[b]-[d]-[g] as a universal triangle as acoustically optimal as [i]-[a]-[u]

Christian Abry

Institut de la Communication Parlée, CNRS UMR 5009, INPG/Université Stendhal

abry@icp.inpg.fr

ABSTRACT

In the world language databases, [b]-[d]-[g] are the most prevalent places for consonants, as are the point vowels [i]-[a]-[u]. Jakobson, Fant & Halle conceived acoustically these places in parallel with the vowels, as a triangular binary representation, until Chomsky and Halle switched to articulatory features. Lindblom's and our endeavor rests on an acoustic space for vowel systems computational prediction. At ICPhS 79, Ohala put forward the false prediction of dispersion theory for a putative 7-consonants systems: ɗ, k', ts, ɬ, m, r, |. But in search of *consonant-place-space* (CPS), we have to leave aside manner, as it is done in predicting oral but not nasal vowel space together. This means that the proposed ɗ, k', ts, ɬ , m, r, | inventory is reduced to ubiquitous labial, coronal and velar places. It is noticeable that Haskins' patterns could have been represented as a triangle of CV transitions – typically toward [a] – in a F2-F3 plane, where, from the "hub" locus of [g] (F2=F3), F3 is rising while F2 is falling, F2-F3 rising both for [b], and falling both for [d]. What could then be the differentiation process, the genesis of the workspaces for consonants and vowels in this common framework of acoustic coordinates? In canonical babbling, F1 is the audio movement corresponding to the carrier of speech, the mandible, with labial [bababa] or coronal [dadada] "frames". When independence of carried articulators (lip and tongue) from the jaw becomes settled, F2-F3 stream carries the information on places of contact. Thus F2-F3 plane is basically orthogonal to F1. When coarticulation emerges, after one year, the [a] vowel can now be produced during the closure phase, before opening, i.e. vocalic F1 is coproduced. Up to 4 years the mastering of the control of the whole vocal tract for [i] and [u] vowels will be in progress, if, like almost all languages, the mother tongue gets them. The differentiation process is thus comparable within the two streams finally overlapping: F1-F2 mainly for universal [i]-[a]-[u] vowels, and F2-F3 for universal [b]-[d]-[g] consonants. So the Dispersion Theory proves efficient for structuring both acoustic planes: this is an answer to the challenge issued by Ohala, an answer which does not call *ab initio* to his joined proposal of compositionality (Maximum Use of Available Features). Additional third dimensions may be used, F3 for vowels (e.g. to contrast [y] vs. [i] in French), or F1 for consonants (contrasting pharyngealized vs. plain segments in Arabic). In these cases our Dispersion-Focalization Theory is better at issue.

NO ACOUSTIC OUTSPACING FOR CONSONANTS?

In the world languages, labial, coronal and dorsal consonants are the most prevalent places, and they prove to be as universals in evidencing a consonant *place* space as it is the case for the point vowels [i]-[a]-[u]. Taking the UPSID 451 language database, we can find, and this the most clearly for plosives (non-nasal), that bilabial, dental, velar places are present in 99 to 100% of the sound inventories. And this is the same order for nasals (95%, 96%, with less velars 53%). Let's dub this triplet as [b]-[d]-[g], knowing that plosives are 64% unvoiced consonants, but that Africa is clearly in favor of voiced consonants (with a ratio of 1.5). As we will detail below the issue is not manner, but a *consonant-place-space* (CPS). Affricates (and to a lesser extent fricatives) clearly do not fit into this three-place scheme, being mostly coronal. But they are clearly from another type of sensorimotor control than ballistic plosives: [s]-type (which bothered phonologists to the point of gaining a special status in syllabification) is the only one where the jaw is directly audible, i.e. controlled as raised and stabilized to use the projection of the air jet on the lower incisors.

Jakobson, Fant & Halle in their *Preliminaries*, conceived clearly this consonant-place-space: (i) acoustically; and (ii) patterning in parallel with the vowels (with more or less tinkering). This was the binary representation: [g] being [+compact] *vs.* others [+diffuse], [b] [+grave] and [d] [+acute] (this triangle becoming somehow a square with palatal plosive [tʃ], as for vowels when contrasting [a] and [ɑ]). Since Chomsky and Halle switched to an articulatory feature representation, followed by the champion of acoustics at the same MIT, Stevens, such a dominance left little success for people to reintroduce

ISBN 1-876346-48-5 © 2003 UAB

acoustic primes. As a whole, there seems to have been no real interest in finding a cross-talk between vowel and consonant *acoustic* spaces, compared to a disputed endeavor concerning the unification of *articulatory* representations for consonant and vowels.

