Perceptual identification of fricatives in low-pass filtered C and CV segments

Santiago Fernández, Sergio Feijóo

Departamento de Física Aplicada
Universidad de Santiago de Compostela
fasanti@usc.es

Abstract

The influence of spectral bandwidth on the perceptual identification of fricatives has been studied. Stimuli were CV syllables formed by /θ, f, s, ʃ/ and /a, e, i, o, u/. Two conditions were tested: 1) isolated fricative noise (C condition), and 2) fricative noise plus 100 ms of the following vowel (CV condition). Several low-pass filtered cutoff frequencies (16, 11, 8, 5.5, 4 and 3 kHz) were considered. Multidimensional scaling provided us with a representation of the perceptual space of every fricative for every spectral bandwidth which helps us to interpret the results of the experiment. Results show that recognition rates are 10% higher in the CV condition than in the C condition for the widest bandwidths and approximately 20% higher for the narrowest bandwidths. Fricatives are well identified with a bandwidth of 6 kHz in the C condition. In the CV condition, bandwidth can be reduced up to 4 kHz without a significant loss of performance.

1. Introduction

In CV syllables the spectral properties of the fricative noises alone seem to be the primary acoustic cues to the perceptual identification of fricatives (see for instance [1]). However, the vocalic part of the syllable is also considered to be important for the distinction between /θ/ and /f/. Some other acoustic characteristics of the friction noise have been studied, for example, duration and amplitude. Although these properties do not play a major role in the identification of place of articulation of fricatives, the amplitude of non-sibilant fricatives tends to be lower than that of the sibilant fricatives, and the duration of the fricative noise required for the perceptual identification of /θ/ is longer than for the rest.

The spectral characteristics of Galician fricatives are similar to those reported for English fricatives. Note that the descriptions of fricatives' spectra are quite different amongst researchers. In general, it is considered that /s/ has spectral energy around 4 and 5 kHz (and above those frequencies), /ʃ/ around 2.5 and 3 kHz, while /f/ and /θ/ have flat spectra with some high frequency peak (around 8 kHz for /f/). Some sex-related differences have also been reported, particularly for /s/ and /ʃ/. The spectra of these fricatives produced by males is displaced down in frequency as compared with those produced by females. There has also been reported a similar effect of the quality of the following vowel on fricatives: the spectrum of the fricative noise produced in the context of a rounded vowel is displaced down in frequency as compared with that produced in the context of an unrounded vowel.

These effects clearly influence the bandwidth of the spectrum neccessary for a correct identification of the fricatives. A study about the perception of fricatives for different bandwidths looks interesting in order to delimite the spectral information necessary for their identification. Moreover, it would be interesting to know what is the effect (in CV syllables) of adding the vowel to the isolated fricative noise for different bandwidths, since this could have implications for ASR systems working with narrow-bandwidth signals.

Recently, Stelmachovicz *e*t al. [2] studied the extent to which high frequencies can provide useful information for fricative perception for normal-hearing and hearing-impared children and adults. The stimuli were low-pass filtered versions of CV and VC syllables formed by /f/, /θ/ or /s/, and /i/. Cutoff frequencies were 2, 3, 4, 5, 6 and 9 kHz. The studied focused on the effect of the stimulus bandwidth on the perception of /s/ in the four kind of aforementioned listeners and thus, confusion matrices were not showed. /f/ and /θ/ were included only because in a pilot study they found that low-pass filtered versions of /s/ tend to be perceived as one of those fricatives. Nevertheless, these results are apparently not available and they seem to be quite interesting if we consider that high frequency energy include important cues for the identification of both /θ/ and /f/. Feijóo and Fernández [3] studied the perceptual and acoustic identification of /f/, concluding that low-frequency energy might contain sufficient cues for the identification of this fricative, especially in the context of /u/.

The objective of this paper is to study the influence of the spectral bandwidth on the perceptual identification of Galician voiceless fricatives in CV syllables. It is usually assumed that a sampling rate of at least 20 kHz is necessary to preserve the important characteristics of the fricative noises, although spectral energy up to 16 kHz (corresponding to a sampling rate of 32 kHz) is present in some fricative noises. On the other hand, some speech recognition systems use a sampling frequency of 16 kHz, which means that some spectral characteristics of the frication noise may be lost during the sampling process. The case of telephone speech is even more exacerbated due to severe reduction in spectral bandwidth. It would be interesting to know which fricatives suffer more from the spectral bandwidth reduction and if the inclusion of the vowel helps to improve identification rates.

