ACOUSTIC CUES FOR THE AUDITORY IDENTIFICATION OF THE SPANISH FRICATIVE /F/

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

This study deals with the distinction of the fricative noises of the spanish fricatives /θ/ and /f/. Previous studies revealed that fricative noises of both phonemes are perceptually similar, auditory identification being significantly dependent on contextual effects: /f/ in the /u/ context is well identified (about 85% correct identification rate), while in the /e/ context identification is much lower (about 60%). Identification of /θ/ is low for every vocalic context (about 60%). These effects were identical for both Hypo and Hyper forms of speech for which perceptual experiments were performed separately. The objective of this paper is to determine which acoustic properties of /f/ in the /u/ context make it a well defined phoneme for the two different forms of speech, in relation to the fricative noises of /θ/ in the /e,u/ contexts, and /f/ in the /e/ context. We conclude that the cues for the identification of the isolated fricative noise of /f/ seem to be in the low frequency region of the spectrum.

1. INTRODUCTION

The voiceless fricatives /θ/ and /f/ are one of the most confusable pairs of phonemes. Acoustic analysis of English fricatives have revealed that their spectra, which are highly similar, show no distinctive peaks (Hughes & Halle, 1956; Heinz & Stevens, 1961; Strevens, 1960), and as a result, fricative noise alone is not sufficient to perceptually distinguish between them, the vocalic part being necessary for their correct identification (LaRiviere *et al.*, 1975). The amplitude of the fricative noises of /f/ and /θ/ is also similar, and tends to be lower than that of the sibilant fricatives (see, for instance, Shadle & Mair, 1996). Very few studies have been devoted to the study of the Spanish fricatives. Studies on the Argentine Spanish have been carried out, but /θ/ is not part of the Argentine Spanish phonetic corpus (Gurlekian, 1981; Borzone de Manrique & Massone, 1981).

In a previous work (Feijóo *et al.*, 1998), we found that the auditory identification of /f/ and /θ/ is significantly dependent on contextual effects for two forms of speech, i.e. Hypo and Hyperspeech, the vocalic part playing an important role in the distinction between those phonemes. The fricative noises of both phonemes turned out to be perceptually similar, but their identification also depended on context: a) /f/ in the /u/ context is well identified (about 85% correct identification rate), while in the /e/ context identification is much lower (60%, roughly above chance); b) /θ/ is poorly identified in every vocalic context (about 60%). Those effects were identical for both Hypo and Hyperspeech, for which the perceptual experiments were performed separately.

Acoustic models for fricatives predict flat spectra for /θ/ (dental place of articulation) and /f/ (labio-dental). Since the length of the front cavity for those fricatives is within the range between 0 and 2 cms, depending upon whether the adjacent vowel is unrounded or rounded, resonances of the order of 8 kHz are expected. Since for /f/ in the /u/ context the front cavity is longer than for any other combination, we expect more notable resonances for this case than for the others, which can be considered to present null lengths for the front cavity.

The objective of this paper is to study which acoustic properties of /f/ in the /u/ context make it a perceptually well defined phoneme for the two different forms of speech, in relation to the fricative noises of /θ/ in the /e,u/ contexts, and /f/ in the /e/ context.

2. MATERIALS AND METHOD

The tokens correspond to citation form two-syllable natural spanish words. The stress falls on the first syllable which was the combination of /θ/ or /f/ with one of the spanish vowels /e,u/. Ten words for each fricative plus vowel combination were recorded with a Rion (type UC-53A) microphone. Two of the authors served as speakers. Although the Hypo/Hyper contrast is usually considered as a within-speaker variable, our view here is broader, in the sense that among a population of speakers, there are a certain number of them which tend to produce rather sloppy pronounciations of speech (weaker consonants, shorter and more reduced elements,etc. . .) while others tend to speak clearly. It is not that simple to make a clear speaker produce Hypospeech, since it is not his/her natural style, as is sometimes very difficult to make a sloppy speaker produce clear and correct pronunciations. That happened with our speakers. One allows for the Hyperspeech form and the other allows for the Hypospeech form. Thus, the total number of tokens was 80 = 2 fricatives × 2 vowels × 10 words × 2 forms of speech. All tokens were digitalized with a DT-2801-A card of 12 bits of precision, sampled at 20 kHz and band pass filtered with cutoff frequencies of 100 Hz and 9.2 kHz.

