ELSEVIER

Speech Communication 18 (1996) 369–379

SPEECH COMMUNICATION

Acoustic parameters for place of articulation identification and classification of Spanish unvoiced stops

M.I. Torres [a,*], P. Iparraguirre [b]

[a] *Dpto. Electricidad y Electrónica, Universidad del País Vasco, Apdo. 644, 48080-Bilbao, Spain*
[b] *Dpto. Arquitectura de Computadores, Universidad del País Vasco, Apdo 649, 20080-San Sebastian, Spain*

Received 27 April 1994; revised 12 March 1996

Abstract

The analysis of the acoustic parameters which best summarize the cues to phone discrimination for the language under consideration should be a previous step in acoustic-phonetic decoding, regardless of the methodology to be used. The Spanish language has not been widely analyzed from this point of view. This work deals with the acoustic discrimination of Spanish stop consonants. Our main goal was to find a reliable and reduced set of parameters for place of articulation identification of Spanish unvoiced stops. On the basis of the obtained parameters, two automatic classifiers were developed and tested. Only the acoustic features of the burst segment, automatically segmented from the speech waveform, were considered in the parameter estimation. The analysis of these features was carried out in both the time and frequency domains over a CV context corpus uttered by 6 speakers. In the first case, the classifier was designed as a procedural form. Alternatively, in the second case a statistical classifier was obtained from a previous automatic discriminant analysis of the parameters. Both classifiers were tested over a CV context corpus uttered by 40 new speakers not included in the analysis corpus, which resulted in a good rate of identification.

Zusammenfassung

Die Analyse der akustischen Parameter, die Phonemen einer Sprache unterscheiden, ist ein vorhergehender Schritt um die akustische-phonetische Dekodifizierung zu verstehen, mit Unabhängigkeit der Methodologie die man benutzt. Die Spanische Sprache ist aber unter diese Perspektive sehr wenig studiert worden. Diese Arbeit handelt über die Unterscheidung der Spanischen okklusiveren Konsonanten. Unser Ziel war eine kleine Menge von bedeutsamen Parameter für die Identifizierung des Gliederungspunktes der Spanischen tauben Okklusiven zu suchen. Von diesen Parameter ausgehend, hat man zwei automatische Klassifizierungsalgorithmen entwickelt und probiert. Um die Parameter abzuschätzen sind nur die akustischen Kennzeichnungen des explosives Segmentes betrachtet worden. Dieser wurde aus der Form der Stimmenwelle herausgezogen. Die analyse der Charakteristiken wurde mit sechs verschiedenen Sprecher auf einem CV Silbencorpus sowohl im Zeit – wie im Frekuentiellbereich ausgeführt. Im ersten Fall hat man die verschiedenen Kennzeichnungen mit einem Regelsystem sortiert. In zweiten Fall hat man gegenüber ein statistischer Sortierer aus einer vorhergehenden Unterscheidungsanalyse der Parameter entwickelt. Um beide Sortierungen zu probieren ist ein CV Silbencorpus von 40 neue Sprecher ausgesprochen worden und man hat gute Identifizierungsprozentsätze erreicht.

* Corresponding author. E-mail: manes@we.lc.ehu.es.

PII S0167-6393(96)00025-8

Résumé

L'analyse des paramètres acoustiques pour la caractérisation et l'identification des phonèmes de la langue à étudier représente la première étape du décodage acoustique-phonétique, indépendamment de la méthode utilisée. On trouve malgré tout très peu d'analyses acoustico-phonétiques de la langue espagnole. Le travail présenté ici traite de la discrimination des occlusives espagnoles. Rechercher un ensemble réduit de paramètres robustes pour identifier le lieu d'articulation des occlusives sourdes a constitué notre premier objectif. A partir de cet ensemble de paramètres, on a ensuite élaboré et évalué deux algorithmes de classification et de reconnaissance automatique. Après une étape de localisation automatique de l'explosion ("burst"), seules les caractéristiques acoustiques de ce segment sont prises en compte pour l'estimation des paramètres. On a mesuré aussi bien les caractéristiques temporelles que fréquentielles sur un corpus de syllables CV prononcées par 6 locuteurs. On a conçu, dans le premier cas, un système procédural à base de règles, pour reconnaître le lieu d'articulation. Pour le traitement des paramètres fréquentiels, on a mis au point un classifieur statistique, développé à partir d'une analyse discriminante préalable des paramètres d'entrée. Un corpus de syllables CV prononcées par 40 nouveaux locuteurs, non utilisé pour la définition et l'analyse des paramètres, a permis d'évaluer les deux systèmes qui ont tous deux conduit à de bons résultats d'identification.

Keywords: Speech recognition; Acoustic-phonetic decoding; Acoustic parameters; Spanish stop consonants; Knowledge-based systems; Discriminant analysis; phone classifiers

1. Introduction

Nowadays the design of accurate acoustic-phonetic modules has been shown to be an important issue in developing Large Vocabulary and/or Continuous Speech Recognition (CSR) systems (Schwartz, 1988; Lee et al., 1990; Haton, 1988). In fact, results on phone-recognition (plain Acoustic-Phonetic Decoding, APD) are interesting not only because of their own applications, but also because good phonetic decoding leads to good word decoding.

