Perception of consonant place of articulation: phonological categories meet natural boundaries

*Willy Serniclaes, *Caroline Bogliotti,**René Carré

*LEAPLE, UMR 8606, CNRS & U. Paris 5, Villejuif, France
wsernic@vjf.cnrs.fr
**Département Traitement du Signal et de l'Image, Unité associée au CNRS, ENST, Paris,
carre@tsi.enst.fr

ABSTRACT

Previous studies reveal that perception of consonant place of articulation is organized around a central reference given by the neutral vowel. In the neutral vocalic context, place boundaries correspond to flat F2-F3 transitions in stimuli where the latter covary. However, when modified separately, the F2 and F3 flat formant transitions subdivide the F2-F3 onset frequency space into four regions. As only three place categories are phonologically relevant for stop consonants in French, the question is to know how the four regions are shared out between categories. The aim of the present study is to specify how natural boundaries are traded-off for operating this share-out. Both labeling and discrimination responses were collected. Results show that even when natural boundaries are located inside phonological categories they can still affect consonant discrimination. Further, non-phonological categories remain perceptible for some adult subjects. The implications of these findings for phonological development are discussed.

1. INTRODUCTION

Previous studies reveal that perception of consonant place of articulation is organized around a central reference given by the neutral vowel (schwa). In the neutral vocalic context, place boundaries tend to correspond to flat F2-F3 transitions, the categories being characterized by rising vs. falling transitions [1]. This suggests that place perception is grounded on a "natural" boundary in the neutral context [2], but which has to undergo specific adjustments in other contexts. The place boundary is shifted towards falling transitions before back rounded vowels, rising transitions before front unrounded vowels, and intermediate positions before front rounded vowels. The radial model of place perception states that the contextual adjustments of the transition boundary follow a rotational movement in the F(onset) – F(endpoint) plane around a central point corresponding to the flat transition in the neutral context, the direction of the boundary line depending on the perceived identity of the following vowel [2]. The aim of the present study is to further specify the relationship between boundaries and categories in the neutral context. The F2-F3 flat formant transitions subdivide the F2-F3 onset frequency space into four regions. As only three place categories are phonologically relevant for stop consonants in French (i.e. /b, d, g/), the question is to know how the four possible regions are shared out between these categories. Here too specific adjustments are necessary in order to accommodate natural settings with linguistic units.

The optimal solution to this problem consists in the emergence of perceptual tradeoffs between F2 and F3 transitions. However, as this is a fairly complex developmental process, natural boundaries might remain present at various degrees in the adult population. Natural boundaries might remain entirely present without any influence of the language categories in some individuals. The perception of the three place categories would then occur in the original four-category space, leaving one of the four regions unexploited. Place perception would be "non-phonological" for these subjects. Non-phonological perception does not imply a lack of phonological processing in speech recognition. Rather, it is specifically the pre-lexical phonology which would be weak or even totally absent. A completely different pattern would consist in the absence of any trace of natural boundaries, the latter being replaced by language-specific ones. The original four-category space would then be equally divided into three regions with a share-out corresponding to the language productive categories. Place perception would be entirely phonological for subjects presenting this perceptual profile. Perception might be entirely phonological for some individuals and entirely non-phonological for others. Another part of the population might lie between these two extreme profiles.

These assumptions about the variable degrees of perceptual phonologisation across individual were investigated by collecting both labeling and AX

ISBN 1-876346-48-5

discrimination responses of adult French listeners to different synthetic /C+schwa/ syllables continua in the F2-F3 space. Place perception was taken as being entirely phonological if: (1) labeling categories were homogeneous, i.e. that they remained confined in definite portions of the acoustic space; (2) there were no more than three discrimination peaks on each continuum; (3) these peaks corresponded to phonological rather than phonetic boundaries. Phonological boundaries will be assessed by measuring 50 % labeling thresholds. Phonetic boundaries are assumed to correspond to the natural landmarks provided by F2 or F3 flat transitions. In doing so we do not imply that phonetic perception is performed by the generalized auditory system. The use of natural boundaries for phonetic purposes might indeed be the end product of and evolutionary process which has left simple traces in the neutral vocalic context. Place perception is however much more complex when the entire vocalic space is taken into account. Speech specific rules, such as those precluded in the radial model, are then required [2]. These rules might result from the adaptation of general purpose mechanisms, such as those reflecting linear constraints [3], to the perception of speech signals [4].

