Gunnar Fant’s contribution to speech perception: forcing the speech code

Willy Serniclaes
Laboratoire Psychologie de la Perception, CNRS & Université Paris Descartes,
45 rue des Sts Pères, F-75270 Paris cedex 06. willy.serniclaes@parisdescartes.fr

A recurring theme in Gunnar Fant’s work was about uncovering the link between the speech signal and the phonological system. He was probably the first one to state that the basic problem in the study of speech communication is to explain the transmission of invariant units, the distinctive features, by a contextually variable acoustic signal. His fundamental contribution to the acoustics of speech made it clear that the quest for invariance in the signal was an enormous challenge.

In this quest he came to the conclusion that variability should also be part of the solution. In Fant’s words: “A main point in my look ahead has been the challenge to force the speech code. [...]The seeming lack of invariance which has discouraged so many investigators ceases to be a problem if we are able to structure the variability as a part of the code” (Fonetik 2000, http://www.speech.kth.se/~gunnar/halfcentury.pdf, p.7). In the present address, I will illustrate the importance of this statement for the study of speech perception with different examples. Supportive arguments will be taken from both empirical findings and key concepts in current theories of speech perception.

1 Features as differential units: a reminder

There is a definite trend in current studies on speech perception for viewing categorization processes in terms of segments, or prototypes, rather than in terms of features, or boundaries. Notably, one of the prevalent models of speech development during infancy, "Neural Magnet" model, is focused on prototypes, even if it acknowledges the contribution on natural boundaries as universal building blocks [7]. It should be stressed that the universal potential is not made of sound categories but rather of auditory boundaries between different sounds categories, i. e. of "basic cuts” in the acoustic space [4]. While there is no question to deny the importance of segmental/ prototypical representations, we are far from having drawn all the implications of features for the study of speech perception.

Features are units of difference, not categories but oppositions between categories ([5], p.130). Features can be defined on psychoacoustic grounds as boundaries between sound categories. Both boundaries and features are differential units, a striking similitude in view of their different epistemological status, boundaries being defined on psychological grounds whereas features are defined on linguistic grounds. This similitude is at the root of the vectorial feature definition by Fant (1967):

“In an integrated view based on all parameters of importance for a distinction the distinctive feature or rather its speech wave correlate can be conceived as a vector perpendicular to the hypersurface constituting the multidimensional boundary” ([2,3], p.166).

This vectorial definition has a great methodological value for the study of speech perception, as previously emphasized by Max Wajskop (Personal Communication). This will be illustrated here with three examples showing how a vectorial conception allowed to progress on three fundamental topics in the study of speech perception: invariance, motor vs. acoustic representations and common vs. separate feature systems for vowels and consonants.

2 Variability is part of the code: new insights on featural invariance

Consider the following example from the current work by colleagues and myself on the perception of place of articulation contrasts. With fricative+vowel stimuli varying in the F2-F3 transition onset space we found that the direction of the perceptual boundary for the f/s contrast changed as a function of the vocalic context ([12]; Figure 1).
This was expected from the classical results at the Haskins Laboratories on the effect of the vocalic context on the perception of place contrasts in stop consonants showing that the F2 transition of a prototypical /d/ sound being rising for /du/ but falling in /di/. Our results show that similar changes occur for fricatives: the boundary moves from a rising transition for /fu/su to a falling transition for /fi/si. However, the interest of our results resides in the convergence of the boundaries towards a single point, suggesting that the effect of the vocalic context is to impose a rotation of the boundaries. This suggests that the perceptual representation of the labial/non labial distinction is isotropic (i.e. invariant by rotation). In order to preserve the directionality of the f/s contrast the rotation has to take place in a three-dimensional space around a rotation axis corresponding to stimuli with flat transitions.

The isotropy of the labial/coronal feature gives a precise content to Fant’s enigmatic statement about variability being part of the code. The fact perceptual boundary varies across contexts as the radius of a circle implies that the feature – the perpendicular to the boundary in Fant’s conception- is invariant just as the points on a circle are invariant.

![diagram](image-url)

**Figure 1: Perceptual s/f boundaries (points collecting 50% /f/ and /s/ identification responses) for synthetic CV syllables with varying F2 and F3 transitions as a function of F2 onset and offset frequencies (F3 covaried with F2; adapted from [12]).** The directions of the boundary lines change according to the vowel context but converge towards a single point corresponding to flat transitions in the neutral vowel context (/a/ vowel or “schwa”). The interrupted line corresponds to stimuli with flat F2-F3 transitions.

3 Perceptuo-motor representation of features: more than an interface

“We would all agree that the categorization inherent at the production end is quite similar to that at the perception end of the speech communication chain but only defenders of a motor theory of speech perception would argue that perception is nothing but the association of the incoming acoustic stimulus with production categories at the listener’s disposal when acting as a speaker. [...] However by introspection we can certainly study our own stored sound images of distinctive features and phonemes some of which we might not be able to produce correctly if they belong to a language we are not so well acquainted with.” ([2,3], p.166).