An important point in the design of an APD is the selection of a sub-lexical unit. This selection should be made on the basis of coverage of the language, context variability and frequency of occurrence in the available training corpus. Several proposals appear in the literature related to this subject: phones (Lee et al., 1990), context-dependent phones (Ney and Billi, 1991), diphones, triphones (Fissore et al., 1991), syllables, etc. The choice of phones as sub-lexical units offers several advantages. The most important is that the size of the set of such units is low enough to obtain a high score of occurrences for each unit in the training set. On the other hand, they are vocabulary-independent so that they could be used to model new words not appearing in the basic training set. The phonological variations of phonemes can be predicted by contextual rules within and between sounds (Haton, 1988; Benedí and Torres, 1992). An in-depth study of the problem, reported in previous works (Torres et al., 1994), was made for Spanish, which resulted in the choice of the phone as the sub-lexical unit.

At this point, an analysis of the acoustic parameters that best summarize the cues to phone discrimination for the language under consideration is a preliminary step. In CSR this kind of study is required in knowledge-based systems (Ederveen and Bores, 1991), and it is also highly recommended in other methodologies (Mariani, 1989) like Stochastic Modelling (Galiano et al., 1994) or Artificial Neural Networks (Bengio et al., 1992). In fact, an in-depth knowledge of the acoustic characterization of phonemes leads to a better choice and application of a given methodology to a specific language.

The Spanish language has not been widely analyzed from this point of view. Only a few specific works focusing on vowel characterization can be found (Torres, 1990). An adaptation of some previous analyses designed for the French language has also been developed leading to poor decoding results (Benedí et al., 1991). A specific analysis of the Spanish language is thus required.

This work deals with the acoustic discrimination of Spanish stop consonants. Our main goal was to find a reliable and reduced set of parameters for identification of place of articulation of Spanish un-

voiced stops. On the basis of the obtained parameters, two automatic classifiers were developed and tested. Only the acoustic features of the burst segment, automatically segmented from the speech waveform, were considered in the parameter estimation. Thus, no information about the vowel transition was included and, as a consequence, the parameters could be considered as context independent. The analysis of these features was carried out in both the time and frequency domains. In the first case, the classifier was designed as a procedural form. Alternatively, in the second case a statistical classifier was obtained from a previous automatic discriminant analysis of the parameters. Both classifiers were tested over a CV context corpus uttered by 40 speakers not included in the analysis corpus, resulting in a good rate of identification.

The paper is organised as follows. In Section 2 we present a brief acoustic characterization of Spanish stop consonants as well as the main cues to be considered for the identification of their place of articulation. Section 3 summarizes the methodology used and the burst segmentation procedure. In Section 4 we report the two approaches presented in both the time and frequency domains. Finally, Sections 5 and 6 present the experimental results and conclusions, respectively.

2. Acoustic characterization

Plosive (or stop) consonants are characterized by a momentary interruption of the sound emission produced by a closure of the vocal channel (Fant, 1973; Quilis, 1989). Three time-consecutive segments can be distinguished in the realization of a plosive consonant, in the CV context:

1. Occlusion. The airstream through the vocal tract is interrupted by the closure of the articulators.

2. Burst. The articulators come apart and the airstream is released producing a brief, turbulent and intense sound. The manner of articulation characterizes this segment. After this moment, a short fricative segment can appear. We do not separate these two segments since many times it is not possible to do so.

3. Transition. Transition segment to the next voiced sound. Plosives usually present very fast vowel transitions. This segment is strongly dependent on the coarticulation effects between both sounds.

Table 1
Classification of Spanish oral stops

	Bilabial	Dental	Velar
Voiced	/b/	/d/	/g/
Unvoiced	/p/	/t/	/k/

This description is accepted by most authors (Fant, 1973; Quilis, 1989). A more general description must include an aspiration segment after the explosion, as is typical in most English stop consonants. Nevertheless, Spanish does not have aspirated stops as occurs in many other languages.

Spanish stops classification considers two main features: voicing and articulation (Martínez Celdrán, 1986; Quilis, 1989). Table 1 shows the classification of Spanish oral stop consonants.

Voicing is produced by the vibration of the vocal cords at the time of occlusion. Thus, a low frequency signal appears before the burst. At this time, unvoiced stops are characterized by a silence. Moreover, Spanish unvoiced stops present more energy; this energy is also more clearly distributed in unvoiced stops.

The cues for perception of the place of articulation are not as clear. There is not complete agreement among authors about the acoustic discriminatory features that characterize the articulation of stop consonants. Some authors consider the burst spectral shape to be the main aspect to be considered (Blumstein and Stevens, 1979; Blumstein et al., 1982; Bush et al., 1983). A spectral pattern would characterize each articulation regardless of the next vowel. However, this theory has been questioned by establishing that the cues for the perception of the stop articulations may be related to dynamic characteristics of the speech signal (Kewley-Port, 1982, 1983; Kewley-Port and Pisoni, 1983; Zahorian et al., 1987; Kobatake and Ohtani, 1987). Thus, the transition to the next vowel would be the main consideration. In

either case, many features have been analyzed to identify place of articulation and then to classify the stop consonants (Kewley-Port, 1982; Pols and Schouten, 1985; Tartter et al., 1983; Repp and Lin, 1989; Gurlekian et al., 1985; Nathan, 1991).