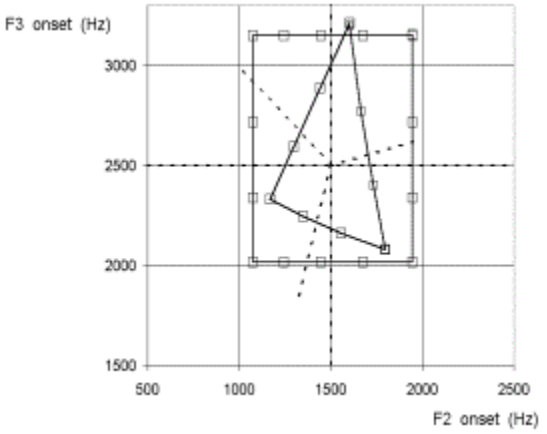

Figure 1. Stimulus continua. The rectangle corresponds to the "phonetic" continuum. The 4 component lines follow directions normal to the natural boundaries at F2=1500 Hz and F3=2500 Hz. The triangle corresponds to the "phonological" continuum. The 3 component lines follow directions normal to the expected phonological boundaries, which divide the acoustic space in 3 equal-sized regions.

The stimulus continua were constructed so as to obtain a fair test of discrimination peaks, either phonological or phonetic. For a given boundary line, perceptual discrimination depends on the direction of the continuum line relative to boundary line in the space of the acoustic cues. Discrimination will be maximal for continua normal to this line because the rate of increase towards and from the boundary is maximal for such continua. It is therefore necessary to use continua which are as close as possible to the normal direction relative to the boundary lines in order to obtain a fair assessment of discrimination peaks. This requires in first instance to have some a priori knowledge on the direction of the boundary lines. The simplest assumptions on this point are as follows. A four-category system would use natural boundaries corresponding to flat F2 or F3 transitions, with onsets at 1500 and 2500 Hz respectively for stops in /C+ schwa/ syllables (Figure 1). A three-category space would optimally subdivide the space into three equally sized regions, as shown in Figure 1. The "void" region corresponding to the place category which is absent in French would be shared by its direct neighbors, i.e. the labial (/b/) and apical (/d/) categories. The dorso-velar category (/g/) would in turn expand symmetrically into the natural labial and apical regions. Following these assumptions, the boundaries would follow three lines crossing at 1500 Hz F2- 2500 Hz F3, separated by 120° angles and with angular coefficients of 15° (/b-g/ boundary), 135° (/d-b/ boundary, and 255° (/b-g/ boundary) relative to the F2 axis. Two stimulus continua were used in this study (Fig.1), one with directions normal to those of the alleged boundaries in a three-category space (i.e. those with 15°, 135° and 255° angular coefficients), the other with directions normal to the alleged boundaries in a four-category space (i.e. those with 0° and 90° angular coefficients).

2. METHOD

23 stimuli CV were generated with a parallel formant synthesizer. F1-F2-F3 transitions ended at 500, 1500 and 1500 Hz respectively after a 27 ms transition. The VOT was of -95 ms, the stable vocalic portion of 154 ms. The stimuli differed as to the onset of F2 and F3 transition. 14 stimuli were generated by separate modification of the F2 and F3 onsets along a "phonetic" continuum as shown in Figure 1. 9 other stimuli were generated by joint modification of the F2 and F3 onsets along a "phonological" continuum (Figure 1). Successive stimuli were 1 Bark apart on each continuum.

10 French-speaking subjects, aged between 17 and 59 years and without known auditory problems, took part in the experiment.

Correct discrimination response scores were calculated for adjacent stimuli on each continuum, either phonological or phonetic, as the mean correct response to both same pairs (e.g. for the first pair on the phonological continuum: O1O1, O2O2) and different pairs (O1O2, O2O1).

ISBN 1-876346-48-5

3. RESULTS

The data were first analyzed separately for each subject. Inspection of the labeling curves showed that five only among the ten subjects identified at least one stimulus as /b/, /d/ and /g/ with scores 100 % scores on both continua. For four other subjects, either the /d/ or the /g/ response scores were generally at or below 50% and never reached 100 %. For the remaining subject, 100% /d/ responses were only present for stimuli along the phonetic continuum and remained at or below 50% on the phonological continuum. The identification data of these five last subjects, with only two consistent labeling regions, will not be further examined.

