TEMPORAL EFFECTS OF PHONETIC INTEGRATION IN CV SYLLABLES

PACS REFERENCE: 43.71.Es

Feijóo, Sergio; Fernández, Santiago
University of Santiago de Compostela
Departamento de Física Aplicada, Facultad de Física
15782 Santiago de Compostela
SPAIN
Phone:+34 981563100 ext.14044
Fax: +34 981520676
E-mail: fasergio@usc.es

ABSTRACT
The possible relationship between the temporal characteristics of fricative-vowel syllables and the perception of place and manner of articulation is studied. A series of perceptual experiments using a gating technique around the release point were carried out. The results show that listeners have difficulties to perceive the place of articulation of the fricative when only the vowel is available. A portion of fricative noise is necessary to correctly perceive the consonantal place of articulation: a brief portion of 20-30 ms triggers a stop percept, while longer durations are necessary for the perception of fricative manner. In sibilant fricatives an intermediate step produces affricate percepts. An interaction between manner and place associated with the duration of fricative noise included in the stimuli was found in some specific cases. This last result can be interpreted as a purely auditory effect and not a consequence of phonetic processing.

INTRODUCTION

In this paper we study the phonetic integration that takes place in Consonant-Vowel syllables, and more specifically, the interaction between the Fricative noise (F) and the accompanying Vowel (V). Auditory identification of the consonant improves when listeners are presented with the FV syllable in relation to the presentation of only the fricative noise. The perceptual interaction between the fricative noise and the vowel is particularly important for the /f-th/ distinction [1]. In this last case, the fricative noise alone does not provide all the necessary information for the auditory identification of the fricative.

A possible explanation for that interaction is based on the articulatory effects that occur in FV syllables. According to that hypothesis, phonetic integration could be the result of:

- The coarticulatory influence of the vowel on the fricative (V-to-F anticipatory coarticulation): the vowel exerts an anticipatory effect on the fricative in such a way that the acoustical characteristics of the fricative are conditioned by the quality of the following vowel. A typical example is the effect produced by lip rounding on the spectral characteristics of fricatives /s/ and /sh/ in the context of vowels /o/, /u/ .
- The coarticulatory influence of the fricative on the vowel (F-to-V carryover coarticulation): The place of articulation of the fricative conditions the following vocalic transition. For instance, /s/ has an alveolar place of articulation, while /f/ has a labial or labiodental place. The articulatory movements that take place within the vowel are

different for both fricative contexts, and, as a result, it is expected that their acoustic characteristics vary accordingly.

In a previous paper, both hypothesis were perceptually tested using conflicting-cue stimuli [2]. Results indicated that F-to-V carryover coarticulation could play a certain role in the FV integration, although some vocalic context effects also showed up. That led us to develop a probabilistic model in which fricative and vowel are evaluated separately, the final result depending on the combination of the probabilities of membership of each segment into the class defined by the fricative's place of articulation. To evaluate the model a series of perceptual experiments were carried out in which the auditory system was used as an auditory *analyzer*. Listeners tried to evaluate the content associated with place of articulation in both the consonant and the vowel [2]. No conclusive results were obtained about the adequacy of the model, mainly because, when listeners were presented with only the vowel, there was not a clear relation between the place of articulation supposedly reflected on the vowel and listeners' percept.

Recently Smits [3], using a gating technique, has studied how the relevant information about place, manner and voicing is temporally distributed in the VCV utterances. His results suggest that place information is more spread-out for fricatives than for stops, and that the transition regions of stops are more informative on place than those of fricatives. Nevertheless, the use of 17 consonants somewhat obscured some particular effects.

In this paper we widen our previous study including different regions around vowel release, in order to assess in which way the effect of phonetic integration on manner and place of articulation is related to the temporal location and duration of those regions.

PERCEPTUAL EXPERIMENTS

Stimuli

The signals used in this study corresponded to FV syllables made up of the fricatives /th, f, s, sh/ with the vowels /a, e, i, o, u/, uttered by one male and one female speakers. Signals were sampled at 32 kHz using a Sound-Blaster audio card (16 bits). The fricative noise plus 100 ms of the vowel were used to construct the stimuli.

Signals were edited to locate the boundary between the fricative noise and the vowel, which will be henceforth called the *release point*. Taking that boundary as a reference, the stimuli were constructed using a gating technique (see figure 1). In forward-gated stimuli the part of the signal following an instant relative to the release point is deleted. In backward stimuli, the part of the signal preceding an instant relative to the release point is deleted. Forward-gated stimuli were denoted by positive offset values from 10 to 40 ms, in 10 ms steps. Backward-gated stimuli were denoted by negative offset values from -10 to -100 ms, in 10 ms steps. Offset 0 corresponded to the presentation of 100 ms of the vowel starting from the release point. Figure 1 shows the segments of signal corresponding to stimuli characterized by different offset values.