In conclusion, each articulation is characterized as follows for Spanish stop consonants (Poch, 1984; Gurlekian et al., 1985; Castañeda, 1986; Torres, 1990; Torres and Iparraguirre, 1993; Benlloch et al., 1992):

– **Bilabial.** Very short burst when detected. Diffuse spectral shape at burst with more relevant energy at low frequency band. Very quick and negative transitions for all the formants.

– **Dental.** Longer burst and very often the presence of a fricative segment. Diffuse spectral shape at burst with strong presence of high frequency energy. Slower transitions, positive at lower formants and negative at higher formants.

– **Velar.** Several bursts followed by friction segments. Very long burst segment. Burst spectral shape compacted around the middle-frequency band.

The main goal of this work was the identification of place of articulation in Spanish unvoiced stops. However, we were interested in context-independent features and, as a consequence, formant transitions, which are strongly related to the articulation of the next vowel, were not considered. Thus, the acoustic features considered and related parameters to be presented in the following sections characterize the burst segment exclusively.

3. Methodology

The objective of the methods to be presented is to parametrize the acoustic features of speech signals in order to obtain classification algorithms for unvoiced stop consonants. A set of parameters representing these features has been defined. After an analysis of the behaviour of the parameters over a set of samples uttered by a small number of speakers, we have designed some recognition procedures. These procedures were then tested over a larger sample set uttered by a new set of speakers.

The analysis procedure was carried out in both the time and frequency domains. The discriminatory ability of a set of parameters was studied, which resulted in two recognition procedures: time and frequency procedures. In the first case, a set of rules forming a decision tree was elaborated. This is a very common decision procedure in knowledge-based systems. However, in the frequency domain, the number of parameters to be analyzed as well as the number of possible discriminatory rules to be elaborated were very high. On the other hand, some of them could include redundant information. Thus, a previous discriminant analysis of the parameters was recommended and resulted in a statistical classifier (Torres and Iparraguirre, 1993). A more general framework such as the design of a knowledge-based acoustic-phonetic decoder could involve alternative discriminant procedures but, in any case, they are not usually too different to those included in this work.

The corpus used for the feature parametrization, parameter analysis and algorithm design procedures, consisted of 15 CV samples (three unvoiced stop consonants along with the five vowels) uttered by 6 speakers, 3 males and 3 females, resulting in a total of 90 samples. For testing purposes, a corpus consisting of the 15 CV samples uttered by 40 new speakers, 20 males and 20 females, was used resulting in a total of 600 samples. This corpus was acquired at 10 Khz and digitized to 12 bits. For the frequency domain procedure, the spectrum amplitude was obtained by Fourier transform (FFT) of segments of 25.6 ms.