Tokens were normalized with respect to their maximum amplitude value. Subsequently, fricative noises were isolated by means of auditory, visual and spectral inspection. In case of doubt, the end of the fricative was determined as the point previous to the rising of the second formant (Soli, 1981).

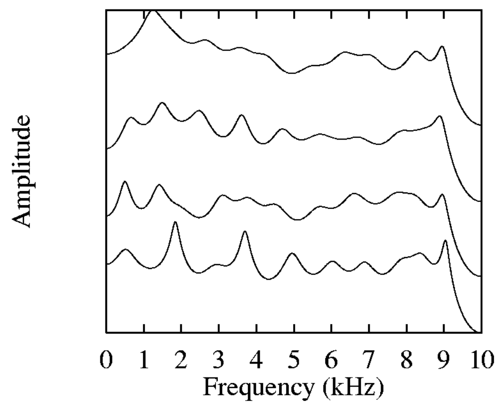

Figure 1: Average LPC spectra of the ten tokens of each of the four combinations for Hypospeech. From top to bottom: /f+u/, /f+e/, /θ+u/ and /θ+e/.

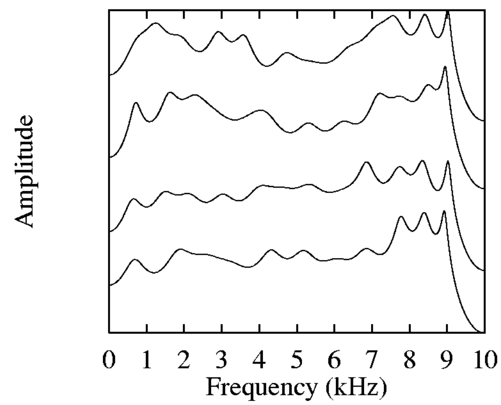

Figure 2: Average LPC spectra of the ten tokens of each of the four combinations for Hyperspeech. From top to bottom: /f+u/, /f+e/, /θ+u/ and /θ+e/.

In order to analyze the acoustic characteristics of the fricative noises they were represented by a window of 25.6 ms at the middle of the fricative noise. Shadle has shown (Shadle & Mair, 1996) that fricative spectra vary a great deal from fricative onset until fricative offset. We decided not to use spectra from fricative onset and offset since the exact begining of /f/ and /θ/ is sometimes difficult to locate due to its very weak amplitude, while the end of the fricative which corresponds to the time when the constriction is released, tends to be influenced by the proximity of the vowel.

3. ACOUSTIC ANALYSIS

As a first step in the acoustic analysis, differences between the two forms of speech were statistically analyzed. Significant duration differences showed up both for fricative noise ($F(1,78) = 249.0, p < 0.0005$) and whole word ($F(1,78) = 204.3, p < 0.0005$). The energy of the fricative noises was significantly different for the two forms of speech ($F(1,78) = 27.7, p < 0.0005$).

Acoustic models for fricatives predict flat spectrums for /θ/ and /f/. Since the length of the front cavity is in the range between 0 and 2 cm, depending upon whether the adjacent vowel is unrounded or rounded, resonances of the order of 8 kHz may be present (see for instance Heinz & Stevens, 1961; Hughes & Halle, 1956). Since for /f/ in the /u/ context the front cavity is longer than for any other combination, we expect more notable resonances for this case than for the others, which can be considered to present null lengths for the front cavity. This would hypothetically explain the fact that /f/ in the /u/ context is perceptually better defined for the two forms of speech than any other combination.

LPC spectral plots revealed resonances of the order of 8 kHz common to both forms of speech for each fricative in both vocalic contexts. A distinct spectral peak of the order of 1.5 kHz was present in the spectra of /f/ in the /u/ context, especially for Hypospeech (see figures 1 and 2).

This peak may be a consequence of coarticulation. The location of this low frequency peak was measured using LPC spectral plots and the spectrum itself. Formant frequencies of the first vocalic pulse were determined by means of LPC spectral plots.

For the acoustic analysis of the fricative noise, the center of gravity and the spectral tilt were calculated. The center of gravity of a spectrum (COG) is in a sense, the "mean" frequency. It is computed as:

$$C.O.G. = \frac{\int fE(f)df}{\int E(f)df} \tag{1}$$

where $E(f)$ is the spectral energy at frequency f. For fricatives the COG is inversely related to the size of the cavity in front of the noise source. The COG is also related to the spectral slope, the steeper the slope, the lower the COG.

