RECOGNITION OF VOWELS IN FRICATIVE CONTEXT

Fernández, S., Feijóo, S., Balsa, R., Barros, N.

Departamento de Física Aplicada
Universidad de Santiago de Compostela
15706 Santiago de Compostela, SPAIN
E-mail: fasergio@usc.es

ABSTRACT

The role of fricative context on vowel recognition in a series of FV syllables being part of natural Spanish words is investigated. Perceptual tests were carried out to assess the recognition of vowels in fricative context, in two conditions: 1) Isolated vowel; 2) Fricative noise + vowel. Analysis of results show that adding the fricative noise improves the recognition of the vowel, while the acoustic analysis reveal that the distribution of the vowels is affected by fricative context. A possible explanation for this improvement, i.e. the coarticulatory influence of the vowel on the fricative, was investigated. The results indicate that coarticulation cannot explain that improvement, since only 7.7% of the cases which improve when the fricative is added, show a clear influence of the vowel on the fricative.

1. INTRODUCTION

Acoustic and perceptual analysis with natural and synthetic speech indicate that both the spectral shape and the first two formants are the main corelates of vowel perception (Peterson & Barney, 1952; Nossair & Zahorian, 1993), although higher formants may play a certain role as well. Nevertheless, formant frequencies of vowels are influenced by consonantal context and speaker characteristics, such as size, sex or age (Nearey, 1989; Strange *et al.*, 1983; Johnson, 1990). Those works show that in identifying vowels, listeners use information from three main sources: 1) The vocalic nuclei; 2) The formant transitions into and out of the vocalic nuclei; and 3) Temporal parameters related to vowel duration.

Another possible source for vowel identification, i.e., the interaction between consonant and vowel, was studied by van Son and Pols (1995). Their results indicate that identification of vowels is also influenced by speech segments beyond the boundaries of the vocalic transitions to neighboring segments. They also found that vowel identification improves more in CV-type tokens than in VC-type tokens, the offset parts of the tokens playing only a minor role in reducing the error rate.

A possible explanation for that improvement might be the CV coarticulation present in the tokens, i.e., significant vowel information may be present in the consonant segments prior to vocalic onset, this information helping in the recognition of the vowel.

In this paper we explore the influence of coarticulation in vowel identification for a series of fricative-vowel syllables of natural two-syllable Spanish words, including also a voiceless affricate. The coarticulatory effects of fricative-vowel syllables are well documented in the literature (see for instance Yeni-Komshian & Soli, 1981), particularly the presence of significant vowel information in the fricative noises. Our purpose was two-fold: First, to assess whether the presence of the fricative noises contribute to enhance vowel perception; and second, to explore the role played by coarticulation in that interaction.

2. MATERIALS AND METHOD

The tokens in our study correspond to citation form two-syllable natural Spanish words whose first syllable was formed by the combination of a Spanish voiceless fricative with one of the five Spanish vowels (/a,e,i,o,u/). The voiceless fricatives are /s,ʃ,f,θ/ and /x/. The voiceless affricate /ʧ/ was also included. The isolated words were pronounced by 5 males and 5 females, all Spanish native speakers (Galician Region). The total number of stimuli was 300=5 vowels × 6 fricatives × 5 speakers × 2 sexes.

All tokens were recorded in a normal office at the Faculty of Physics with a Rion microphone (type UC-53A), the whole process being supervised by one of the authors to ensure that tokens were natural and correctly pronounced. Then, tokens were sampled at 20 kHz using a DT-2801-A card of 12 bits of precision, and band pass filtered with cutoff frequencies of 100 Hz and 9.2 kHz.

Fricative noise and the first 51.2 ms of the following vowel were isolated by means of visual, auditory and spectral inspection.

3. PERCEPTUAL EXPERIMENTS

Perception tests were carried out in order to assess the importance of the phonetic FV integration for the recognition of vowels. 12 subjects acted as listeners for course credits (Spanish native speakers). The perceptual tests were carried out over two segments: a) In condition V, 51.2 ms of the vowel beginning at the vocalic onset after the fricative noise; b) In condition FV, the whole fricative noise + the following 51.2 ms of the vowel mentioned above.

Listeners were instructed in the nature of the perceptual experiments and in the use of the computer program which controls the whole process. Tokens were presented to the listeners through SONY MDR-CD570 headphones in random order. For each stimulus one repetition is allowed, after which the listener has to choose one of the possible options: /a,e,i,o,u/ and “other”.