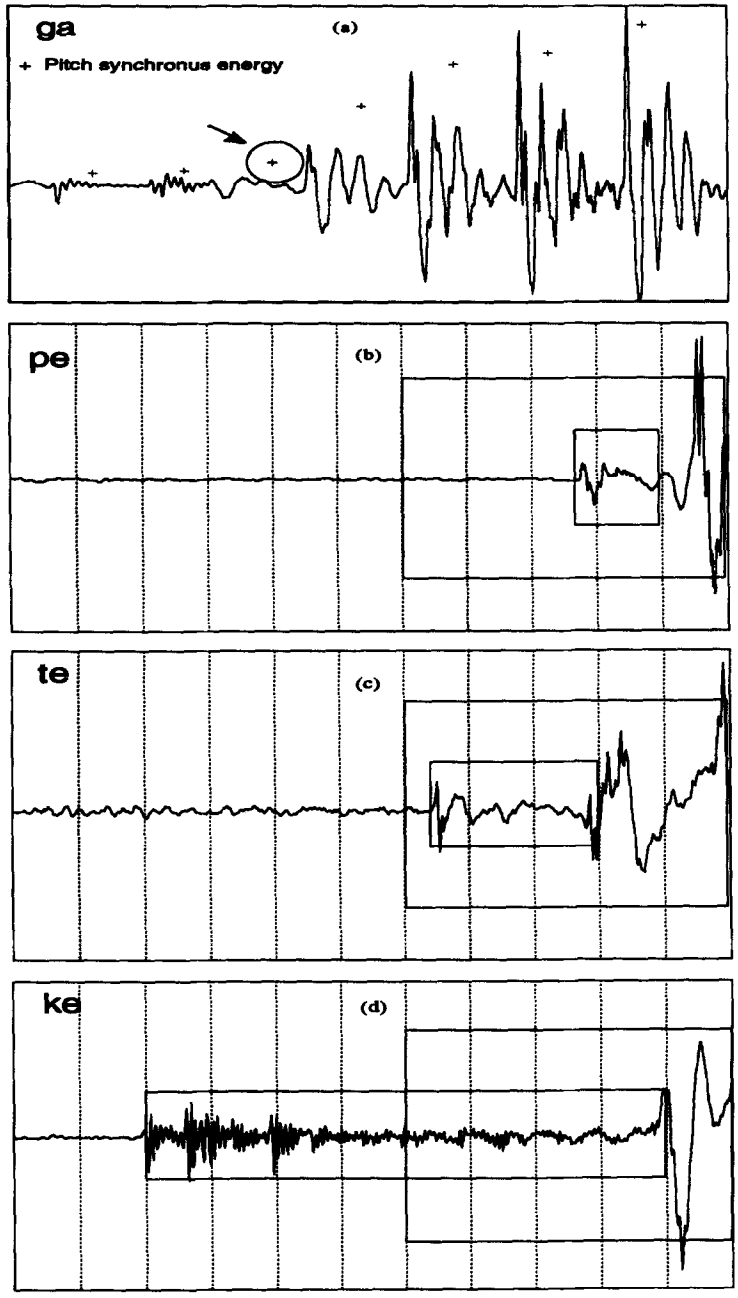
3.1. Burst extraction

In this work, the unvoiced stop consonant discrimination was carried out by modelling the time and frequency acoustic features of the burst segment only. As a consequence, a robust localization of this segment was needed. On the other hand, the burst segment should be automatically identified and, thus, any manual segmentation will be avoided.

The general procedure was based on a pitch synchronous analysis of the signal. Thus, a specific procedure (Torres, 1990) was previously used to estimate the pitch period. Then, the burst localization procedure was developed in two steps. In the first step, the normalized average energy was calculated in pitch synchronous frames. When an abrupt change of the average energy was detected, a fixed length segment of 25.6 msg located at this point was se-

lected. In the second step, the final burst segment was obtained by setting a minimum and a maximum energy threshold normalized to the vowel maximum energy. Some examples of this procedure are shown in Fig. 1.

The validity of this procedure was manually tested



(a) First step of the procedure: detection of an abrupt change of the average and pitch synchronous energy.
(b) (c) (d) Final burst localization for the three Spanish unvoiced stops uttered by the same speaker.

Fig. 1. Some examples of the burst detection procedure.

over the whole analysis set. The burst segments automatically located were analysed by three human experts which validated the procedure. No sample of the final test corpus was used in this test in order to preserve the absolute independence between the analysis and the test corpora.

This procedure allowed us to obtain automatically a burst segment of variable length, according to the specific features of the sample to be analyzed. The burst length could then be considered as a classification parameter.

4. Discriminative procedures

Two kinds of procedures based on both the time and frequency analysis of speech signals were designed. In both cases a set of parameters characterizing the burst shape needed to be previously defined. These parameters represent the following acoustic features observed in the burst of Spanish unvoiced stops in CV context:

/p/ – very short segment or no burst detected. Energy increases uniformly from occlusion to vowel. Low frequency oscillations are present.

/t/ – longer burst. Strong presence of fricative segment. Uniform energy distribution in the whole burst segment. High frequency oscillations are present.

/k/ – very long segment with multiple bursts. Very high difference between energy values in bursts and fricative segments. Strong presence of middle frequency oscillations. High frequency oscillations are also observed.

These features agree largely with those reported in previous sections characterising each place of articulation (Gurlekian et al., 1985; Castañeda, 1986; Quilis, 1989; Torres, 1990; Benedí et al., 1991).

4.1. Time domain

The first procedure was aimed at obtaining a set of discriminatory parameters from the speech waveform to identify the three articulatory points. Thus, the energy distribution as well as the zero crossing rate along the already selected analysis segment were the main features to be considered. The burst segment previously detected was then divided into frames of 5 msg. The energy normalized to the vowel maximum value and the zero crossing rate were computed in each frame. Based on these measures, a set of new parameters was defined and calculated for the burst segment:

zi: length (number of frames) of the burst segment. A very short segment can only correspond to a bilabial. On the contrary, a segment which is too long should be identified as velar.

maxcc: maximum zero crossing rate. It measures the maximum frequency in the segment to some degree.

nmax: number of energy peaks of the burst segment. The energy distribution along the segment could be represented by this parameter. Thus, a large number of energy peaks represents the multiple burst sequence of a velar.

In many samples only one significant peak of energy was found. For such a case, three more parameters were also computed:

ccmax: zero crossing rate at the frame of maximum energy. Thus, the form of oscillations in the maximum energy frame of the burst is also considered.

pend: energy slope after the energy peak. It allows us to identify short segments of high energy that appear typically in velars but not in dentals with a large fricative segment.

dvoc: distance (number of frames) between the energy peak and the beginning of the vowel. It measures the length of the low energy segment before the vowel onset, when present.

An analysis of the statistical distribution of these parameters over all the samples of the analysis set was carried out. Fig. 2 shows the corresponding histogram.

After the study of the behaviour of all parameters over the analysis samples, a set of discriminatory rules was elaborated. In this procedure the more discriminative parameters were first considered. These rules conform to the final decision algorithm shown in Fig. 3. The threshold values needed by the rules were established from the parameter distribution over the analysis sample set shown in Fig. 2. Alternative decision trees based on different rule order were also tested over the analysis set. However, the decision algorithm in Fig. 3 achieved the best classification scores.

