PERCEPTION OF PLACE-OF-ARTICULATION DISTINCTIONS: COMMON REPRESENTATION FOR VOWELS AND CONSONANTS

Willy Serniclaes[a,b] *& Jacqueline Salinas*[b]

[a]CNRS, France;
[b]Laboratoire Psychologie de la Perception, Université Paris Descartes, France
willy.serniclaes@parisdescartes.fr; salinas.jacqueline@gmail.com

ABSTRACT

We look for a common perceptual representation of place-of-articulation distinctions between stop consonants and those between closed vowels. This representation can be either acoustic or articulatory in nature but not both acoustic-articulatory because the acoustic consequences of front to back articulatory changes are *inverted* for stops and vowels. Identification data show that the perceptual boundaries for vowels and consonants do not match in an acoustic representation. However, vowel and consonant boundaries almost coincide after *rotation*, suggesting an articulatory-wise isotropic representation of place distinctions.

Keywords: vowels, consonants, invariance

1. INTRODUCTION

The distinction between vowels and consonants occupies a central position in feature systems. This distinction is present in all languages and delineates two fairly large classes of speech sounds. The articulatory properties of vowels differ in several aspects from those of consonants, mainly in the degree of vocal tract opening, although none of these properties alone is sufficient to distinguish these classes of sounds [2, 3, 8].

While the main focus of previous research is about *differences* between the two classes of sounds, there has also been a debate about possible *similarities* in their internal structure. The question raised was whether the articulatory differences between vowels and those between consonants can be represented with the same features. Arguments have been advanced either in support of a common representation [5] or against [4] but the question remains. Also notice that there are both common points and differences in the definition of consonant and vowel features in the "*Preliminaries to speech analysis*" [6]. Two binary features are used for capturing the place distinctions between consonants and closed vowels, one being common to both classes (grave/acute for b/d and u/y) and the other specific to each class (compact/diffuse for g/b; plain/flat for i/y). In this paper the common representation of vowels and consonants is addressed with a new paradigm based on the perceptual boundaries between places of articulation for the two classes of speech sounds. We will conclude that there is a common representation if the place boundaries between vowels correspond to those between consonants.

The problem raised by the notion of a common representation is that the acoustic relationships between stops and vowels do not match their articulatory relationships. Acoustically, the front stop /b/ corresponds to the back vowel /u/, both being characterized by low-frequency energy. Conversely, the back stop /d/ corresponds to the front vowel /i/, both being characterized by high-frequency energy. The acoustic relationships between consonants and vowels are thus *inverted* with respect to their articulatory relationships. This may be called "Jakobson's paradox" ([5] p.148). So we are faced with a twofold question. First, is there common perceptual representation of consonants and vowels? Second, is this common representation acoustic or articulatory in nature? The aim of the present study is to investigate these two points.

The acoustic differences between consonant places of articulation mainly reside in the spectrum of static noise segments and in the direction of F2 and F3 formant transitions. The present study is limited to differences in formant frequencies in order to simplify the comparison with vowels. For the same reason, we compare voiced stop consonants with closed vowels because these two subclasses of consonants and vowels are fairly close on both acoustic and articulatory grounds. Voiced stops have a relatively weak noise component in comparison with fricative and voiceless stops, and the differences in F2 and F3 frequencies between closed vowels are larger and more similar to those between stops than are those between open vowels. Finally, these comparisons will be performed in French, which has both three voiced stops (/b, d, g/) and three closed vowels (/i, y, u/).

The question raised here is whether the stop boundaries match the vowel ones in the acoustic F2-F3 transition space. If this does not happen there remains another possibility for a common representation of consonants and vowels, although it is more complex. Supposing that there is a common *articulatory* representation of the two classes of sounds, the /b,d,g/ consonant percepts should then correspond to the vowel percepts in the /i,y,u/ order. However, this is only possible after some rotation of the acoustic representations at some higher level of perceptual processing. The crucial test for assessing the perceptual relevance of a common articulatory representation of place differences for stops and vowels is whether the perceptual boundaries between b/d/g match those between i/y/u after a rotation of the acoustic space.

In a two-dimensional stimulus space, such as the F2-F3 transition onset space, each boundary is characterized by a line with a specific direction, starting from the onset frequencies and ending at the offset (stable) frequencies. When offset frequencies are fixed (same vocalic context for CV transitions), the possible boundaries form a circle around a central point corresponding to them. In order to test the relevance of each possible boundary with the highest possible precision, i.e. such that the slope of the identification function is as large as possible around each potential boundary, a circular continuum orthogonal to the direction of each boundary should be used.

