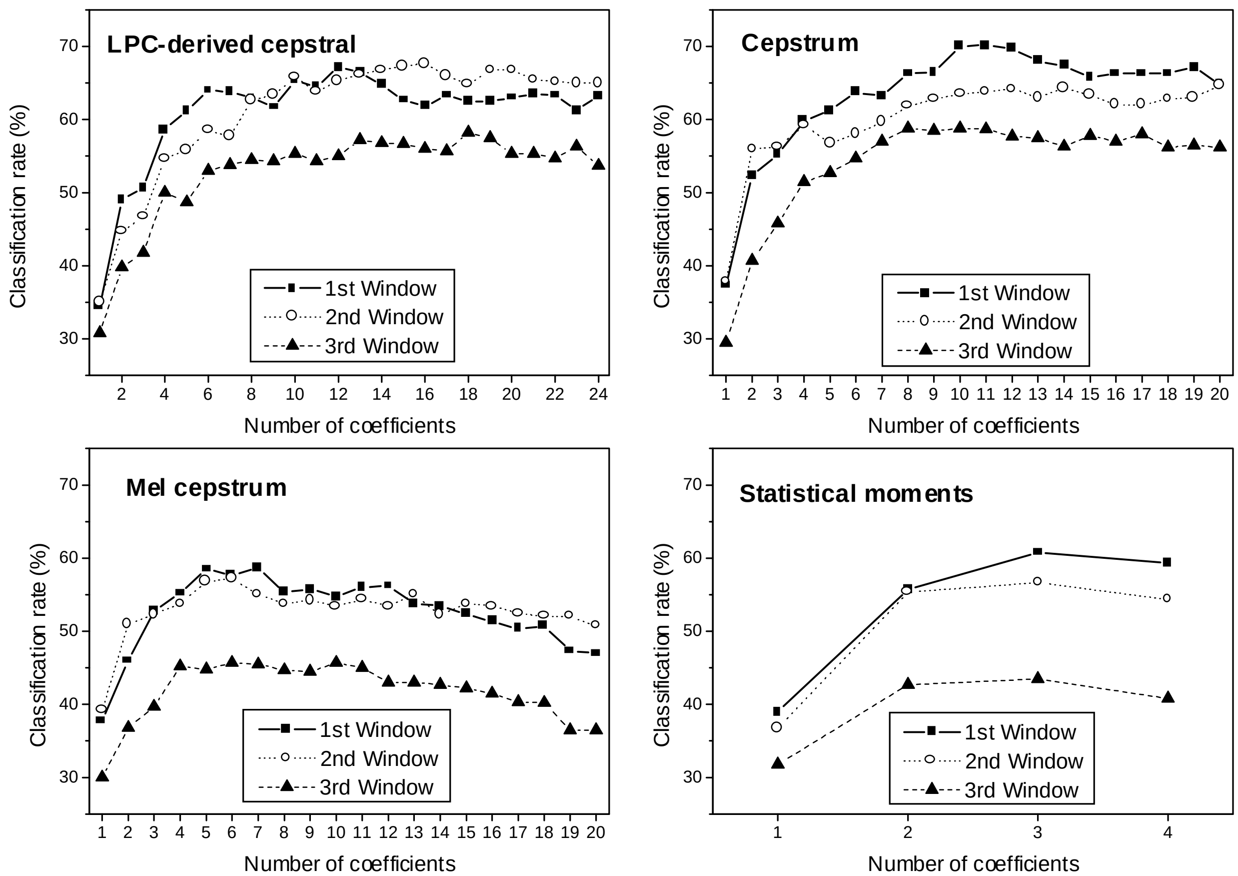

Figure 1. Classification results for each acoustic representation, window and number of coefficients.