4.2. Frequency domain

In this case the main goal was to obtain a small set of values that would be able to represent the spectral shape of the previously detected burst sample. As mentioned above, the segment to be analyzed was not of fixed length. Thus, a Fourier Spectrum of 25.6 msg, 12.8 msg or 6.4 msg was computed according to the actual length of the burst sample. In all the cases, the Spectrum obtained was reduced to

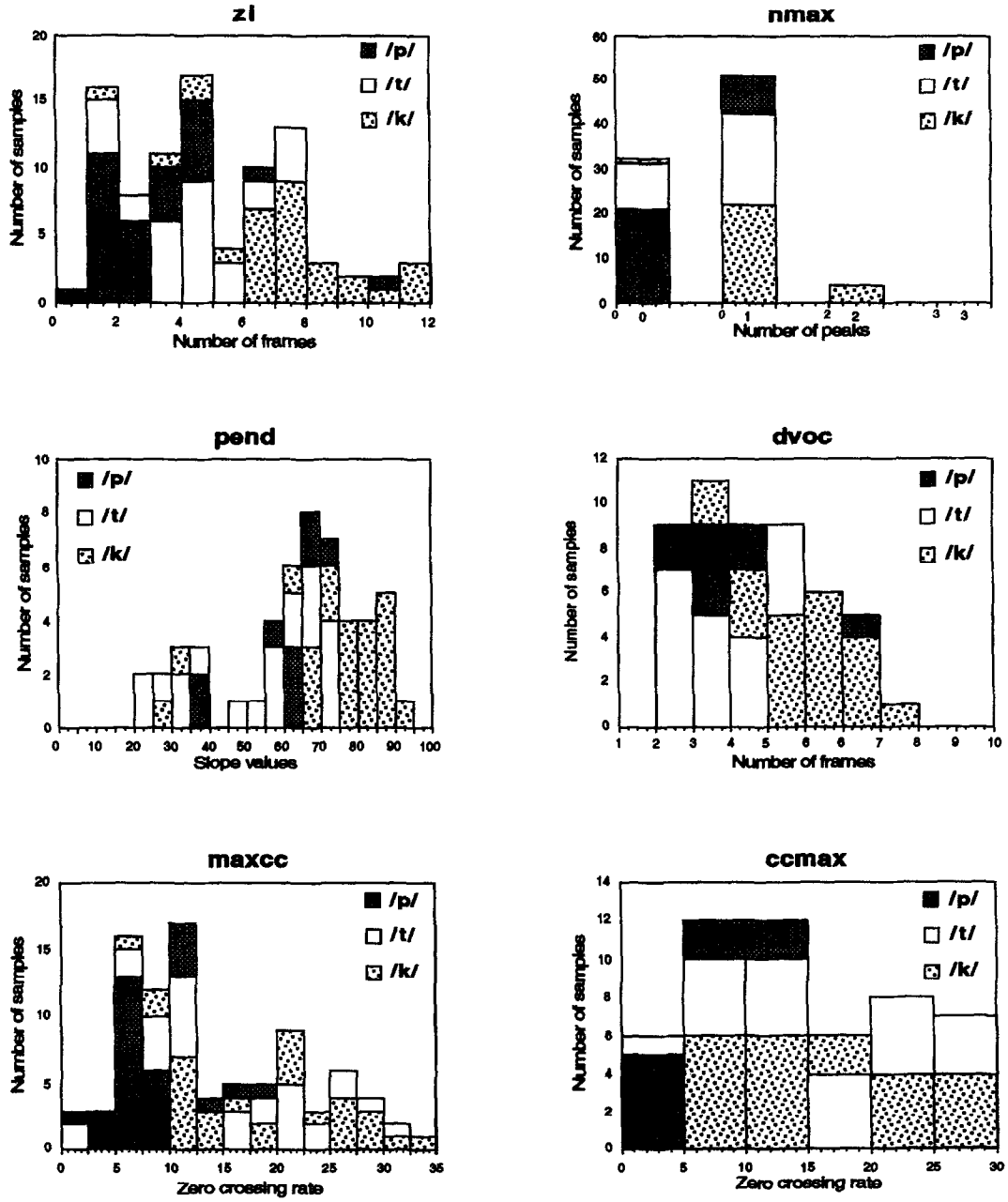


Fig. 2. Distribution of the time domain parameters over all the samples of the training set.

```
zi ≠ 0 => nmax classification
        nmax > 1 => /k/
        nmax = 0 => [/p/-/t/]
                        zi = 1 => /p/
                        zi ≠ 1 => [/p/-/t/]
                                        maxcc < 10 => /p/
                                        maxcc ≥ 10 => /t/
        nmax = 1 => [/p/-/t/-/k/]
                        pend > 75 % => /k/
                        pend < 75 % => [/p/-/t/-/k/]
                                        maxcc < 10 => [/p/-/t/]
                                                        ccmax ≤ 5 => /p/
                                                        ccmax > 5 => /t/
                                        maxcc ≥10 => [/t/-/k/]
                                                        dvoc ≥ 4 => /k/
                                                        dvoc < 4 => /t/
```

Fig. 3. Decision algorithm for the unvoiced stops classification with time domain parameters.

64 points. Samples shorter than 6.4 msg were not computed and were directly classified as bilabial /p/.