2. Materials

The tokens for this study were CV syllables formed by the combination of one of the Galician voiceless fricatives /θ, f, s, ʃ/ with one of the five vowels /a, e, i, o, u/. Two native speakers of Galician (Northwest Spain), one male and one female, pronounced the syllables carefully but in a natural way. Therefore the tokens may be considered to correspond to clear speech (hyperspeech). Hyperspeech tokens include high-frequency energy to a greater extent than not so carefully pronounced tokens [3]. Recorded stimuli showed an important amount of energy at high frequencies, as compared with casually pronounced stimuli. The distance to the microphone (Shure BG3.0 omnidirectional), approximately 15 cm, was selected so as to take advantage of the whole 16 bits quantization range of the digital card (MAXI-Sound 64). Then, each speaker uttered the series of 20 CV syllables = 4 fricatives × 5 vowels, which was sampled at 32 kHz. According to previous experiments with large databases of fricatives the most relevant source of variance for the identification of fricative noises in CV syllables is vocalic context. Moreover this is only true for /f/. Neither speaker nor sex introduced significant differences regarding perception of the friction noises. Nevertheless, as different speakers may use different strategies, high frequency regions may be more important for some speakers than for others. This effect was minimized by using hyperspeech tokens produced by two speakers of different sex in different vocalic contexts, since gender and vocalic context are well known to introduce differences in the spectral content of the fricatives. While the use of a large data set would have provided us with a broad view of the processes involved, it nevertheless obscures particular details that may be important. A reduced and selected set, on the other hand, allows us to access more detailed information.

The speech signals of both speakers (i.e. two series of 20 CV syllables) were, firstly, band pass filtered with six different upper cutoff frequencies in order to obtain six different experimental conditions; secondly, normalized; and lastly, segmented. Let us take a closer look at each one of these steps.

The signals were recorded with a sampling frequency of 32 kHz and filtered at the Nyquist frequency. For the first condition the signals were high pass filtered with a cutoff frequency of 100 Hz in order to eliminate undesired low-frequency noises, such as breathing noises. For the following conditions they were band pass filtered with a lower cutoff frequency of 100 Hz and upper cutoff frequencies of 11, 8, 5.5, 4 and 3 kHz. These spectral bandwidths were chosen because they correspond to some of the most frequently used sampling rates: 32, 22, 16, 11, 8 and 6 kHz from the first to the sixth experimental condition respectively. Due to the large number of conditions it was necessary to limit the number of speakers in order to carry out the perceptual experiments. Frequencies above the cutoff frequency of each condition are not heard by the listeners because the attenuation in the stopbands was approximately 70 dB. The ripple of the passband was 0.005 dB at most. The transition bands were aproximately 32 Hz wide. A Chebyshev window was employed. The total number of tokens was 240 = 4 fricatives × 5 vowels × 2 sexes × 6 bandwidths.

Secondly, each of the 12 low-pass filtered series = 2 sexes × 6 LPF cutoff frequencies, consisting of 20 CV syllables, were normalized so as to reach their maximum value at 75% of the maximum quantization level in order to minimize differences in amplitude between speakers. Since the aforementioned ripple of the passbands was very low the energy of the stop bands was never substantially increased by the normalization procedure.

The last step involved segmenting the fricative noises from their corresponding syllables. The end of the friction noise was taken as the begining of the first clear vocalic pulse. Because the beginning of a vowel in fricative context is sometimes very noisy, in case of doubt, the spectrum of the conflicting region was examined looking for a steep rise of the second formant which marks the begining of the vowel. Lastly, the beginning and end of the friction noise segments were smoothed with a 10 ms long cosine-type window to prevent undesired clicks. The stimuli are: a) the fricative noises of the 240 tokens (henceforth C condition), and b) the fricative noises plus the first 100 ms of the following vowel (CV condition).

3. Method

The perceptual experiments took place in a quiet office at the Faculty of Physics. 37 Galician speaking students participated for course credits. The nature of the experiment was explained to them (to identify the fricative) as well as the whole procedure, but they did not know that some of the stimuli were filtered versions of other stimuli. For the C and CV condition separately, the 240 tokens were presented binaurally to each subject through SONY MDR-CD570 headphones in different random order. A random order was selected to limit the loudness cues that might be present if each low-pass filtered condition was tested separately. The presentation level was approximately 70 dBA. Subjects were provided with five different options, /θ, f, s, ʃ/ and *another sound*, in order to avoid guessing, especially in the case of the tokens corresponding to the lower low-pass filter (LPF) cutoff frequencies. The orthographic symbols used were: /θ/-z, /f/-f, /s/-s, /ʃ/-x.

The whole process was controlled by a computer program developed at our laboratory. The listeners performed the experiment at their own pace. They just had to mouseclick the appropriate button on the screen to listen to the next token, to repeat the present token (only one repetition is allowed) or to mark their response. Once the listener had finished the experiment, the responses were automatically saved as *response profiles* and the confusion matrices and recognition rates were computed. Response profiles were submitted to an analysis of variance and confusion matrices were submitted to a multidimensional scaling analysis.