To avoid the presence of *clicks* due to a sudden onset or offset, the endpoints of the stimuli were smoothed using a 10 ms semicosine-type window. The total number of stimuli was 600= 2 talkers x 4 fricative x 5 vowels x 15 conditions.

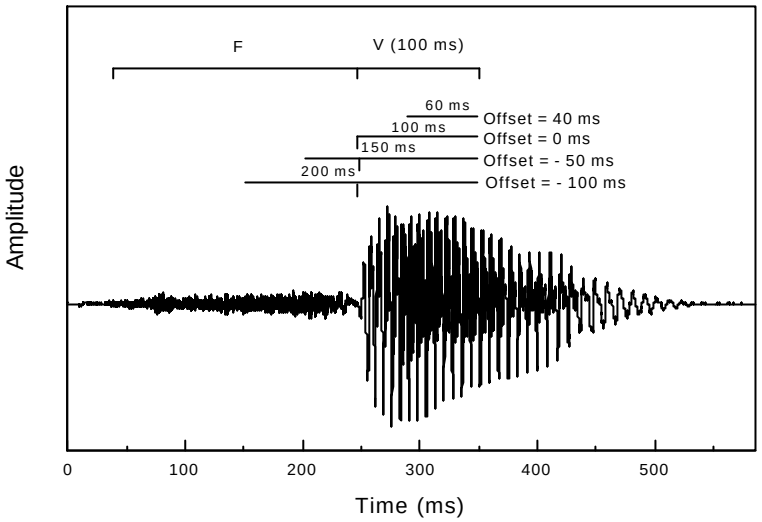

Figure 1. Speech waveform showing the temporal boundaries for different offset conditions.

Listeners

Fifty phonetically naive students with normal hearing and ages between 20 and 30 years old, participated in the experiments for course credits in the subject of Acoustics. All of them were bilingual, i.e., they understood and spoke both Galician and Spanish.

Presentation of the stimuli and responses

Listeners were explained that their task was to indicate the initial consonant in the stimuli. The response labels were: P, B, T, D, CH (for /tsh/), Z (for /th/), F, S, X (for /sh/), vowel, and diphthong. The last two responses were indicated for the case in which listeners do not perceive any initial consonant. Stimuli were presented via headphones in a quite room at a comfortable intensity level (around 70 dBA). Listeners were allowed one repetition of each stimuli, after which it was mandatory to select an answer. The experiments were self-paced and stimuli were presented blocked by speaker. The order of presentation was random and different for each listener.

RESULTS

Due to the size of the collected data, only the more significant results are presented in figure 2. To obtain a clearer picture, the responses corresponding to the stops with same place of articulation and different voicing were collapsed into a single category: labial (/p/, /b/), dental (/t/, /d/).

Figure 2 (top) show the results for /f/, separately for the male and the female speakers. Perception of a labial stop is predominant between 10 and -30 ms. Perception of a labial fricative prevails from -40 or -60 ms, depending upon the particular speaker. The perception of the vowel occurs quickly within the vocalic part, around 10 ms for the male and 0 ms for the female, indicating that the influence of the place of articulation of the consonant vanishes quite rapidly within the vowel. Overall, the results of the /f/ gated stimuli show that the perception of place of articulation when only the vowel is available is ambiguous. It was expected that in the 0 ms condition a clear labial percept was produced. Nevertheless, it was necessary to include some amount of fricative noise, of a duration similar to that of a natural plosive, to obtain a labial percept.