Among the five subjects who identified /b/, /d/ and /g/ with scores 100 % scores on both continua, two exhibited /g/ above threshold (50 %) identification responses in two different regions on both the phonetic and phonological continua, either below or above 1500 Hz F2. These two subjects will be referred to as the "four category" subjects, the three others as "three category" ones. The data of the two groups were analyzed separately and the /g/ responses covering the two frequency regions were re-labeled differently (g1 and g2) and considered as exemplars of two different phonetic categories in the statistical analyses.

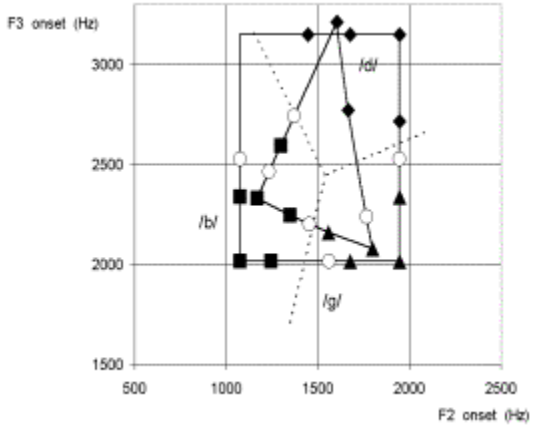

Figure 2. Labeling and discrimination results for the three category subjects. Stimuli collecting at least 75% /b/, /d/ or /g/ responses are indicated by squares, diamonds and triangles respectively. Broken lines represent category boundaries. White circles indicate above chance discrimination peaks.

The identification responses of the three and four category subjects were analyzed separately by multivariate logistic regression with F2 and F3 onset values (in Barks) as predictors. For three category subjects both F2 and F3 effects were significant ($p<.002$) for each category distinction, i.e. /d-g/, /b-d/ and /b-g/. Figure 2 gives the boundary lines between categories as well as the stimuli with at least 75% responses in one of the three categories for these subjects. The F2-F3 space is shared into three regions of roughly the same size, the angular coefficients of the boundary lines being of 16°, 131°, 250°. These values are not very far from those of the 15°, 135° and 255° values corresponding to an optimal share out of the acoustic space into three equally sized regions.

For the four category subjects both F2 and F3 effects were significant ($p<.002$) for each category distinction, at the exception of F3 ($p=.13$) for /b-g1/, and F2 ($p=.12$) for /d-g1/. Figure 3 shows the boundary lines and stimuli with 75% category responding for these subjects. Category regions are not equally sized. The /b/ and /d/ regions are larger than the /g1/ and /g2/ ones. The /b-d/, /d-g2/, and /b-g1/ boundaries follow a F2 = 1500 Hz line. The two other boundaries depend both on F2 and F3 frequencies.

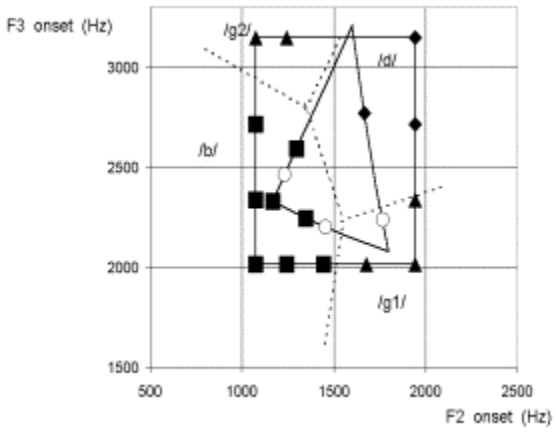

Figure 3. Labeling and discrimination results for the four category subjects (see Figure 2 for symbols).

Discrimination results were also different for the three vs. four category subjects. These results will be summarized here by considering discrimination peaks significantly ($p<.05$) above the 50 % chance level. Three category subjects exhibited 4 peaks along the phonological continuum and located as follows (Figure 2): in the vicinity of the /b-g/ boundary; inside the /b/ region close to the crosspoint with the 2500 Hz F3 line; at the /b-g/ boundary; inside the /g/ region. There were 3 further discrimination peaks along the phonetic continuum for these subjects, among which two were located at the crosspoint with the 2500 Hz F3 line and one close to the crosspoint with the 1500 Hz F2 line. Four category subjects only exhibited 3 discrimination peaks and all were along the phonological continuum (Figure 3) and which corresponded to 3 out of the 4 peaks already found for three category subjects along this continuum. One of them was the inside the /b/ region close to the crosspoint with the 2500 Hz F3 line and the two others were located in the neighborhood of the either the /b-g1/ or /d-g1/ boundaries.