	/a/	**/e/**	**/i/**	**/o/**	**/u/**	**other**
/a/	78.5	8.2	0.3	7.2	0.0	5.8
/e/	1.1	80.1	11.7	2.4	1.8	2.9
/i/	0.0	11.3	87.8	0.0	0.1	0.8
/o/	3.6	6.8	1.0	67.2	13.3	8.1
/u/	0.1	0.6	5.7	5.8	83.9	3.9
/a/	96.0	2.5	0.0	1.2	0.0	0.3
/e/	1.0	89.0	6.4	0.4	1.0	2.2
/i/	0.0	5.7	93.9	0.0	0.1	0.3
/o/	5.0	2.4	0.3	78.0	11.7	2.6
/u/	0.0	0.1	0.6	1.5	95.6	2.2

Table 1: Confusion matrices. Top: V condition, bottom: FV condition.

	/a/	**/e/**	**/i/**	**/o/**	**/u/**	**TOTAL**
/θ/	79.2	80.0	92.5	77.5	95.0	84.8
/f/	84.2	89.2	81.7	85.8	85.0	85.2
/s/	90.8	85.0	93.3	66.7	92.5	85.7
/ʃ/	57.5	68.3	90.0	37.5	72.5	65.2
/x/	95.8	84.2	93.3	89.2	90.0	90.5
/ʧ/	63.3	74.2	75.8	46.7	68.3	65.7
TOTAL	78.5	80.2	87.8	67.2	83.9	79.5
/θ/	93.3	91.7	97.5	75.8	98.3	91.3
/f/	98.3	89.2	91.7	85.0	93.3	91.5
/s/	100.0	80.8	99.2	77.5	96.7	90.8
/ʃ/	87.5	83.3	95.8	64.2	95.0	85.2
/x/	100.0	97.5	95.8	90.8	95.0	95.8
/ʧ/	96.7	91.7	83.3	75.0	95.0	88.3
TOTAL	96.0	89.0	93.9	78.1	95.6	90.5

Table 2: Percent of correct responses for each fricative and vowel. Top: V condition, bottom: FV condition.

3.1. Results

Confusion matrices for the V and FV conditions are showed in Table 1. In condition V, listeners correctly identified 79.5% of the five vowels. Adding the fricative noise (condition FV) improved the recognition of the vowels, raising the correct percent to 90.5%. The percent of correct responses for each fricative can be seen in Table 2. For both conditions, vowels in the context of /ʃ/ and /ʧ/ have the lowest identification percents, whereas vowels in the /x/ context attain the highest identification rates. /o/ is the worst identified vowel, particularly in the context of /ʃ/ and /ʧ/.

Analysis of variance on correct responses with fricative, vowel, sex and condition as factors, showed a significant main effect for condition ($F(1, 598) = 52.0, p < 0.0005$) indicating that there is a significant improvement in vowel recognition when the fricative is added. Significant main effects for fricative ($F(5, 594) = 14.8, p < 0.0005$) and vowel ($F(4, 595) = 18.4, p < 0.0005$) altogether with the fricative × vowel ($F(20, 579) = 2.3, p < 0.001$), fricative × condition ($F(5, 594) = 4.7, p < 0.0005$), fricative × sex ($F(5, 594) = 3.2, p < 0.008$) and vowel × sex ($F(4, 595) = 8.1, p < 0.0005$) interactions showed up. No other interactions were significant.

In condition V, both the fricative context ($F(5, 294) = 15.1, p < 0.0005$) and the vowel ($F(4, 295) = 8.8, p < 0.0005$) showed a significant effect on listeners' responses, plus significant fricative × vowel ($F(20, 279) = 1.8, p < 0.02$) and vowel × sex ($F(4, 295) = 7.1, p < 0.0005$) interactions. The effect of the sex was only significant for /a/ ($F(1, 118) = 5.3, p < 0.02$) and /i/ ($F(1, 118) = 20.8, p < 0.00005$), whereas the effect of vowel was only significant for women: ($F(4, 145) = 10.9, p < 0.00005$). A Scheffé test revealed that women's /i/ was recognized better than /o/. In the case of the significant fricative × vowel interaction, the effect of the vowel was only significant for /s/ ($F(4, 45) = 3.37, p < 0.02$) and /ʃ/ ($F(4, 45) = 6.02, p < 0.0006$, /o/ is significantly worse identified than /i/). The effect of the fricative was only significant for the vowels /a/ ($F(5, 54) = 5.80, p < 0.0002$), /o/ ($F(5, 54) = 7.33, p < 0.00005$) and /u/ ($F(5, 54) = 3.03, p < 0.02$). A Scheffé test revealed that vowels were best recognized in the context of /x/, while identification was worst in the context of /ʃ/ and /ʧ/.