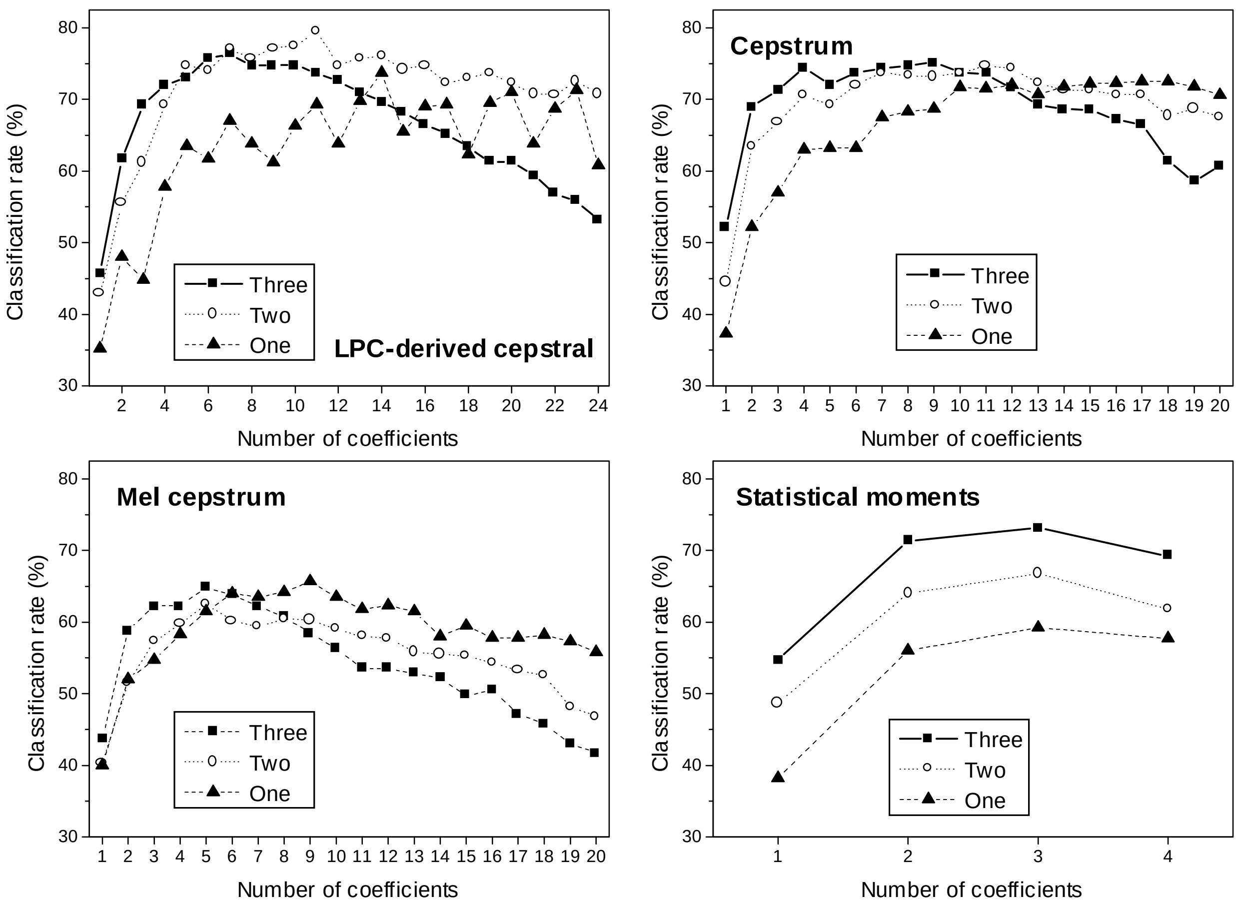

Figure 2. Classification results for each acoustic representation, number of coefficients and cosine series expansion: Squares: the three windows; circles: two cosine coefficients; triangles: one coefficient.

that way, the classification results are more realistic and close to the true classification rates.

First of all, the importance of the dynamics of the fricative noises for their classification will be studied. The classification results for each window were computed varying the number of coefficients, N, included in the analysis. For the LPC-derived cepstral coefficients this means that they have to be computed for each particular order, N. The results are shown in *Figure 1*. The first two windows gave similar results while the third window gave poorer classification rates. The best spectral representations are the LPC-derived cepstral coefficients and the cepstrum coefficients.

To obtain classification rates for the whole fricative noise the information of the three windows was included in the analysis: the value of each parameter for each window was expanded using a cosine series expansion [5]:

$$P(n) = \sum_{k=1}^{M} C_k \cos\left(\frac{(k-1)\pi(n-0.5)}{L}\right)$$

where L is the total number of windows (three), P(n), $1 \le n \le L$, is the parameter value for window n, M is the number of cosine coefficients and C_k are the cosine coefficients. Thus, the cosine coefficients encode the smoothed trajectory of a speech parameter. C_1 is the average value of a parameter, C_2 is a measure of the tilt over time of a parameter trajectory and higher number terms encode additional details of a parameter trajectory. In our case the number of cosine coefficients was varied from one to three, three cosine coefficients being equivalent to consider the value of the parameter for every window.