Our goal was to reduce the spectral distribution shape to a small set of parameters. A fixed number of points (three centroids) was chosen to represent the Spectrum. The centroid positions were calculated by minimizing the error of such a simplification (Crowe and Jack, 1987). Therefore, the spectral distribution was divided into three energy bands of variable length adapted to the specific spectral shape of the sample being considered. Finally, a set of parameters that characterized the spectrum was obtained:

k_1, k_2, k_3: position (Hz) of the three centroids.
c_1, c_2: band limits calculated by the algorithm.
k_{21}, k_{32}, k_{31}: distances between centroids.
c_{21}: distance between band limits.
em_1, em_2, em_3: average energy of each band.
em_{12}, em_{23}, em_{13}: ratios between average energies.

Fig. 4 shows a representation of these parameters for several samples.

If these parameters were applied to voiced segments, a relationship between them and formant energy and position could surely be found. However, in this case no assumption was made about the number of energy peaks, where they should be found or the band limits position. The band limits are adapted to each spectral shape and the ratios between average energies allow us to define the number of peaks actually found in the sample. Thus, the spectral energy distribution of each articulation was well-represented by these parameters since the way they were obtained was adapted to each specific sample shape.

A discriminant analysis of these parameters was carried out using the statistical software SPSS-X. Thus, a discriminant function was calculated as a linear combination of the selected parameters. The analysis set of samples was exclusively used in this procedure. A discriminant punctuation could then be computed from the evaluation of the discriminant functions for each sample and a Bayes classifier could then assign a probability to each sample and group. Such a classifier was finally tested over the test set of samples (see experimental evaluation).

Actually, the process was a two-level procedure. In the first step, the sample was assigned to one of the three two-consonant groups: [p–t], [t–k] or [p–k]. Then, a second discriminant function led to the final classification. In the first step the variables k_{21}, k_{32}, k_{31} and c_{21} were excluded by the discriminant analysis. In the second step, the variables selected by the system to calculate the discriminant function were:

/p/–/t/ discrimination: k_3, k_{21} and c_{21}.
/t/–/k/ discrimination: em_{13}, k_{31}, em_{23}, em_{12} and c_{21}.
/p/–/k/ discrimination: k_2, c_2, k_1, em_1, k_{31}, em_{12} and em_{13}.

It is important to note the different contribution of each variable to the final characterization of each articulation. Thus, the discrimination bilabial/dental could be achieved without any energy value. The relationship between the spectral shape at low and

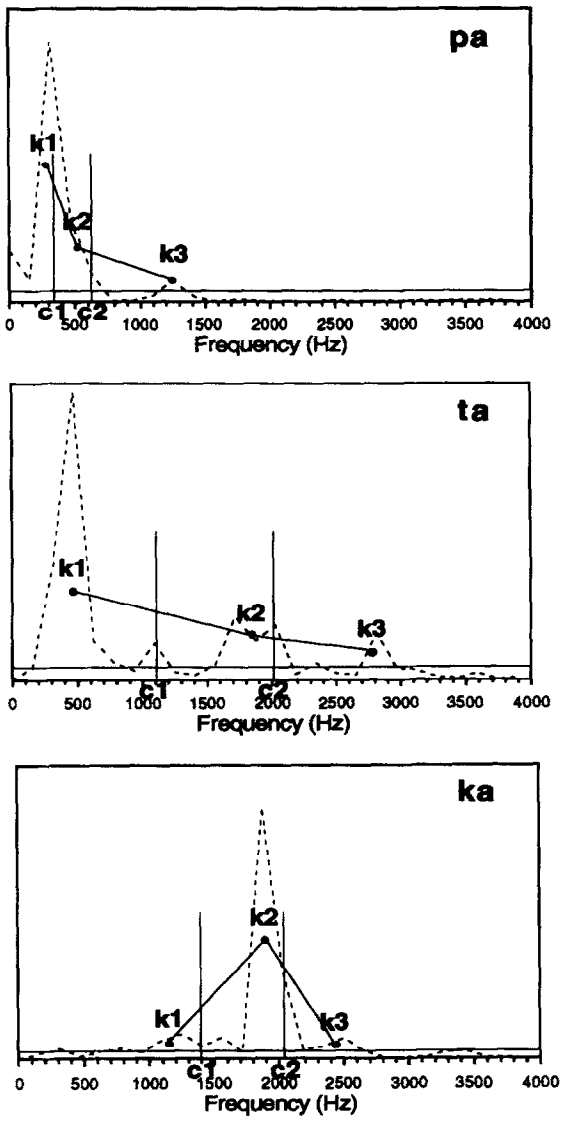


Fig. 4. Spectral distribution of the three unvoiced stops represented by three centroids.

middle frequencies seemed to be the most significant variable to be considered. On the other hand, dental/velar and bilabial/velar discrimination needed a higher number of variables. In the first case, ratios between average energies were mainly considered. In the second one, parameters related to the first and third bands were the most significant.

5. Experimental evaluation

In this section, we report the results obtained through the methodologies described in Section 4. The evaluation corpus consisted in 15 CV (unvoiced stop consonant–vowel) samples uttered by 40 new speakers which were not included in the analysis corpus previously used, resulting in a total of 600 samples. Thus, even if the analysis procedures were carried out over a reduced number of speakers, the validity of the proposed parameters was tested over a large number of speakers. The size of the test corpus

Table 2
Confusion matrix obtained after the application of the time domain recognition procedure to the test corpus (Fig. 3)

Consonant uttered	Identified consonant /p/	/t/	/k/	rate of recognition
/p/	151	31	18	75.5 %
/t/	20	118	62	59.0 %
/k/	4	19	177	88.5 %
Average rate of recognition:				74.33 %

Table 3
Confusion matrix obtained after the application of the frequency domain classifier to the test corpus

Consonant uttered	Identified consonant /p/	/t/	/k/	rate of recognition
/p/	150	40	10	75.0 %
/t/	31	147	22	73.5 %
/k/	17	33	150	75.0 %
Average rate of recognition:				74.5 %

was, then, nearly seven times the size of the analysis one.