In condition FV only a significant effect for the vowel showed up ($F(4, 295) = 11.5, p < 0.00005$). This effect was due mainly to the vowel /o/, which was very poorly recognized, especially in the contexts of /ʃ/ (64.2%) and /ʧ/ (75.0%). This vowel was even more poorly recognized when the fricative was removed (condition V) attaining very low percent correct scores (37.5% in /ʃ/ context, and 46.7% in the /ʧ/ context for the V condition).

The analysis of the perceptual experiments revealed a significant improvement in condition FV with respect to condition V, vowel /o/ being poorly identified in both conditions. For condition V, women stimuli attained lower percents than those of men. The effect of the context was also significant in this condition, vowels in the context of /ʃ/ and /ʧ/ being poorly identified whereas vowels in the context of /x/ attained high correct percents.

Fricative noise alone certainly may contribute to vowel identification to a certain extent because of coarticulation. A perceptual test in condition F was carried out to assess whether there exist enough vowel cues in the fricative noise to explain the improvement of condition FV with respect to condition V. Two experienced listeners evaluated the tokens, trying to identify the vowel associated to a particular fricative noise by listening to only the fricative noise (condition F).

For condition F, results were analyzed in a different way. First, cases were divided into two groups by means of a statistical procedure: 1) Those for which the vowel identification improves in the FV condition with respect to the V condition, and 2) Those for which there is no improvement. If coarticulation is the main factor explaining the improvement in the FV condition with respect to the V condition, we expect that the percent of cases which belongs to group 1 and for which the vowel was correctly identified by both listeners in condition F should be high. That would indicate that the coarticulatory influence of the vowel on the fricative would explain the improvement in condition FV with respect to condition V. Nevertheless, in only 7.7% of the cases of group 1 the vowel was correctly recognized from the fricative noise alone (condition F). For group 2, in 19.7% of the cases the vowel was correctly recognized from the fricative noise alone. Thus, coarticulation explains a little percent of the improvement in the FV condition with respect to the V condition.

4. ACOUSTIC ANALYSIS

For the acoustic analysis the fricative was represented by three windows of 25.6 ms: one at the begining of the fricative noise, one at the middle, and one at the ending of the fricative noise. The 51.2 ms of vowel was represented by two consecutive windows of 25.6 ms. The outputs of filter band spectra were computed for every window given a set of 23 outputs per window. The filter bank integrates the spectra with a mel-frequency scale in the manner described by Davis *et al*. (1980).

Two analysis were carried out on the filter outputs: 1) A Linear Discriminant Analysis (LDA), both for the V condition (23 outputs × 2 windows) and for the FV condition (23 outputs × 5 windows), along with the correlation of this acoustic representation and the perceptual representation for both conditions; 2) A Principal Components Analysis (PCA) for the V condition. Only 18 filter outputs per vocalic window, which integrate the spectra from 200 Hz to 4.5 kHz, were considered for the PCA.

4.1. Linear Discriminant Analysis

LDA is a classification procedure which involves forming linear combinations of the independent variables to determine the classification group for each case. The number of classification groups must be indicated to the procedure. From the distribution of cases and groups in the classification space it is possible to compute for each case *a posteriori probabilities* (APP) of membership in each group using Bayes' theorem.

In the V condition a LDA was performed with the 46 (23 outputs × 2 windows) output filters as independent variables attaining a correct classification percent of 93.7%. In the FV condition a LDA was performed with the 115 (23 outputs × 5 windows) output filters as independent variables, attaining a correct classification percent of 99.3%. The most interesting outcome is that the LDA is able to integrate the acoustic properties of both segments in the FV condition and so the correct classification percent in the FV condition improves with respect to the V condition.