There certainly have been major advances in our conception of the perceptuo-motor link since Fant made this statement. Mainly, the discovery of mirror neurons which can be triggered either by external stimuli or by internal motor activity [1] boosted the interest for motor theories of perception. However, the basic question raised by Fant about the nature of feature representation remains: are featural representations simply an interface between motor commands and acoustic cues or are they more abstract with some degree of autonomy relative to both the motor and the auditory system?

In this view it is interesting to note that in the results presented above (Figure 1), the center of the isotropic representation corresponds to a syllable with flat formant transitions and with formant frequencies corresponding to a vowel sound produced by a uniform vocal tract (with F2 and F3 values at 1.5 and 2.5 kHz). This means that the labial/coronal boundaries in different vocalic contexts converge towards the neutral vowel context (the /a/ vowel or “schwa”), and that the boundary in this context corresponds in a change in the upward/downward direction of formant transitions. Now, enhanced sensitivity to changes in formant transition direction has been evidenced for non-speech sounds (sinewave analogues of ba/da/go syllables: [10]), which indicates that they constitute natural psychoacoustic boundaries. The central point of the labial/coronal representation is thus psychoacoustic in nature. However, psychoacoustics cannot account for the contextual variations of the boundary, indicating that there is a speech specific component in representations of place of articulation features (for similar evidence about the voicing feature cf. [11]).

It would thus seem that the core of the perceptual representation of speech features is psychoacoustic in nature and is directly operational in a neutral articulatory context. However, the generalization of this core principle to other articulatory contexts is achieved through speech specific rules. More than a simple interface between motor commands and acoustic cues, featural representations are autonomous structures integrating both articulatory mappings and psychoacoustic landmarks.

4 Feature systems for vowels and consonants: a common mapping

According to Roman Jakobson [5] differences in place-of-articulation between consonants and vowels are basically similar and should be represented with the same features. However, given that articulatory changes on the front/back dimension have inversed acoustic changes for consonants and vowels, there can be no common representation of the two sound classes of sounds in both articulatory and acoustic terms (Figure 2). A common acoustic
representation will distort the articulatory relationships between the two sound classes (e.g. the front/back b/g contrast will be equivalent to the back/front u/i contrast). Conversely, a common articulatory representation will distort the acoustic relationships between the two sound classes (e.g. the grave/acute b/g contrast will be equivalent to the acute/grave u/i contrast).

In the early investigations on speech perception at the Haskins Laboratories, Liberman, Delattre, Gerstman & Cooper [8] found that the lengthening of formant transitions progressively changed a stop+vowel percept (/b,d,g/+V) into a semivowel+vowel percept (/w,t,j/+ V) and then into a vowel+vowel percept (/u,y,i/+V). This suggests that the place contrasts between stops and those between vowels are conveyed by similar acoustic differences between formant frequencies. However, these similarities are fairly coarse because they are based on the prototypical values of the vowel and consonant categories, not on the boundaries between categories. Prototypical values correspond to the endpoints of the S-shaped identification function and can therefore not be assessed with the same precision as the perceptual boundaries which are located at the middle of the function (50% response point), where the slope of the S-curve is the steepest. Whether or not distinctions between vowels and those between consonants share the same boundaries is the acid test for assessing perceptual similarities.

We recently found evidence in support of a common articulatory representation of vowel and consonant boundaries [13]. Using synthetic /stop+schwa/ and /vowel+schwa/ syllables varying along a circular continuum in a F2-F3 transition onset space, and collecting identification responses by adult French listeners, we found that the identification boundaries between stops and those between closed vowels did not match in the acoustic space (Figure 3). However, the b/d/g and i/y/u boundaries coincided after a 151° rotation of the acoustic space (Figure 4), suggesting a common articulatory representation of the two classes of sounds. Similar results were obtained in different groups of French children [14] and for adult Spanish and Swedish listeners [9].

More than an interface, the autonomy of featural representations pops up again, in accordance with Fant’s view. The representations are based on acoustic dimensions but the topological equivalence between vowels and consonants is based on articulation. However, contrary to Fant’s opinion, there seemingly is a common representation of vowel and consonant features. But it is precisely the autonomy of featural representations, as advocated by Fant, which renders such a common representation possible.
Figure 3: Perceptual boundaries for synthetic /stop+schwa/ and /vowel+schwa/ syllables varying along a circular continuum in a F2-F3 transition onset space. The identification boundaries between stops (top) and those between closed vowels (bottom) do not correspond in the acoustic space [13].

Figure 4: After a rotation of about 150° consonant boundaries coincide with the vowel boundaries [13].

References