We investigated possible similarities between consonants and between vowels in the perceptual representation of place differences in the following way. Both /Cə/ and /Və/ syllables were generated by systematically changing the F2 and F3 transition onset values along a circular continuum centered on the F2 and F3 frequencies of the final /ə/ vowel (Fig. 1; for the choice of the neutral context see [1, 10, 11]). The 16-step continuum was composed of 16 linear segments, each orthogonal to the F2-F3 transition direction in the middle of these segments, with each step corresponding to a difference of one Bark in the combined F2 and F3 onset frequencies. This allowed for a fair test of all the potential boundaries, i.e. those corresponding to all possible combinations of F2 and F3 formant transition directions.

According to the “acoustic” hypothesis, the following relationships are expected between vowel and stop categories: /i/ would correspond to /d/, /y/ to /g/ and /u/ to /b/. According to the “articulatory” hypothesis, the following relationships are expected: /i/ would correspond to /b/, /y/ to /d/ and /u/ to /g/ after a rotation of their acoustic representation in the F2-F3 transition onset space.

2. METHOD

16 [stop +ə] segments and 16 [vowel + ə] segments were generated with a parallel formant synthesizer (by R.Carré http://pagesperso-orange.fr/ren.carre). The endpoint frequencies were 500, 1500 and 2500 Hz respectively for F1, F2 and F3. The onset formant frequencies were identical for the stop and vowel stimuli, and they were generated so as to obtain 16 stimuli with 1-Bark intervals between two adjacent values on a circular continuum centered on the neutral vocoid values (Fig. 1) The F1 onset was fixed at 300 Hz and the F0 was constant at 100 Hz for all the stimuli. Total duration was 170 ms. The stops had a negative VOT of 65 ms, a 10-ms noise burst, a 20-ms transition, and a stable vocalic portion of 75 ms. The fricative burst formants were placed at the same frequencies as the voiced ones. For the vowels, the initial and final vocalic segments were each 50 ms in duration and the transition was 70 ms.

22 adult French speakers without known hearing problems participated in the experiment. The stimuli were presented over headphones (Beyerdynamic DT290) and identification responses (forced choice between B, D, G or between I, U, OU) were given on a computer keyboard. Each stimulus was presented 10 times in random order in two different experimental series, one for the stops and the other for the vowels. The order of presentation of stops and vowels was balanced between participants. Boundaries were calculated separately for each participant and for

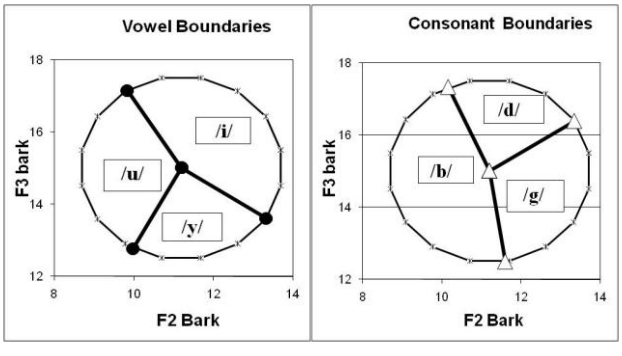
each contrast (b/d, b/g, d/g, i/u, u/y, y/u). A boundary was defined as the value along the circular stimulus continuum, in degrees between 0 and 360°, corresponding to equal percentage responses of the two most frequent response categories. Individual data were discarded whenever one category did not collect response frequencies above 50% for at least one stimulus.

3. RESULTS

Of the 22 participants, 17 displayed all the vowel and consonant categories, with at least one stimulus eliciting more than 50% responses for each category. The results of the 5 remaining participants, those who did not collect 50% for at least one category due to the use of simplified synthetic stimuli, were discarded for boundary calculations.

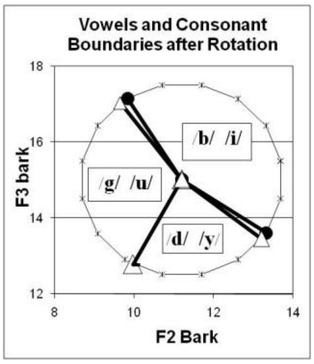
The match between the consonant and vowel category *boundaries* is quite poor. The mean consonant and vowel boundaries in the F2-F3 transition onset space are presented in Fig. 1. On acoustic grounds one would expect a match between the following boundaries: d/b and i/u (rising/falling), g/d and y/i (divergent/falling), b/g and u/y (rising/divergent). The d/b and i/u boundaries are fairly close, but the two other pairs of consonant and vowel boundaries (g/d and y/i, b/g and u/y) do not match at all. When tested with a repeated measures ANOVA with Class (consonant vs. vowel) and Acoustic Contrast (rising/falling, divergent/falling, rising/divergent) as fixed factors and participant as a random factor, the Class x Acoustic Contrast interaction was highly significant ($p<.001$). When tested separately for each contrast, the effect of Class was significant for u/y vs. b/g and for y/i vs. g/d (both $p<.001$). However, the effect of Class was not significant for i/u vs. d/b ($p=.15$).