In Tables 2 and 3, we present the results of the experiments carried out with the time domain and frequency domain methods, respectively.

The comparison between time and frequency parameters should be carefully made since the decision procedure was not the same in both classifiers. However, there were some issues that should be noted and analysed. The performance of both classifiers was quite similar. Nevertheless, the distribution of the rate of recognition among consonants was homogeneous in the second classifier but not in the first one. In this case, dental /t/ achieved quite a low rate of recognition due to the decision rule features. The parameters analyzed in the time domain characterized velar and labial articulations. In fact, we did not find a clearly discriminatory parameter for dental articulation. This effect was not present in the frequency domain analysis. Finally, it is important to note that both classifiers obtained very high rates of recognition for bilabial /p/. This articulation was the least identified in human intelligibility tests (Torres, 1990).

The classification errors shown in Tables 2 and 3 could also include some burst localization errors due to the automatic procedure reported in Section 3. However the experiments resulted in good identification rates. Only the burst shape was considered and no contextual information was supplied to identify place of articulation in the reduced, but highly confused, set of Spanish unvoiced stops (Quilis, 1989; Torres, 1990). A combination of both kinds of parameters could be considered in a more general framework like a knowledge-based acoustic-phonetic decoder (Benlloch et al., 1992; Benedí et al., 1994).

6. Concluding remarks

The main goal of this work was to find a reliable and reduced set of parameters for place of articulation identification of Spanish unvoiced stops. Two sets of parameters characterizing the burst segment have been proposed in both the time and the frequency domain. Then, two automatic classifiers for the Spanish unvoiced stops have been presented for evaluation purposes. Both classifiers were tested over a corpus consisting of 15 CV context samples uttered by 40 speakers which were not included in the analysis corpus (600 samples) and resulted in a good rate of identification.

Some important features were shared by the two methodologies presented:

– The burst segment was automatically localized and segmented. No hand segmentation was used.

– Only the information supplied by the burst was quantified and then used in the algorithm design step.

– No contextual information was used. Both methods classified unvoiced Spanish stops regardless of the vowel context.

Both classifiers obtained a similar average rate of identification in speaker-independent tests. However, the distribution of the rate of recognition among consonants was much more homogeneous when a discriminant analysis of the frequency domain parameters was used. Thus, the centroids of the burst spectral shape seemed to characterize the three places of articulation better than the parameters obtained after an analysis of the segment in the time domain. However, the acoustic features of the burst segment do not seem to be enough to achieve a complete identification of the place of articulation of Spanish stop consonants. Features based on dynamic spectra as well as vowel transition and coarticulation effects should also be considered.

Unvoiced stops are usually well detected but difficult to discriminate in continuous speech. Thus, the proposed sets of parameters, based simultaneously on acoustic-phonetic knowledge and statistical assessment, constitutes an interesting proposal to characterize burst segments in isolated words or continuous speech. In fact, a combination of both kinds of parameters is being used in a knowledge-based acoustic-phonetic decoder for Spanish (Benlloch et al., 1992; Benedí et al., 1994). The automatic burst localization has also demonstrated to be robust enough in such a more general framework.

Acknowledgements

The authors would like to thank the reviewers for their comments and suggestions that have surely improved the quality of the first version of the paper.

References

J.M. Benedí and I. Torres (1992), The contextual factors of phonetic variation: Coarticulation and word junction effect, ESPRIT II Project D13-IV ROARS report.

J.M. Benedí, I. Benlloch, I. Torres, J.A. Gomez, M.J. Castro and J.A. Puchol (1991), Acoustic-phonetic knowledge for the Spanish version of the ROARS system, ESPRIT II Project, D13 Part II ROARS Report, WP1-T6.

J.M. Benedí, M.J. Castro and J.A. Gomez (1994), "Decodificación acústico-fonética para el Castellano en el sistema ROARS", *Novática*, Vol. 56, pp. 27–31.

Y. Bengio, R. De Mori, G. Flammia and R. Kompe (1992), "Phonetically motivated acoustic parameters for continous speech recognition using artificial networks", *Speech Communication*, Vol. 11, Nos. 2–3, pp. 261–271.

I. Benlloch, M.J. Castro, I. Torres, J.A. Gomez, J. M. Benedí and J.A. Puchol (1992), Acoustic-phonetic knowledge for the Spanish version of the ROARS system, ESPRIT II Project, D13 Part II ROARS Report, WP1-T7.

S.E. Blumstein and K.N. Stevens (1979), "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants", *J. Acoust. Soc. Amer.*, Vol. 66, No. 4, pp. 1001–1017.

S.E. Blumstein, E. Isaacs and J. Mertus (1982), "The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants", *J. Acoust. Soc. Amer.*, Vol. 72, No. 1, pp. 43–50.