In order to correlate both classifications, the APP were computed from the LDA: each case is represented by a vector of probabilities of membership in each group. Perceptual experiments represented each case by a vector of perceptual distances to each vowel. The overall correlation coefficient for each condition was computed over the classification vector and the perceptual vector formed by 1500=300 cases × 5 vowels. The correlation coefficient for each vowel and condition was computed over the 300 cases vector. Overall correlation coefficients were 0.95 for the FV condition and 0.89 for the V condition. Correlation coefficients for each vowel were: a) For the V condition: 0.92 for /a/, 0.85 for /e/, 0.93 for /i/, 0.83 for /o/ and 0.91 for /u/; and b) For the FV condition: 0.97 for /a/, 0.95 for /e/, 0.97 for /i/, 0.93 for /o/ and 0.95 for /u/.

4.2. Principal Components Analysis

PCA is a procedure for estimating which dimensions in the classification space contribute to the classification of the tokens to a larger extent. This involves computing the eigenvalues and eigenvectors from the covariance matrix of the independent variables and then, selecting those which explain a larger percent of the variance. It has been shown that the first two dimensions of such analysis are highly correlated with the first two formants of the vowels (Bakkum *et al.*, 1993; Pols *et al.*, 1969).

Before PCA is applied, the independent variables are submitted to a normalization procedure in the manner described by Bakkum *et al.* (1993). In order to have a significant number of tokens, ten additional speakers (five men and five women) were included, allowing for 300 new tokens following the guidelines described in section 2. Thus, 600 tokens = 5 vowels × 6 fricatives × 10 speakers × 2 sexes were used in the PCA. Our independent variables are the 18 filter outputs per vocalic window, which integrate the spectra from 200 Hz to 4.5 kHz. The normalization procedure is as follows: first, the level in decibels of every filter output is calculated; second, as differences in overall level are not of interest, all calculated spectra are level-normalized by substracting the average level of all 18 bands from the separate band levels; and third, as speaker-specific characteristics determined by the size of the vocal tract and the shape and functioning of the sound-production source are not of interest, a speaker normalization procedure was applied. The speaker normalization procedure is as follows: for each speaker and fricative context, the average of all five vowels calculated spectra was substracted from the five separate spectra. Prior to the speaker normalization procedure, the average of the calculated spectra of the two vocalic windows was computed. Then, the speaker normalization procedure was applied to the three calculated spectra. The PCA of the two vocalic windows will be used to obtain vowel trajectories, whereas the PCA of the average of the vocalic windows will be used to obtain overall distributions of the vowels.

PCA was applied separately for each sex to the initial, final and average normalized vocalic windows to obtain a two dimensional subspace for the vowels. The two dimensions of the subspace were Varimax rotated, such that the variance of their direction cosines was maximal. The direction cosines indicated that the first dimension was most sensitive to spectral variations in the F2 region, whereas the second dimension was most sensitive to spectral variations in the F1 region.

The overall distribution of vowels are plotted as ellipses, where the long axes correspond with the directions in which the interindividual variance for each vowel is maximal and the orthogonal short axes represent the remaining variance in two dimensions. The centers of the ellipses represent the average values for all speakers of each group. These centers were computed for the initial and final vocalic windows and plotted as the extremes of the trajectories. The two dimensional plots show considerable overlapping among the different vowels. This overlap is more marked in the context of /ʃ/ and /ʧ/, and less marked in the context of /x/ (see for instance figures 1 and 2).

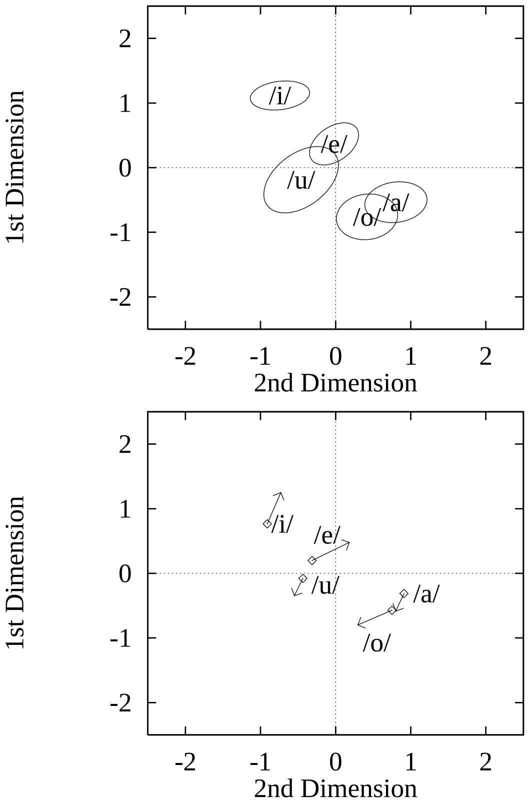

Figure 1: Distributions and trajectories in the two dimensional space of the vowels pronounced by men in the context of /ʃ/.