Multidimensional scaling (MDS) is a mathematical tool that enables us to obtain the underlying dimensions which are perceptually relevant to the subjects. It permits to represent the stimuli spatially without a priori knowledge of their attributes. All we need is a set of measures of the similarity or dissimilarity between the stimuli. The mathematical algorithm will represent the stimuli taking into account that these perceptual distances must be kept. The researcher has to choose the dimensionality of that perceptual space and give an interpretation to those dimensions.

Several kinds of MDS algorithms are available. The choice depends to a large extent on the kind of data to be studied. Confusion matrices are the usual way of representing the relationship among stimuli with a kind of inverse distance or dissimilarity. Van der Kamp and Pols [4] found no significant differences in the vocalic perceptual space obtained with MDS either from similarity judgements or from confusion matrices. In our particular case, the differences among the six confusion matrices corresponding to the six LPF cutoff frequencies are of interest. Besides, differences amongst the stimuli that form those confusion matrices should be studied. Also, notice that confusion matrices show substantial asymmetry and response bias. Therefore an asymmetric individual differences model (ASYNDSCAL) was chosen (the term individual differences stands for differences among, in this particular case, the six confusion matrices). Thus, the perceptual map for the C and CV conditions will include every fricative for every LPF cutoff frequency. This is clearly a very simplistic model for the broad problem of perceptual identification and categorization. Nevertheless, our primary purpose was to find the most relevant perceptual dimensions to the subjects, which agrees with the fundamental idea of MDS.

4. Results

4.1. Analysis of variance

If every factor was taken into account for the analysis of variance (fricative, vowel, speaker/sex, LPF cutoff frequency and condition) just one token for each group would be available. Therefore, firstly, a one-way analysis of variance was carried out for every factor on arcsine transformed correct responses which leads to a better normal distribution and normalize the variance across performance levels. Every factor was significant except speaker/sex that will be discarded in the following analyses.

An ANOVA was carried out with condition, LPF cutoff frequency, fricative and vowel as factors. Main effects were all significant: fricative ($F(3, 476)=57.97$, $p<0.0005$), vowel ($F(4, 475)=21.25$, $p<0.0005$), LPF cutoff frequency ($F(5, 474)=71.82$, $p<0.0005$) and condition ($F(1, 478)=199.94$, $p<0.0005$). The following interactions were also significant: fricative × vowel ($F(12, 467)=8.44$, $p<0.0005$), fricative × LPF cutoff frequency ($F(15, 464)=20.26$, $p<0.0005$), fricative × condition ($F(3, 476)=14.28$, $p<0.0005$), vowel × condition ($F(4, 475)=4.17$, $p<0.003$), fricative × vowel × condition ($F(12, 467)=6.68$, $p<0.0005$) and LPF cutoff frequency × condition ($F(5, 474)=3.20$, $p<0.008$).

The effect of condition is highly significant but it depends on the particular fricative and LPF cutoff frequency. For each fricative an ANOVA was carried out with condition, LPF cutoff frequency and vowel as factors. For /θ/, condition ($F(1, 118)=109.41$, $p<0.0005$), vowel ($F(4, 115)=4.32$, $p<0.004$), LPF cutoff frequency ($F(5, 114)=20.09$, $p<0.0005$) and LPF cutoff frequency × condition ($F(5, 114)=4.48$, $p<0.002$) were significant. For /f/, condition ($F(1, 118)=27.21$, $p<0.0005$), vowel ($F(4, 115)=33.29$, $p<0.0005$), LPF cutoff frequency ($F(5, 114)=4.80$, $p<0.001$) and condition × vowel ($F(4, 115)=10.85$, $p<0.0005$) were significant. For /s/, condition ($F(1, 118)=39.78$, $p<0.0005$) and LPF cutoff frequency ($F(5, 114)=136.14$, $p<0.0005$) were significant. For /ʃ/, condition ($F(1, 118)=31.49$, $p<0.0005$) and vowel ($F(4, 115)=4.36$, $p<0.004$) were significant. Therefore, for /θ/, the effect of the condition if different for each LPF cutoff frequency. For /f/ and /s/ the effect of the condition is independent of the particular LPF cutoff frequency. These fricatives are always better identified in the CV condition than in the C condition, recognition rates decreasing as the LPF cutoff frequency decreases. The fricative /ʃ/ is also better identified in the CV condition than in the C condition, the effect of the LPF cutoff frequency being non-significant.