ISBN 1-876346-48-5

4. DISCUSSION AND CONCLUSIONS

The stimuli used here were synthesized without burst which may explain why half of the subjects did not hear good exemplars of each of the three places of articulation categories. Perceptual weighting of acoustic cues vary across subjects and both the transitions and the burst convey important cues for perceiving stop place of articulation. It is therefore probable that the subjects who did not consistently perceive exemplars of the three place categories with the burstless stimuli used in this experiment rely more on burst than on transitions for place recognition.

The other half of subjects were able to make coherent place distinctions on the basis of transitions alone but they differed in other important aspects. The formant transition space was shared out into three categories for 3 among the 5 subjects who could rely on transitions for labeling place of articulation. These subjects complied with the first phonologisation criterion posited here, concerning the homogeneity of labeling categories, i.e. the fact that they remain confined in definite portions of the acoustic space. Further, as expected, the acoustic space was shared out into three fairly equally sized regions thereby revealing an optimal exploitation of the phonetic potential offered by formant transitions. However, contrary to what would be expected for a phonological profile, the number discrimination peaks along the phonological continuum was larger than the number of labeling boundaries (4 vs. 3). The extra peak was located in the vicinity of a natural boundary (at 2500 Hz F3), which supports a phonetic interpretation. Further, the 3 peaks located on the phonetic continuum were close to natural boundaries (either 2500 Hz F3 or at 1500 Hz F2) and one among them (at 2500 Hz F3) was clearly apart from labeling boundaries. This suggests that even subjects with three labeling categories rely on natural boundaries in their discrimination behavior. Four-category subjects had some boundaries in much the same direction as the 1500 Hz F2 line. As expected, these subjects used a natural boundary for performing labial-apical place distinctions. However, other boundaries were not aligned on the 2500 Hz F3 line. Further, there were no discrimination peaks along the phonetic continuum for these subjects. Finally, the peaks found long the phonological continuum were fairly similar to those found for the three category subjects. The discrimination behavior of the four category subjects was then certainly not less phonological than for three category ones.

In summary, the discrimination data collected in this study show that subjects keep some trace of the sensitivity to natural boundaries in the processing of transitional cues for the perception of stop place of articulation. Discrimination peaks were found outside the phonemic boundaries in regions corresponding to flat formant transitions. This indicates that adult subjects still make use of natural boundaries for perceiving speech, thereby revealing a pre-phonological stage of processing.

The labeling data showed that the more than three place categories are sometimes present in the listener's repertoire, even in a language with only three such categories. When four categories are present, some of the place boundaries tended to align along natural boundaries. While the fourth category remains to be specified in relationship with speech production models [5], the present results suggest that there are different pathways for accommodating phonological categories with phonetic constraints in the course of perceptual development. Perception of within-category variants (such as those found here for stimuli with low F2-high F3 at onset, see Figure 3) does probably not raise problems for speech perception. But it might constitute an obstacle for the acquisition of written language [6]. Although none of the subjects in the present experiment reported to have experienced problems in learning to read further investigations are needed to clarify the relationship between phonologisation of speech categories and reading deficits.

5. ACKNOWLEDGMENTS

This research was supported by the French Project ACI "Cognitique".

6. REFERENCES

[1] R. Carré, J.-S. Liénard, E. Marsico and W. Serniclaes (2002) "On the role of the "schwa" in the perception of plosive consonants". 7th *ICSLP,* 1681-1684.

[2] W. Serniclaes & R. Carré (2002). Contextual effects in the perception of place of articulation: a rotational hypothesis. *7th ICSLP*, 1673-1676.

[3] H. M. Sussman, D. Fruchter, J. Hilbert, and J. Sirosh (1998) "Linear correlates in the speech signal: The orderly output constraints" *Behav. and Brain Sciences* 21, 241-259.

[4] A. M. Liberman (1998) "Three questions for a theory of speech," Proc. ICA-ASA '98 meeting.

[5] M. Mrayati, R. Carré, and B. Guérin, "Distinctive region and modes: A new theory of speech production," *Speech Com.*, vol. 7, pp. 257-286, 1988.

[6] W. Serniclaes, L. Sprenger-Charolles, R. Carré, & J.F. .Démonet (2001) "Perceptual discrimination of speech sounds in dyslexics" *J. Speech Lang. Hrng. Res.* 44, 384- 399.