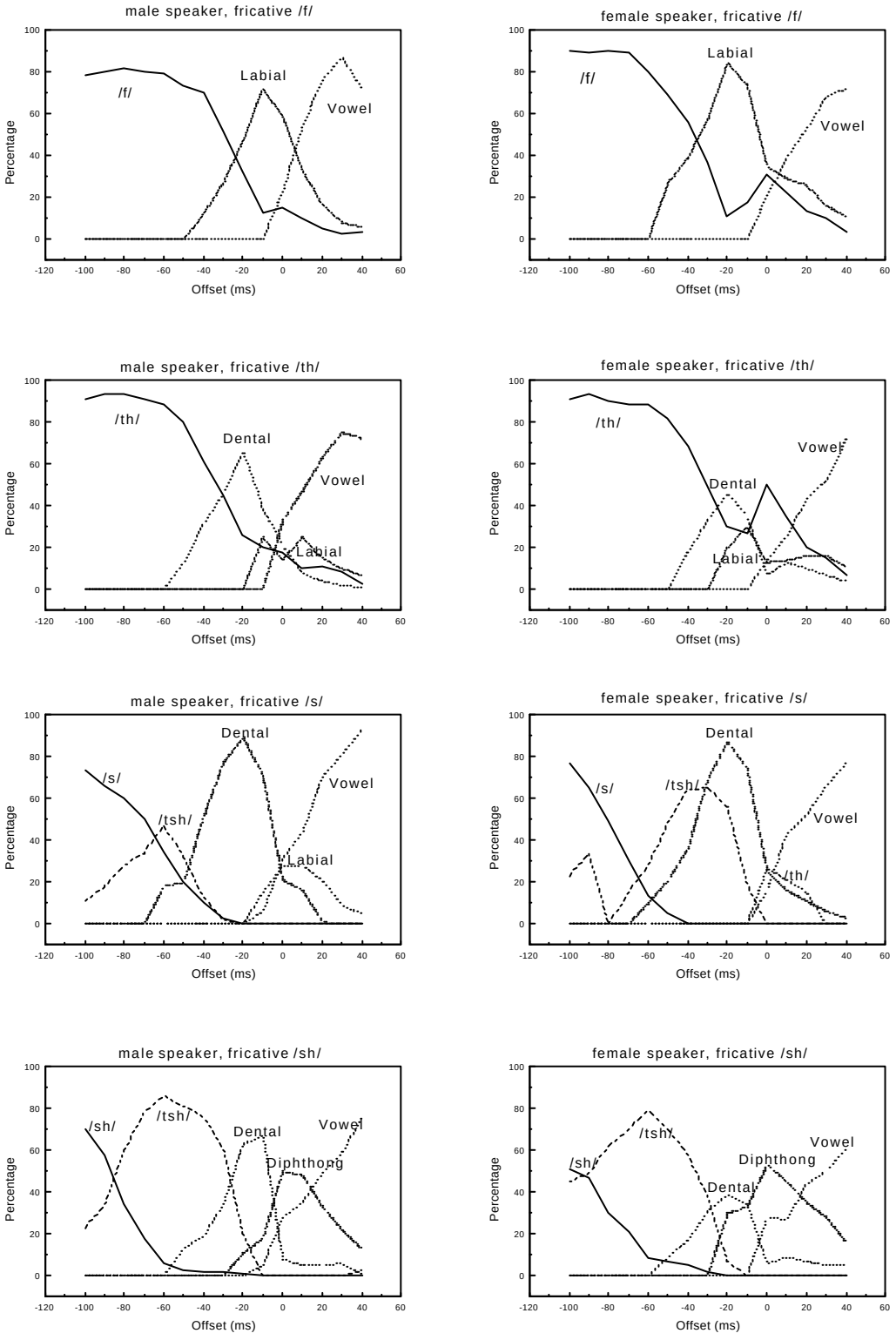

Figure 2 Perceptual identification for different offset conditions of (from top to bottom) /f/, /th/, /s/ and /sh/ pronounced by a male (left column) and female (right column) speakers.

Figure 2 (second row) show the results for /th/. The first somewhat unexpected result is that the forward-gated stimuli produced more labial than dental percepts. Nevertheless, the prevailing response for those stimuli was vowel. To have a clear dental percept it was necessary to have a duration of fricative noise of approximately 20-30 ms. This fact has been sometimes interpreted as the result of a bias in favor of a labial percept whenever the noise is absent, and only the vowel is available. As happened in other studies [4], there seems to be an additional source of information about fricative identity within the vowel, which is especially important in the case of the female speaker: a detailed inspection of the stimuli revealed that some fricative noise was superimposed on the vowel, within 10-20 ms from the release point. The perception of the fricative is stabilized from -60 ms.

Figure 2 (bottom row) show the results for /sh/. When only the vowel is available (offset 0 ms), the prevailing percept is that of a diphthong starting with vowel /i/. The presence of /i/ is limited to the first 20 ms of the vowel. Nevertheless, and contrary to what happens to /th/, the labial percept is very weak (data not included in the figure). A dental percept comparable to that of fricative /th/ is obtained when some noise of a duration between 10-30 ms is included in the stimuli. As more noise is included in the stimuli, the prevailing percept is that of an affricate, 80-90 ms of noise being necessary to induce a fricative percept. The perception of the affricate is due to the shortening of the original fricative noise, which is spectrally very similar for naturally produced /sh/ and /tsh/. It may be that the duration of natural /tsh/ noises is within the limits of offset conditions -40 and -80 ms. An alternative hypothesis is that the sudden rise in amplitude for those conditions (that roughly correspond to the part of the noise with higher amplitude) makes the rise time of the stimuli comparable to that of a natural affricate. It is difficult to say which of the two hypothesis is correct, since no satisfactory explanation of the differences between /sh/ and /tsh/ is yet available.