For each LPF cutoff frequency an ANOVA was carried out with condition, fricative and vowel as factors. For 16 kHz only the effect of condition ($F(1, 78)=20.88$, $p<0.0005$) and vowel ($F(4, 75)=3.18$, $p<0.02$) were significant. For 11 kHz the effect of condition ($F(1, 78)=20.48$, $p<0.0005$), fricative ($F(3, 76)=3.34$, $p<0.03$) and vowel ($F(4, 75)=3.06$, $p<0.03$) were significant. For 8 kHz the effect of condition ($F(1, 78)=22.84$, $p<0.0005$), fricative ($F(3, 76)=11.10$, $p<0.0005$) and vowel ($F(4, 75)=3.09$, $p<0.03$) were significant. For 5.5 kHz the effect of condition ($F(1, 78)=40.83$, $p<0.0005$) and fricative ($F(3, 76)=15.06$, $p<0.0005$) were significant. For lower LPF cutoff frequencies more significant effects show up. For 4 kHz, condition ($F(1, 78)=33.66$, $p<0.0005$), fricative ($F(3, 76)=19.16$, $p<0.0005$), vowel ($F(4, 75)=4.61$, $p<0.004$) and condition × fricative ($F(3, 76)=4.28$, $p<0.01$) were significant. For 3 kHz, condition ($F(1, 78)=71.70$, $p<0.0005$), fricative ($F(3, 76)=103.89$, $p<0.0005$), vowel ($F(4, 75)=7.5$, $p<0.0005$), condition × fricative ($F(3, 76)=12.80$, $p<0.0005$) and fricative × vowel ($F(12, 67)=3.58$, $p<0.001$) were significant. This analysis show that fricatives are better identified in the CV condition than in the C condition for all LPF cutoff frequencies except for 3 kHz, for which the difference between conditions depends on the particular fricative.

Once differences between conditions have been studied, the analyses focused on each particular condition. For the CV condition, one-way analysis of variance revealed that the effect of speaker/sex was again not significant. An ANOVA was carried out for the CV condition with fricative, vowel and LPF cutoff frequency as factors. Main effects as well as some interactions were significant: fricative ($F(3, 236)=20.36$, $p<0.0005$), vowel ($F(4, 235)=16.81$, $p<0.0005$), LPF cutoff frequency ($F(5, 234)=23.51$, $p<0.0005$), fricative × vowel ($F(12, 227)=11.50$, $p<0.0005$) and fricative × LPF cutoff frequency ($F(15, 224)=9.82$, $p<0.0005$). Scheffé test showed that /ʃ/ was better identified than the rest of fricatives. Also that fricatives in the context of /u/ and /o/ were better identified than fricatives in the context of /e/, and that fricatives in the context of /u/ were also better identified than fricatives in the context of /i/. Scheffé tests revealed that, for a LPF cutoff frequency of 3 kHz, fricatives were worse identified than for the rest of LPF cutoff frequencies.

To study further the importance of the significant interactions, an ANOVA was carried out for each fricative and LPF cutoff frequency separately. For /θ/, /f/ and /ʃ/ the only significant effect was vowel ($F(4, 55)=4.76$, $p<0.004$; $F(4, 55)=41.46$, $p<0.0005$ and $F(4, 55)=5.70$, $p<0.002$; respectively). Scheffé tests showed that: 1) /θ/ is better identified in the context of /o/ and /u/ than in the context of /a/; 2) /f/ is better identified in the context of /a/, /o/ and /u/ than in the context of /e/ and /i/. Besides /f/ is better identified in the context of /i/ than in the context of /e/; 3) /ʃ/ is worse identified in the context of /i/ than in any other context. For /s/ there was only a significant main effect of the LPF cutoff frequency, fricative /s/ being worse identified for 3 kHz than for higher LPF cutoff frequencies.

ANOVAs for each LPF cutoff frequency did not show significant effects either for the highest LPF cutoff frequencies (16 kHz, 11 kHz and 8 kHz) or for 4 kHz. For 5.5 kHz a significant main effect for fricative showed up ($F(3, 36)=5.83$, $p<0.005$), fricative /ʃ/ being better identified than fricative /f/. For 3 kHz there was a significant main effect of fricative ($F(3, 36)=36.38$, $p<0.0005$) and also a significant main effect of vowel ($F(4, 35)=5.12$, $p<0.005$). Fricative /s/ is worse identified than the rest of fricatives. The effect of the vowel was significant only for fricative /f/ ($F(4, 5)=9.26$, $p<0.02$). Fricative /f/ is better identified in the context of /o/ than in the context of /e/.

For the C condition, one-way analyses of variance revealed that the effect of the fricative was significant ($F(3, 236)=18.04$, $p<0.00005$) and a Scheffé test showed the following subgroups from the worst identified fricative to the best identified one: 1) /θ/, 2) /s/ and /f/, and 3) /ʃ/. The effect of the vowel was significant at the 0.05 level as well ($F(4,235)=2.79$, $p<0.027$) but the Scheffé test did not reveal significantly different groups. The effect of the LPF cutoff frequency was also significant ($F(5,234)=20.91$, $p<0.00005$) and the following subgroups showed up: 1) 3 kHz, 2) 4 kHz, 5.5 kHz and 8 kHz, and 3) 5.5 kHz, 8 kHz, 11 kHz and 16 kHz. Thus, the identification rate of the fricatives for a LPF cutoff frequency as low as 5.5 kHz was not significantly different from that for a higher LPF cutoff frequency, at least up to 16 kHz.