The spectral tilt is the slope of the linear regression of the spectra in decibels. To compute the spectral tilt, the spectral peaks were emphasized by weighing each squared local error by the local linear amplitude value. Thus, the squared error E^2 minimized in the linear regression was:

$$E^2 = \sum_i |\mathcal{F}(\omega_i)| \left(10 \log |\mathcal{F}(\omega_i)| - a_0 - a_1 y\right)^2 \tag{2}$$

where $\mathcal{F}(\omega_i)$ is the local linear amplitude value, and a_0 and a_1 are the y intercept and slope of the linear regression function.

3.1. Results

An analysis of the COG and spectral tilt was carried out for each fricative in each vocalic context. Differences between the two forms of speech were significant both for COG and spectral tilt, COG being lower for Hypospeech than for Hyperspeech. The spectral tilt was negative for Hypospeech and positive for Hyperspeech (see figures 3 and 4).

This results indicate that the two forms of speech are acoustically different, Hyperspeech having more spectral energy at high frequencies, while for the low frequency region Hypospeech has more spectral energy. Nevertheless, in a previous work (Feijóo *et al.*, 1998) was concluded that clear speech is not necessarily

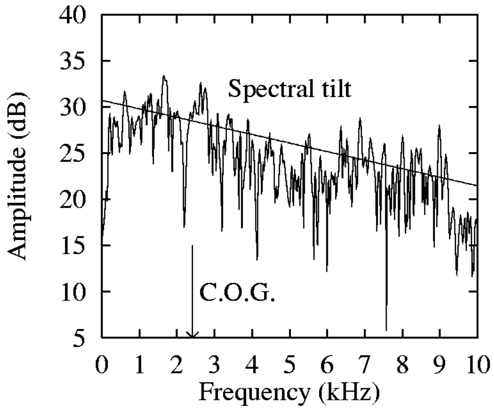

Figure 3: Spectral tilt and C.O.G. of the spectrum of Hypospeech (/f/ in the context of /u/).

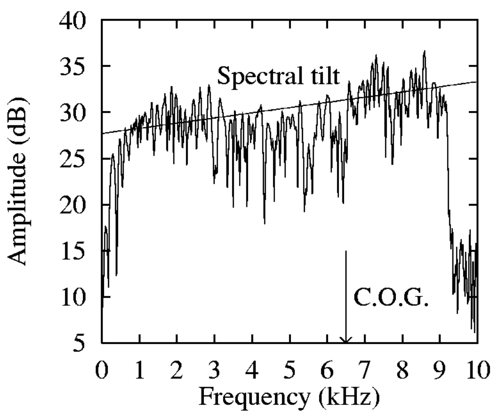

Figure 4: Spectral tilt and C.O.G. of the spectrum of Hyperspeech (/f/ in the context of /u/).

more intelligible than Hypospeech, which is in contrast to other researchers' results (see for instance Lindblom, 1996): it was found that the perceptual identification of these fricatives was not significantly different for both forms of speech, /f/ in the /u/ context being slightly better identified for Hypospeech (87.9%) than for Hyperspeech (83.6%).

An acoustic analysis with the center of gravity as variable and "combination" as factor was carried out separately for each form of speech. The four levels of the factor were the four possible fricative + vocalic context combinations. For the Hypospeech form the differences were significant ($F(3,26) = 10.3, p < 0.0001$). A Scheffé test revealed that the COG of /f/ in the context of /u/ was significantly different from the COG of /θ/ in the same context, whereas for the Hyperspeech form no significant effect showed up.

The same analysis with spectral tilt as variable revealed no significant effect for the Hyperspeech form and a significant effect for the Hypospeech form ($F(3,36) = 10.0, p < 0.0001$), spectral tilt of /f/ in the /u/ context being significantly steeper than the spectral tilt of the other combinations (Scheffé test), which indicates that the spectral energy in the low frequency region is larger for /f/ in the /u/ context than for any other combination.