Therefore, both the resolution of the dynamic and static spectral information was modified: i.e. the number of cosine coefficients and the number of parameters or order, respectively. The results can be seen in *Figure 2*. For every acoustic representation, including the three windows gives the best result, except for the LPC-derived cepstral coefficients which perform best for the two cosine coefficients expansion. An order of aproximately 4 seems adequate to represent most of the acoustic properties of the fricatives. The best acoustic characterizations are: a) The LPC-derived cepstral coefficients: a representation of order 11 and 2 cosine coefficients for a total of 22 parameters attained a classification rate of 79.5 %; b) The cepstrum coefficients: a representation of order 9 for the 3 windows for a total of 27 parameters attained a classification rate of 75.2 %; c) The first 3 statistical

	/θ/	/f/	/s/	/ʃ/	/x/	/tʃ/
/θ/	**70**	18	5	0	4	3
/f/	32	**65**	0	0	3	0
/s/	0	1	**89**	8	1	1
/ʃ/	0	0	14	**75**	0	11
/x/	0	3	0	0	**97**	0
/tʃ/	3	0	3	12	1	**81**

	/θ/	/f/	/s/	/ʃ/	/x/	/tʃ/
/θ/	**57**	27	8	0	6	2
/f/	27	**63**	3	0	5	2
/s/	2	1	**81**	13	0	3
/ʃ/	0	0	20	**69**	0	11
/x/	1	1	0	1	**97**	0
/tʃ/	3	0	7	6	0	**84**

Table 2. Classification matrices for the two best characterizations: Top: LPC-derived cepstral coefficients (11^{th} order and 2 cosine coefficients); bottom: cepstrum coefficients (9^{th} order and 3 cosine coefficients). Mean classification rates: Top: 79.5 %; bottom: 75.2%.

moments for the 3 windows for a total of 9 parameters attained a classification rate of 73.2 %. The classification matrices for these first two characterizations can be seen in *Table 2*.

The resulting *a posteriori* probabilities were used to correlate these first two acoustic characterizations with the responses of the listeners to the perceptual experiment. The overall correlation coefficients were 0.78 and 0.75 for the aforementioned cepstral and cepstrum representations, respectively. For each fricative, they were higher for the perceptually best defined phonemes as, for example, /x/: 0.94 for both the cepstral and cepstrum characterizations. The correlation coefficients can be seen in *Table 3*.

5. CONCLUSION

Summarizing, the perceptual identification of the Spanish isolated fricatives has been studied together with several of the possible spectral characterizations of these phonemes. A notable correlation has been found between the perceptual identification and the spectral characterization of the fricatives. A low order representation, aproximately four coefficients, gives satisfactory results and seems to represent most of the acoustic properties of the fricatives. This is the case for the LPC-derived cepstral coefficients, cepstrum coefficients and even for the statistical moments. The mel cepstrum coefficients gave poorer results. Nevertheless, for the LPC-derived cepstral coefficients a slightly higher order can give even better results. Thus, in spite of the inability of the LPC method to model zeros, the LPC-derived cepstral coefficients are a

	/θ/	/f/	/s/	/ʃ/	/x/	/tʃ/	Overall
LPC	0.62	0.66	0.84	0.76	0.94	0.80	0.78
FFT	0.51	0.61	0.74	0.75	0.94	0.84	0.75

Table 3. Correlation coefficients for the two best characterizations: Top: LPC-derived cepstral coefficients; bottom: cepstrum coefficients.

good acoustic representation for fricatives. On the other hand, the dynamics of the fricatives are well represented by the three considered windows. A smoother representation, say two cosine coefficients, attain lower rates for every acoustic representation except for the LPC-derived cepstral coefficients, for which both acoustic representations attain similar rates. The representation with one cosine coefficient is always inadequate.

6. REFERENCES

[1] Jongman, A., Sereno, J. A., Wayland, R. and Wong, S. (1998), Acoustic properties of English fricatives. *Proceedings of the ICA/ASA'98*, pp. 2935–2936.

[2] Shadle, C. H. and Mair, S. J. (1996), Quantifying spectral characteristics of fricatives. *Proceedings of ICSLP'96*, vol. 3, pp. 1521–1524.

[3] Forrest, K., Weismer, G., Milenkovic, P. and Dougall, R. N. (1988), Statistical analysis of word-initial voiceless obstruents: Preliminary data. *J. Acoust. Soc. Am.*, vol. 94, pp.115-123.

[4] Davis, S. B. and Mermelstein, P. (1980), Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-28, pp. 357–366.

[5] Zahorian, S. A. and Jagharghi, A. J. (1993), Spectral-shape features versus formants as acoustic correlates for vowels. *J. Acoust. Soc. Am.*, vol. 94, pp. 1966-1982.