Therefore, spectral energy above 6–8 kHz is of little importance for the identification of place in fricatives, while spectral energy up to 5 or 6 kHz contains enough spectral cues for its correct identification. The effect of the speaker, which includes the effect of sex, was not significant. Therefore this factor was discarded and an ANOVA was carried out with fricative, vowel and LPF cutoff frequency as factors in order to obtain a more accurate estimate of the importance of each factor and its interactions. Every main effect was significant (fricative: $F(3{,}236)=51.61$, $p<0.0005$, vowel: $F(4{,}235)=9.40$, $p<0.0005$, and LPF cutoff frequency: $F(5{,}234)=51.26$, $p<0.0005$) along with the following 2-way interactions: fricative × vowel ($F(12{,}227)=3.68$, $p<0.0005$) and fricative × LPF cutoff frequency ($F(15{,}224)=12.16$, $p<0.0005$). Therefore, it is clear that the identification of the fricative is affected by the vocalic context as well as by the spectral bandwidth.

In order to gain an insight into these effects an ANOVA was carried out separately for each fricative with vowel and LPF cutoff frequency as factors. For /θ/ only the effect of the LPF cutoff frequency was significant ($F(5{,}54)=27.89$, $p<0.0005$) for which a Scheffé test revealed the following subgroups: 1) 3 kHz and 4 kHz, 2) 5.5 kHz, 8 kHz, 16 kHz and 11 kHz. The same happenned to /s/ ($F(5{,}54)=80.94$, $p<0.0005$) with subgroups: 1) 3 kHz, 2) 4 kHz and 5.5 kHz, 3) 5.5 kHz, 11 kHz, 16 kHz, and 8 kHz. For /f/, only the effect of the vowel was significant ($F(4{,}55)=13.29$, $p<0.0005$) with subgroups (from the context with the worst identification rate of the fricative to that with the best): 1) /a/ and /e/, 2) /e/, /i/ and /o/, and 3) /o/ and /u/. No significant effects showed up for /ʃ/. Therefore, /θ/ and /s/ are the fricatives most affected by the reduction in spectral bandwidth, especially for the lowest ones, whereas /f/ and /ʃ/ are less affected. The only fricative showing a major effect of the vocalic context is /f/ which is especially well identified in the context of /u/, this effect being constant across spectral bandwidths.

Lastly, an ANOVA was performed separately for each LPF cutoff frequency on arcsine transformed correct responses with fricative and vowel as factors. For the two highest LPF cutoff frequencies, i.e. 16 and 11 kHz, no significant effects showed up. For the other LPF cutoff frequencies, the effect of the fricative was always significant. For 8 kHz ($F(3{,}36)=7.16$, $p<0.002$) the following subgroups appeared: 1) /f/ and /θ/, 2) /θ/ and /ʃ/, and 3) /ʃ/ and /s/. For 5.5 kHz ($F(3{,}36)=10.48$, $p<0.0005$) the subgroups were: 1) /θ/ and /f/, 2) /f/ and /s/, and 3) /s/ and /ʃ/. For 4 kHz ($F(3{,}36)=26.75$, $p<0.0005$) they were: 1) /θ/, 2) /s/ and /f/, and 3) /ʃ/. For 3 kHz ($F(3{,}36)=123.78$, $p<0.0005$): 1) /s/ and /θ/, and 2) /f/ and /ʃ/. For this particular LPF cutoff frequency the interaction fricative × vowel was also significant ($F(12{,}27)=3.72$, $p<0.005$).

4.2. Confusion matrices

Table 1 shows the confusion matrices for each LPF cutoff frequency in the C and CV conditions separately. Chance level is 20%. For the C condition, reducing the spectral bandwidth generally lowered the identification rate, although there are some exceptions. The fricatives less affected by the reduction were /f/ and /ʃ/. For /θ/ the identification rate decreased continuously with the LPF cutoff frequency, implying that only twenty percent of the signals corresponding to this particular fricative were correctly identified with a LPF cutoff frequency of 3 kHz. /s/ is poorly identified only in the 3 kHz condition. Confusions primarily took place between the pairs /θ/-/f/ and /s/-/ʃ/, although /s/ was sometimes identified as /θ/, especially for the following LPF cutoff frequencies: 5.5, 4 and 3 kHz. Confusions tended to increase as the LPF cutoff frequency was decreased, but some other patterns of confusions also appeared, especially for the lower LPF cutoff frequencies which were more confusing.