M.A. Bush, G.E. Kopec and V.W. Zue (1983), "Selecting acoustic features for stop consonant identification", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Boston, MA, pp. 742–745.

M.L. Castañeda (1986), "El VOT de las oclusivas sordas y sonoras españolas", *Estudios de Fonética Experimental II*, Facultad de Filología Barcelona, pp. 91–110.

A. Crowe and M.A. Jack (1987), "A globally optimising formant tracker using generalised centroids", *Electronics Lett.*, Vol. 23, No. 19, pp. 1019–1020.

D. Ederveen and L. Bores (1991), "Knowledge-based phoneme recognition", *Proc. Eurospeech*, Genova, pp. 42–45.

G. Fant (1973), "Stops in CV syllables", in *Speech Sounds and Features* (MIT Press, Cambridge, MA), pp. 111–139.

L. Fissore, E. Giachin, P. Laface and G. Micca (1991), "Selection of speech units for a speaker-independent CSR task", *Proc. Eurospeech*, Genova, pp. 1389–1392.

I. Galiano, E. Sanchis, I. Torres and F. Casacuberta (1994), "Acoustic-phonetic decoding of Spanish continuous speech", *Internat. J. Pattern Recognition Artificial Intelligence*, Vol. 8, No. 1, pp. 155–180.

J.A. Gurlekian, M. Guirao and H.E. Franco (1985), "Acoustic characteristics and perception of Spanish stop consonants", *J. Acoust. Soc. Japan*, Vol. 65, No. 36, pp. 271–278.

J.P. Haton (1988), "Knowledge-based aproaches in acoustic-phonetic decoding of speech", in *Recent Advances in Speech understanding and Dialog Systems*, ed. by H. Niemann, M. Lang and G. Sagerer (Springer, Berlin).

D. Kewley-Port (1982), "Measurement of formant transitions in naturally produced stop consonant–vowel syllables", *J. Acoust. Soc. Amer.*, Vol. 72, No. 2, pp. 379–389.

D. Kewley-Port (1983), "Time-varying features as correlates of place of articulation in stop consonants", *J. Acoust. Soc. Amer.*, Vol. 73, No. 1, pp. 322–335.

D. Kewley-Port and D.B. Pisoni (1983), "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants", *J. Acoust. Soc. Amer.*, Vol. 73, No. 5, pp. 1779–1793.

H. Kobatake and S. Ohtani (1987), "Spectral transition dynamics of voiceless stop consonants", *J. Acoust. Soc. Amer.*, Vol. 81, No. 4, pp. 1146–1151.

C.H. Lee, L.R. Rabiner, R. Pieraccini and J.G. Wilpon (1990), "Acoustic modeling for large vocabulary speech recognition", *Computer Speech and Language*, Vol. 4, pp. 127–165.

J. Mariani (1989), "Recent advances in speech processing ", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Glasgow, Invited paper, pp. 429–440.

E. Martínez Celdrán (1986), *Fonética* (Teide, Madrid).

K.S. Nathan (1991), "Comparison of formant transition based stop classifiers: Time-varying and time-invariant signal models", *Proc. Eurospeech*, Genova, pp. 147–150.

H. Ney and R. Billi (1991), "Prototype systems for large-vocabulary speech recognition: Polyglot and Spicos", *Proc. Eurospeech*, Genova, pp. 193–200.

D. Poch (1984), "Datos acústicos para la caracterización de las oclusivas sordas del español", *Folia Phonetica*, Vol. I, pp. 89–106.

L.C.W. Pols and M.E.H. Schouten (1985), "Plosive consonant identification in ambiguous sentences", *J. Acoust. Soc. Amer.*, Vol. 78, No. 1, pp. 322–335.

A. Quilis (1989), *Fonética acústica de la lengua española* (Gredos, Madrid).

B.H. Repp and H.B. Lin (1989), "Acoustic properties and perception of stop consonant release transients", *J. Acoust. Soc. Amer.*, Vol. 85, No. 1, pp. 379–396.

R.M. Schwartz et al. (1988), "Acoustic-phonetic decoding of speech", in *Recent Advances in Speech understanding and Dialog Systems*, ed. by H. Niemann, M. Lang and G. Sagerer (Springer, Berlin).

V.C. Tartter, A.G. Samuel and B.H. Repp (1983), "Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions", *J. Acoust. Soc. Amer.*, Vol. 74, No. 3, pp. 715–725.

I. Torres (1990), Contribución al reconocimiento automático de vocales y consonantes oclusivas del Castellano, PhD Thesis, Universidad del País Vasco.

I. Torres and P. Iparraguirre (1993), "Acoustic-phonetic decoding of Spanish occlusive consonants", *Proc. Eurospeech*, Berlin, pp. 457–460.

I. Torres, F. Casacuberta and L. Sánchez (1994), "Linguistic decoding of Spanish continuous speech with hidden Markov models", in *Advances in Pattern Recognition and Applications* (World Scientific, Singapore), pp. 207–217.

S.A. Zahorian, Z.B. Nossair and R.F. Coleman (1987), "Evidence against acoustic invariance in initial voiced stop consonants", *Meeting Acoust. Soc. Amer.*, Vol. 81, pp. 1–36.