Figure 2 (third row) show the results for /s/. When the whole vowel is available (0 ms condition), listeners perceive mainly a vowel. As was the case with /th/, for the stimuli of the female speaker, some fricative noise is included in the first 30 ms of the vowel, giving rise to a /th/ percept. The perception of a dental stop is predominant between conditions 0 and -40 ms, and is much clearer than for the stimuli constructed from fricative /th/, although the place of articulation of /th/ is dental in Spanish and Galician, as is the place of /t/ or /d/, while /s/ has an alveolar place. Nevertheless, the formation of a medial groove in the tongue for /th/ may have an important impact on the acoustic characteristics of the transition, questioning the validity of place of articulation as an index for categorizing the acoustic signal. A surprising finding is the perception of the affricate /tsh/ for conditions between -40 and -60 ms. Although the fricative noises of /s/ and /sh/ have some common characteristics, previous experiments have shown that they were not confused by the listeners. A certain interaction between the spectral characteristics of /s/ and the duration of the noise may have occurred, triggering the affricate percept for lesser noise durations than in the case of /tsh/.

DISCUSSION

The first evidence that comes out of the results is that the perception of place of articulation is ambiguous when listeners only have the information contained in the vowel. The vocalic transition seems to affect the first 10-20 ms of the vowel, and its perceptually little informative on its own. When some fricative noise, of a duration between 10 and 40 ms, is added to the stimuli, listeners perceive a stop. In fact, the durations of fricative noise necessary for stop percetion are similar to the reported duration of stop noises in Spanish [5]. The stop percept was strong for /f/ and /s/; for /th/ and /sh/ it was ambiguous.

With between 40-90 ms of noise the stimuli constructed from /sh/ are perceived as affricates, which is also in accordance with the durations of affricate /tsh/ included in a database of fricatives and affricates of our laboratory. For /s/, the affricate percept prevails for noise

durations between 40-60 ms, indicating that in this case the affricate manner induces a change in perceived place. The minimal noise durations that correspond to the perception of fricative manner are 30-40 ms for /f/, 30 ms for /th/, and 60-90 ms for /s/ and /sh/.

Another interesting outcome of our experiments is that a change in perceived manner was associated with a change in perceived place in some cases. Since the labels were constant throughout all conditions, that change in perceived place cannot be attributed to a phonetic processing effect, as suggested by Carden et al. [4]. It is more likely that some temporal detection mechanism is associated with the perception of manner, and that the resulting interaction between the noise and the vowel influences the final percept, in which case the manner-place interaction would be a purely auditory effect. In this sense, it seems that in order to obtain a certain manner percept certain temporal characteristics that define the morphology of the waveform are of great importance.

Lastly, our results argue against the separate processing of cues in speech perception. If that were the case, the role of the vowel in the integration process would be minimal, according to our experiments, since little trustworthy information about place is present. The change in perceived place across conditions for some fricatives further support the integration of both noise and transitions into a unitary representation, as suggested by Kurowski & Blumstein [6], among others.

A final question remains. What is then the role played by the vocalic transition? Several studies using conflicting-cue stimuli have shown dramatic changes in fricative perception when the vocalic transitions of /th/ and /f/ are traded against each other. Nevertheless, listeners were unable to detect a clear percept distinguishing both transitions. A proper perception of place required some noise, which improved the place results particulary for /f/, despite the fact that /f/ and /th/ noises are quite ambiguous and easily confused by listeners. Further resarch is still needed to clarify that question.

ACKNOWLEDGMENTS
This work was financed by Xunta de Galicia under project PGIDT00PXI20608PR.

REFERENCES

1. Harris, K. S., “Cues for the discrimination of American English fricatives in spoken syllables,” Lang. Speech, 1:1-7, 1958.
2. Fernández, S., Feijóo, S., Balsa, R. and Barros, N., “Perceptual effects of coarticulation in fricatives,” Proceedings of the ICASSP2000, 2000.
3. Smits, R., “Temporal distribution of information for human consonant recognition in VCV utterances,” J. of Phonetics, 27:111-135, 2000.
4. Carden, G., Levitt, A., Jusczyk, P. W. and Walley, A., “Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place,” Percep. & Psychoph., 29(1):26-36, 1981.
5. Feijóo, S., Fernández, S. and Balsa, R., “Acoustic and perceptual study of phonetic integration in Spanish voiceless stops,” Speech Communication, 27:1-18, 1999.
6. Kurowski, K. and Blumstein, S. E., “Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants,” J. Acoust. Soc. Amer., 76(2):383-390, 1984.