Since the vocalic context had a significant effect in the analysis of variance, the identification rates were divided by context (results not shown). Both /θ/ and /s/ showed little variation with the vocalic context across bandwidths. The identification rate of /ʃ/ did not vary with the vocalic context, except for a LPF cutoff frequency of 3 kHz, this effect beginning to appear at a LPF cutoff frequency of 4 kHz. For that LPF cutoff frequency, the identification of /ʃ/ was better for the front vowels (/o/ and /u/) than for the back vowels (/a/, /e/ and /i/). Since the low frequency region of the spectrum of /ʃ/ seems to include very important cues for its identification, we expect lower identification rates for a LPF cutoff frequency of 3 kHz because those cues have been partially deleted. This is the case with /ʃ/ in the context of /a/, /e/ and /i/, but not for /o/ and /u/ because it seems that lip-rounding lowers partially the frequency of those important cues and therefore they are still present for a LPF cutoff frequency of 3 kHz, the identification rate being similar in this case and for higher LPF cutoff frequencies. Differences in the identification rate of /ʃ/ across vocalic contexts were clear, although the vocalic context effects were not significant. Notice that few cases where available for the analysis of variance ($F(4{,}5)=4.14$, $p<0.08$). /f/ presented the same contextual effect as /ʃ/ but the difference is that for /f/ it was already present for the highest LPF cutoff frequency (16 kHz). Therefore the explanation for /ʃ/ is not valid for /f/, and it looks even more complicated for this particular case. /f/ was especially well identified in the context of /u/ for all LPF cutoff frequencies. It was also well identified in the context of /o/ (except for a LPF cutoff frequency of 5.5 kHz) and in the context of /i/ for the three highest LPF cutoff frequencies (16, 11 and 8 kHz). This effect has been already studied by [3], who found that a low-frequency broad peak (centered at 1.5 kHz approximately) of /f/ in the context of /u/ could be the reason of the high identification rate for this particular fricative (sampling frequency of 20 kHz).

Adding the vowel to the stimuli in the CV condition improves recognition rates for every fricative. This improvement is approximately 10% for the highest LPF cutoff frequencies, and around 20% for the lowest LPF cutoff frequencies. For example, for 4 kHz (the bandwidth used in telephones), identification rates are 64.2% in the C condition and 84.5% in the CV condition. The only important drop in performance takes place for a LPF cutoff frequency of 3 kHz and fricative /s/ which is only accurately identified 24.5% of the times. Due to the presence of the vowel, which is widely assumed to have similar vocalic transitions for /s/ and /θ/, fricative /s/ is uncorrectly identified as /θ/ most of the times, while in the C condition /s/ obtained a lower identification rate (10.6%) and was uncorrectly identified both as /θ/ and as /f/. The presence of the vowel has also a great effect in the identification of /θ/ for a LPF cutoff frequency of 3 kHz. In the CV condition identification rate is 76.4% while in the C condition /θ/ is uncorrectly identified as /f/ most of the times, attaining a recognition rate as low as 19.2%.

To sum up, in the C condition fricatives are correctly identified with a LPF cutoff frequency as low as 5.5 kHz, while in the CV condition fricatives are well identified even with a LPF cutoff frequency of 4 kHz. Overall, fricatives in the CV condition attain 10% to 20% higher identification rates than in the C condition. The fricatives /f/ and /ʃ/ are the least influenced by the LPF cutoff frequency in the C condition. In the CV condition, every fricative is well identified despite the decrease in bandwidth except /s/ for a LPF cutoff frequency of 3 kHz. Con-

Table 1: *Confusion matrices and total recognition scores (in %) for each low-pass filter cutoff frequency. Results are presented for the C and CV conditions separately. In each matrix, rows indicate the intended fricative and columns indicate listener's responses.*