Since the most distinctive feature of the spectrum of /f/ in the context of /u/ (in relation to /f/ in the context of /e/ and /θ/ in the contexts of /e,u/) is a low frequency spectral prominence, the relation of that frequency with the vowel formants was investigated. Since the resonance of the front cavity cannot explain the presence of that low frequency spectral prominence, it was hypothesized that the spectral prominence may be related to the frequency locations of the /u/'s F2 and F3, due to coarticulation. The frequency location of the peak for the fricative noise of /f/ in the context of /u/ was determined through visual inspection of the spectra, together with the LPC-spectra. The frequency of the prominence was manually located. The location of the peak falls between F2 and F3 of /u/. Mean values of the formants for Hypospeech were 458 Hz for F1, 1.0 kHz for F2, 2.4 kHz for F3 and 1.5 kHz for the peak. For Hyperspeech they were 496 Hz for F1, 995 Hz for F2, 2.5 kHz for F3 and 1.9 kHz for the peak. Then, the Pearson correlation between the location of the peak and the /u/'s F2 and F3 was calculated. Only for Hypospeech, the correlation coefficient attained a significant value of $r = -0.70$ (two-tailed $p = 0.02$) for the correlation between the peak and F3. Then, there is a certain inverse relationship between the peak location and F3 in Hypospeech, although it is not easy to determine the possible source of that relationship.

In order to assess the importance of the low frequency spectral resonances for the better identification of /f/ in the /u/ context versus the other combinations, two perceptual experiments were carried out. The cues for the identification of the place of articulation of /f/ seems to be in the low frequency region, especially in the context of /u/. Thus, two conditions were considered: 1) Fricative noises were low pass filtered at 3 kHz, and 2) Fricative noises were high pass filtered at 3 kHz.

The original words were band pass filtered with cutoff frequencies of 100 Hz and 3.0 kHz, which corresponds to condition 1, and cutoff frequencies of 3.0 kHz and 9.2 kHz, which corresponds to condition 2. The atenuation of the stopbands was about 70 dB and the transition bands were less than 60 Hz wide. Fricative noises were isolated following the procedure of section 2.

In view of the fact that filtered tokens may sound as unnatural phonemes or noises two experienced listeners carried out the perceptual experiments. They were performed separately for Hypo and Hyperspeech tokens. The stimuli was presented through SONY MDR-CD570 headphones in random order. For each stimulus one repetition was allowed.

In experiment 1, both listeners agree that every token was perceived as natural sounding /f/ stimuli, whereas in experiment 2, every token was perceived as strident and noisy /ʃ/-like phonemes. This indicates that the cues for the identification of the place of articulation of /f/ may be in the low frequency region of the spectrum, i.e. below 3 kHz.

4. DISCUSSION

The results from the acoustic analysis seem puzzling. On the one hand, the cues for the identification of the place of articulation of /f/, especially in the context of /u/, seem to be in the low-frequency region of the spectrum. This is true for the two forms of speech but for Hypospeech, the analysis of the spectral tilt showed a more prominent low-frequency region for /f/ in the /u/ context than for /f/ in the /e/ context and /θ/ in the /e/ and /u/ contexts. On the other hand, the spectrum of /f/ in the /u/ context presents a distinct low-frequency peak for the Hypo form of speech which has a difficult interpretation.

Heinz *et al.* (1961) found that "... some of the /f/ spectra were characterized by broad low-frequency noise in addition to the high-frequency peaks ...". They considered that a possible source of this low-frequency noise could be the turbulence at the lips beyond the constriction, this source being relatively uncoupled to the vocal-tract cavities.

It is well know that the air is forced through the constriction at high velocity, and turbulent flow occurs in the vicinity of the constriction. Since the place of articulation of /f/ is very close to the lips, the air flow is radiated at high energy in the case of this phoneme (and also in the case of /θ/ to a lesser extent). This would explain the presence of a broad low-frequency noise in the case of /f/. Besides, the radiated energy could be higher for /f/ in the context of /u/ because of lip-rounding. Lip-rounding provides the vocal tract with a short front cavity of the order of 1 or 2 cm. This cavity, apart from being the reason of a soft spectral resonance of the order of 8 kHz, amplifies the radiated energy because it is a poor acoustic resonator tube. This would explain that /f/ in the /u/ context is a perceptually well defined phoneme with respect to the fricative noises of /f/ in the /e/ context and /θ/ in the /e,u/ contexts for the two forms of speech. LPC-spectrum did not show clearly this differences because pre-emphasis was applied. For Hyperspeech, the values of the C.O.G. and spectral tilt are controlled by the high-frequency spectral energy and thus, they obscure the presence of this broad low-frequency energy. For Hypospeech, this global acoustic characteristics are controlled by the distinct low-frequency peak. Thus, to determine the importance of the broad low-frequency energy, Hyperspeech tokens and detailed low-frequency cues seems more suitable.