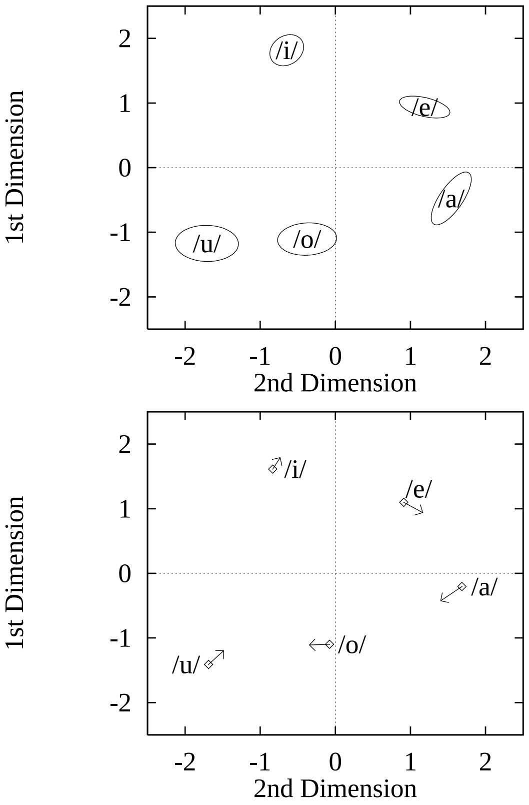

Figure 2: Distributions and trajectories in the two dimensional space of the vowels pronounced by men in the context of /x/.

5. CONCLUSION

Overall, the results show that fricative consonants play a role in the identification of vowels, which is not limited to the formant transitions within the vowel. It may be possible that the presence of those transitions, which help to preserve the acoustic continuity within the syllable, point to certain consonantal environments without which the perception of the vowel is incomplete.

The data presented in this paper clearly show that recognition of vowels in fricative FV context benefits from the presence of the fricative noise. This interaction between fricative and vowel enhances the perceptual characteristics of the vowel. The acoustic analysis show that the distribution of vowels varies a great deal with the fricative context, as is also the case with the perceptual identification of the isolated vowels. Coarticulatory effects, i.e. the influence of the vowel on the fricative, do not explain the improvement in vowel identification when the fricative is added.

6. REFERENCES

1. Bakkum, M.J., Plomp, R., Pols, L.C.W., "Objective analysis versus subjective assessment of vowels pronounced by native, non-native, and deaf male speakers of Dutch," *J. Acoust. Soc. Am.*, vol. 94, pp. 1989–2004, 1993.
2. Davis, S.B., Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 357–366, 1980.
3. Johnson, K., "Contrast and normalization in vowel perception," *Journal of Phonetics*, vol. 18, pp. 229–254, 1990.
4. Nearey, T.M., "Static, dynamic and relational properties in vowel perception," *J. Acoust. Soc. Am.*, vol. 85, pp. 2088–2113, 1989.
5. Peterson, G.E., Barney, H.L., "Control methods in a study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, pp. 175–184, 1952.
6. Pols, L.C.W., Van der Kamp, L.J.Th., Plomp, R., "Perceptual and physical space of vowel sounds," *J. Acoust. Soc. Am.*, vol. 46, pp. 458–467, 1969.
7. Strange, W., Jenkins, J.J., Johnson, T.L., "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.*, vol. 74, pp. 695–705, 1983.
8. van Son, R.J.J.H., Pols, L.C.W., "How transitions and local context affect segment identification," *IFA Proceedings*, vol. 19, pp. 51–69, 1995a.
9. Yeni-Komshian, G.H., Soli, S.D., "Recognition of vowels from information on fricatives: perceptual evidence of fricative-vowel coarticulation," *J. Acoust. Soc. Am.*, vol. 70, pp. 966–975, 1981.
10. Zahorian, S.A., Jagharghi, A.J., "Spectral-shape features versus formants as acoustic correlates for vowels," *J. Acoust. Soc. Am.*, vol. 94, pp. 1966–1982, 1993.