		C condition					**CV condition**				
		/f/	**/θ/**	**/s/**	**/ʃ/**	**another**	**/f/**	**/θ/**	**/s/**	**/ʃ/**	**another**
16 kHz	**/f/**	81.7	14.2	1.1	0.2	2.8	88.3	10.6	0.0	0.0	1.1
	/θ/	11.9	78.6	2.8	1.1	5.6	5.8	92.5	0.6	0.0	1.1
	/s/	0.0	3.6	86.7	8.0	1.7	0.0	1.9	94.4	2.6	1.1
	/ʃ/	0.0	0.3	11.7	87.5	0.5	0.0	0.3	1.7	98.0	0.0
					Total:	83.6				**Total:**	93.3
11 kHz	**/f/**	77.2	15.6	2.2	0.0	5.0	87.8	12.2	0.0	0.0	0.0
	/θ/	11.1	78.6	5.0	0.0	5.3	5.0	93.9	0.3	0.3	0.5
	/s/	0.0	3.3	86.1	8.6	2.0	0.0	0.6	96.1	2.5	0.8
	/ʃ/	0.0	0.0	9.4	90.6	0.0	0.0	0.0	3.3	96.4	0.3
					Total:	83.1				**Total:**	93.6
8 kHz	**/f/**	67.8	24.7	1.9	0.0	5.6	81.4	18.3	0.3	0.0	0.0
	/θ/	9.7	72.2	8.9	1.1	8.1	3.6	92.2	2.5	0.3	1.4
	/s/	0.0	2.5	90.6	5.5	1.4	0.0	1.1	95.0	3.1	0.8
	/ʃ/	0.0	0.6	9.4	89.4	0.6	0.0	0.0	2.8	96.9	0.3
					Total:	80.0				**Total:**	91.4
5.5 kHz	**/f/**	57.5	27.8	3.6	0.6	10.6	76.2	20.8	0.8	0.0	2.2
	/θ/	24.4	57.8	8.1	0.3	9.4	7.5	87.2	2.8	0.3	2.2
	/s/	0.6	4.7	77.2	13.1	4.4	0.0	1.9	95.0	2.8	0.3
	/ʃ/	0.0	1.1	9.4	88.9	0.6	0.0	0.0	3.9	95.8	0.3
					Total:	70.3				**Total:**	88.6
4 kHz	**/f/**	68.9	16.1	3.3	1.1	10.6	81.4	16.1	0.8	0.3	1.4
	/θ/	44.7	33.6	7.2	0.8	13.6	17.2	78.1	2.8	0.0	1.9
	/s/	5.3	9.2	64.4	11.7	9.4	0.0	10.8	81.7	3.9	3.6
	/ʃ/	0.3	0.6	8.9	89.7	0.6	0.0	0.0	2.8	96.9	0.3
					Total:	64.2				**Total:**	84.5
3 kHz	**/f/**	68.1	14.2	1.7	3.1	13.1	76.1	18.9	1.4	0.0	3.6
	/θ/	54.7	19.2	3.9	2.2	20.0	12.8	76.4	6.7	0.0	4.1
	/s/	36.7	28.9	10.6	5.0	18.9	4.7	58.9	24.5	5.0	6.9
	/ʃ/	1.9	1.9	13.3	75.6	7.2	0.0	0.0	4.4	93.9	1.7
					Total:	43.3				**Total:**	67.7

fusions take place between /θ/ and /f/, and between /s/ and /ʃ/, although for the lowest LPF cutoff frequencies, /s/ is also confused with both /θ/ and /f/ in the C condition, and only with /θ/ in the CV condition due to the presence of the vowel. Clear vocalic effects appear only for /f/ which is best identified, for every LPF cutoff frequency, in the context of /o/ and, especially, /u/. Vocalic effects were less marked in the CV condition than in the C condition. No significant differences between speakers/sexes showed up. This results extend and corroborate the pilot study by [2].

4.3. Multidimensional scaling

The asymmetric individual differences analysis was carried out using the alternating least squares scaling (ALSCAL) program [5, 6] which includes, among others, the ASYNDSCAL model. The confusion matrices for each LPF cutoff frequency were entered into the analysis, and thus, the patterns of confusions that differ among LPF cutoff frequencies will be assigned to separate dimensions. The MDS analysis program requires some other information about the characteristics of the data and the relationships between the different matrices which primarily depends on the possibility of being able to interpret the resulting perceptual space. Maybe the most important of those variables is the number of dimensions considered to be adequate for the perceptual space. A two-dimensional space gave interpretable results in our case, explaining 99.5% of the variance in the sample both in the C condition and in the CV condition. The MDS-derived perceptual space for both conditions has been plotted in figure 1. The location in that space of each fricative for each LPF cutoff frequency is represented by the name of the fricative (/θ/, /f/, /s/ or /ʃ/) plus the corresponding LPF cutoff frequency. Dimension 1 corresponds primarily to the distinction between sibilant and non-sibilant fricatives. Dimension 1 also contributes to the distinction between /f/ and /θ/. This is true at least for the conditions in which the bandwidth is wider. Dimension 2 distinguishes between /ʃ/ and the other fricatives. The distance between phonemes in the perceptual space represents the degree of confusion between those phonemes.