For fricatives, the sound source occurs at the point of constriction in the vocal tract. The vocal tract behind the point of constriction (back cavity) has less influence on the acoustics of the sound than the part of the vocal tract in front of the constriction, the noise spectrum being mainly controlled by whatever resonance chamber exists in the front cavity. The influence of the back cavity on the spectrum of the sound depends on the size of the constriction's opening (Olive *et al.*, 1993). Thus, for a vocal-tract configuration with a narrow constriction the output spectrum is approximately unchanged if the back cavities are neglected (Heinz *et al.*, 1961).

This simplification is valid for Hyperspeech fricatives and their spectrum does not show any peak owing to the coupling of the front and back cavities. For Hypospeech, the vocal tract is not so tightly constricted and the front/back cavity coupling can not be neglected. Fant (1970) found an extra soft formant at 1.5 kHz in the spectrum of /f/ which he attributes to a large coupling to the trachea system. In our case, the low-frequency peak of /f/ in the /u/ context is more marked. The effect of lip-rounding provides the vocal tract with a front cavity which allows the front/back cavity coupling. For /f/ the tongue tip is not required to form the constriction: the back cavity close to the place of articulation is bigger in the context of the back vowel /u/ than in any other context. Thus, because of back/front cavity coupling an extra formant showed up in the /f+u/ spectrum. This extra formant will be in a different location in relation to the vocalic formants, but its location can be related to the location of some of the vocalic formants because of coarticulation. In our case, the extra formant was correlated with F3. This may be due to the F3-dependency of the cavity system anterior to the source (Fant, 1970).

The importance of the low frequency region of the spectrum in the identification of /f/ is being corroborated by the results of new experiments with Hyperspeech fricatives, for which coupling resonances may be neglected. Low frequency energy, though, is still present in the /f/ spectra in the context of rounded vowels.

5. REFERENCES

1. Borzone de Manrique, A.M., Massone, M.I., "Acoustic analysis and perception of the Spanish fricative consonants," *J. Acoust. Soc. Am.*, vol. 69, pp. 1145–1153, 1981.
2. Fant, G., "Acoustic theory of speech production," *Mouton & Co.*, 2nd printing, 1970.
3. Feijóo, S., Fernández, S., Balsa, R., "Context effects in the auditory identification of the Spanish fricatives /f/ and /θ/: Hyper and Hypospeech," *Proceedings of the ICA/ASA '98*, pp. 2025–2026, 1998.
4. Gurlekian, J.A., "Recognition of the Spanish fricatives /s/ and /f/," *J. Acoust. Soc. Am.*, vol. 70, pp. 1624–1627, 1981.
5. Heinz, J.M., Stevens, K.N., "On the properties of voiceless fricative consonants," *J. Acoust. Soc. Am.*, vol. 33, pp. 589–596, 1961.
6. Hughes, G.W., Halle, M., "Spectral properties of fricative consonants," *J. Acoust. Soc. Am.*, vol. 28, pp. 303–310, 1956.
7. LaRiviere, C., Winitz, H., Herriman E., "The distribution of perceptual cues in English prevocalic fricatives," *Journal of Speech and Hearing Research*, vol. 18, pp. 613–622, 1975.
8. Lindblom, B., "Role of articulation in speech perception: Clues from production," *J. Acoust. Soc. Am.*, vol. 99, pp. 1683–1692, 1996.
9. Olive, J.P., Greenwood, A., Coleman, J., "Acoustics of american english speech," *Springer-Verlag*, 1993.
10. Shadle, C.H., Mair, S.J., "Quantifying spectral characteristics of fricatives," *Proc. Intl. Conf. on Spoken Lang. Proc., ICSLP*, vol. 3, pp. 1521–1524, 1996.
11. Soli, S.D., "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," *J. Acoust. Soc. Am.*, vol. 70, pp. 976–984, 1981.
12. Strevens, P., "Spectra of fricative noise in human speech," *Language and Speech*, vol. 3, pp. 32–49, 1960.