Figure 1: Consonant and vowel boundaries.



Although there are important differences between the absolute locations of consonant and vowel boundaries, there is also a striking correspondence in their relative positions. The similarity between the vowel and consonant spaces becomes evident when a 152 rotation is applied to the consonant space (Fig.2).

Figure 2: Consonant boundaries after 152° rotation and vowel boundaries.



The vowel and consonant boundaries almost coincide after rotation (Fig.2) in accordance with the articulatory hypothesis: i/u goes with b/g, u/y goes with g/d, and y/i goes with d/b. When tested with a repeated measures ANOVA with Class and Articulatory Contrast as fixed factors and participant as a random factor, the Class x Articulatory Contrast interaction was at all not significant ($F<1$).

4. DISCUSSION AND CONCLUSIONS

In summary, these results show that (1) the perceptual boundaries between vowels and those between stop consonants do not match in the F2-F3 transition onset space, and (2) the i/y/u vowel boundaries almost coincide with the b/d/g stop boundaries, in that order, after the latter undergo a 152° rotation. These are remarkable findings for two reasons. First, the fact that the place boundaries between vowels do not match with those between stops in the F2-F3 acoustic space means that there is no common representation of the two classes of speech sounds at an acoustic-auditory processing level. Second, the fact that the boundaries between vowels along the front/back dimension (from /i/ to /u/) almost coincided with those between stops along the same articulatory dimension (from /b/ to /g/) after rotation suggest that their acoustic locations are governed by a common representation at an articulatory-wise level.

The locations of the place boundaries for vowels in the F2-F3 transition onset plane differ from those of consonants, indicating that place perception varies as function of the articulation class in an acoustic representation. Remember that on acoustic grounds the /b,d,g/ categories should correspond to /u,i,y/, in that order. Instead, what we found is that the consonant categories are the mirror image of the vowel categories, and with /b,d,g/ corresponding to /i,y,u/ in that order (Fig. 2). These reversals are quite systematic, as evidenced by the equivalence between the vowel and consonant boundaries after rotation. The rotational equivalence between these boundaries suggests that the perceptual system has integrated the inversion of the acoustic consequences of articulatory changes along the front/back dimension between the two classes of speech sounds (see Introduction: "Jakobson's paradox") in order to build up a common representation.

That knowledge about speech production contributes to speech perception is evident from the results of various studies [9]. However, the nature and origin of this knowledge is an open question. In the present results, the perceptual equivalence between the b/d/g and i/y/u contrasts represents the common articulatory topology of the two classes of sounds along the front-back dimension, but it does not fully represent the articulatory relationships between consonant and vowel categories. For instance, the labial/non-labial consonant contrast (b/g) is equivalent to the unrounded/rounded vowel contrast (i/y), which does not make much sense on articulatory grounds. This suggests that perceptual representations of articulatory gestures are not physiological in nature, as some motor theories of speech perception would predict (notably [7]). Instead, perception seems to call upon topological landmarks of sound articulation without retaining the contribution of the individual articulators.

In conclusion, the present study suggests that the perception of place distinctions is based on a representation which is common to consonants and vowels and reflects the topological relationships between consonants and vowel along the front/back dimension without incorporating articulatory details.

5. REFERENCES

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

[2] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English.* New York: Harper and Row.

[3] Fant, G. 1969. Distinctive features and phonetic dimensions. *STL-QPSR, KTH* Stockholm 10, 2-3, 1-18.

[4] Fant, G. 1973. *Speech Sounds and Features*. Cambridge, Mass: MIT Press.

[5] Jakobson, R. 1973. *Essais de Linguistique Générale*. Paris: Editions de Minuit.

[6] Jakobson, R., Fant, G., Halle, M. 1952. *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates.* Cambridge Mass: MIT Press.

[7] Liberman, A.M., Mattingly, I.G. 1985. The motor theory of speech perception revised. *Cognition* 21, 1-36.

[8] MacNeilage, P.F. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499-546.

[9] Schwartz, J.-L., Basirat, A., Ménard, L., Sato, M. In press. The perception for action control theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics.*

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

[11] Serniclaes, W., Geng, C. 2009. Cross-linguistic trends in the perception of place of articulation in stop consonants: A comparison between Hungarian and French. In Pellegrino, F., Marsico, E., Chitoran, I., Coupé, C. (eds.), *Approaches to Phonological Complexity*. The Haghe: Mouton, 241-266.