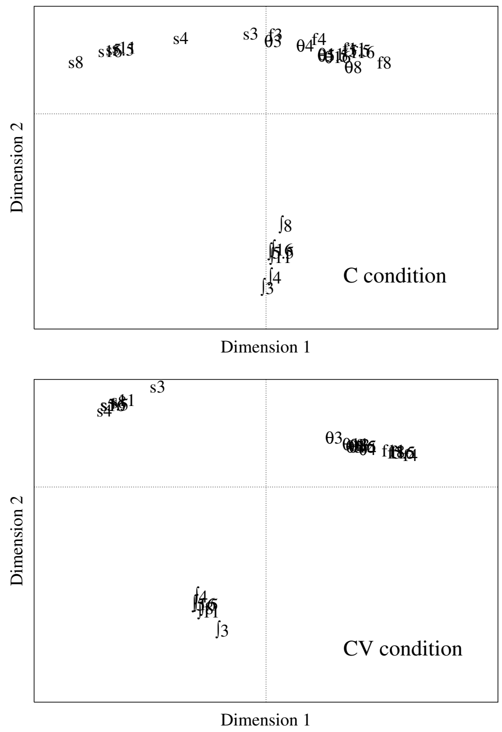

Figure 1: *Perceptual space in the C (top) and CV (bottom) conditions.*

For the C condition the perceptual space varies a lot with the reduction in spectral bandwidth. For some of the highest LPF cutoff frequencies (16, 11 and 5.5 kHz) the location of the fricatives in the perceptual space is exactly the same. For 8 and 4 kHz that location does not vary very much, but the phonemes start to move erratically towards their position for 3 kHz. There is a point in the process of reduction of the bandwidth where the phonemes lose their characteristics becoming very ambiguous or even sounding just like noise. That boundary point for these particular fricatives could be below a LPF cutoff frequency of 5.5 kHz, although it is different for each fricative, /ʃ/ being well characterized for every LPF cutoff frequency. The obtained perceptual space coincides quite well with that obtained by [7] for English fricatives sampled at 20 kHz. She used a different type of perceptual experiment (paired comparisons) and a different approach to MDS analysis. She also included one more category (/h/), and thus, interpreted dimension 2 as place of constriction. For the CV condition the perceptual space is similar for every LPF cutoff frequency. The groups of fricatives are well separated and the interpretation of the perceptual dimensions is clearer than in the C condition. Fricatives for different LPF cutoff frequencies are tightly clustered. Only for 3 kHz the location of fricatives in the perceptual space varies, with /s/ and /θ/ approaching each other.

Overall, the perceptual spaces show that in the CV condition fricatives are perceptually better defined than in the C condition.

5. Conclusion

The perception of fricatives is influenced by the spectral bandwidth. Nevertheless, satisfactory identification rates can be achieved with a low-pass filtered cutoff frequency as low as 6 kHz for fricatives in isolation. When the vowel is added to the fricative noise of CV syllables, LPF cutoff frequency can be lowered to 4 kHz without a significant reduction in identification rates. Moreover, in this condition, identification rates are 20% higher than in the isolated fricative condition.

Listeners used primarily the following distinctions: sibilant versus non-sibilant, and /ʃ/ versus /f, θ and s/. Confusions tend to occur between /f-θ/ and /s-ʃ/, and they increase as the bandwidth is decreased. Fricatives /f/ and /ʃ/ are the least influenced by the bandwidth reduction. For these fricatives the spectral energy for frequencies up to 3 kHz contains most of the information neccessary for their identification. In the CV condition fricative /θ/ is also well identified for every LPF cutoff frequency; only fricative /s/, and for a LPF cutoff frequency of 3 kHz, is not identified satisfactorily.

The perceptual space in the C condition varied a lot as the LPF cutoff frequency was lowered. Nevertheless, for the CV condition the perceptual space is very similar for every LPF cutoff frequency. Phonetic categories form better defined groups in the perceptual space of the CV condition than in the C condition.

6. Future work

The perceptual space can be used as reference for acoustic analyses aimed at determining which spectral regions are more important for the identification of fricatives (see for instance [8]). It would also be interesting to study whether a ASR system can obtain the same benefit from adding the vowel to the fricative noise than listeners, especially for the lowest LPF cutoff frequencies.

7. References

[1] Harris, K. S. "Cues for the discrimination of American English fricatives in spoken syllables", Lang. Speech, 1:1–7, 1958.

[2] Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M. and Lewis, D. E. "Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults", J. Acoust. Soc. Amer., 110:2183–2190, 2001.

[3] Feijóo, S., Fernández, S., Barros, N. and Balsa, R. "Context effects and acoustic cues for the auditory identification of Spanish fricatives /f/ and /θ/", Acta Acustica united with Acustica, 88:113–126, 2002.

[4] van der Kamp, L. J. TH. and Pols, L. C. W. "Perceptual analysis from confusions between vowels", Acta Psychologica, 25:64–77, 1971.

[5] Shiffman, S. S., Reynolds, M. L. and Young, F. W., Introduction to multidimensional scaling: Theory, methods, and applications, Academic Press, Inc., London, 1981.

[6] Young, F. W., Takane, Y. and Lewyckyj, R. J., ALSCAL multidimensional scaling program, http://forrest.psych.unc.edu/, 1977.

[7] Tokuma, W. "A multidimensional scaling study of fricatives; a comparison of perceptual and physical dimensions", Proceedings of EuroSpeech (Aalborg), 2001.

[8] Feijóo, S. and Fernández, S. "Location of spectral cues for the perceptual identification of fricatives", Proceedings of Forum Acusticum (Sevilla), 2002.