This trend towards a unification in the computation of representations for speech sounds in acoustics was stopped about a quarter of a century ago, at ICPhS 79, by John Ohala. Being a promoter of the perceptual (*vs.* articulatory) dominance of the sounds of speech, he moderated a symposium on "Phonetic universals in phonological systems and their explanation" (vol. 3, pp. 181-194), where he made several deeply programmatic statements under the heading "A specific problem in phonological universals" (pp. 184-185). His problem deserves detailed quoting.

Ohala: "The notion of vowel "space" has been used in phonetics for about 2 centuries but it is only recent evidence which points to this space having acoustic-auditory correlates. The research of Lindblom and his colleagues suggests that the placement of vowels in this space in various languages is dictated by the principle of maximal perceptual difference, i.e., that however many vowels there are in the system, they tend to arrange themselves in the available space in such a way as to maximize their distance from each other. This principle seems to adequately predict the arrangement of systems with approximately 7 or 8 vowels." That's what continued research demonstrated steadily from Liljencrants & Lindblom to Schwartz and colleagues in our Dispersion-Focalization Theory, etc.

Ohala (cont.): "It would be most satisfying if we could apply the same principles to predict the arrangements of consonants, i.e., posit an acoustic-auditory space and show how the consonants position themselves so as to maximize the interconsonantal distance. Were we to attempt this, we should undoubtedly reach the patently false prediction that a 7 consonant system should include something like the following set: ɗ, k', ts, ɬ, m, r, | ".

Ohala (cont.): "[…] Rather than maximum differentiation of the entities in the consonant space, we seem to find something approximating the principle which would be characterized as "maximum utilization of the available distinctive features" ". This MUAF will form part of the endeavor of Lindblom in predicting phonetic structures beyond vowels.

Ohala (end): "[…] Does this mean that consonant inventories are structured according to different principles from those which apply to vowel inventories? […] Or does it mean that we are asking our questions about segment inventories in the wrong way?"

The purpose of this "yes" answer, to the last question, 25 years late, is the following. (i) To limit the whole question to the *consonant-place-space* (CPS), we have to leave aside manner, as it is done in predicting oral but not nasal vowel space together. Thus the proposed ɗ, k', ts, ɬ, m, r, | inventory reduces to labial, coronal and velar places. This means that there is a mechanism – a pure perceptual, or an inverse mechanism to recover articulatory control from integrated perception (a robotic Motor Theory of Speech Perception), or a mechanism for control recovery understanding the perceptual goal (our last PACT, Perception-for-Action Theory, gaining some support from motor evoked excitability for coronals, attributed to specific tongue mirror neurons, thanks to the recent work of Fadiga and colleagues, from Rizzolatti's group). This does not mean that non-place information has no effect on place. The documented case of confusion of labial and coronal places under nasal conditions, [m] and [n], by blind people, shows, among many other examples, that place is sensitive to manner. (ii) To the first question we will answer that consonants (plosives or ballistic consonants) are different from the vowels, but intertwined in the same acoustic space, along overlapping flows. This does not mean that they are articulatorily of the same control types: to be short, plosives rest on a contact-pressure sensorimotor information; whereas vowel positions or configurations, use a postural (or figural) control of the whole vocal-tract; sibilants being both orotactile and oropostural.

AN ACOUSTIC CONSONANT PLACE SPACE (CPS)

Just remember the patterning of vowels in the F1-F2-F3 space (Fig. 1 builds on typological elaborations and articulatory-model-to-acoustic-model simulations from our lab.). The classical F1-F2 space is sufficient to express the distances between the point vowels [i]-[a]-[u]. But obviously not for [i] *vs.* [y], which we called a *focal* vowel. This focal vowel is not present in all sound systems (8%). Like focal colors, focal vowels have universal properties but not ubiquity. Perceptual transforms in this 3D space can account for focalization, i.e. formant grouping (hence salience): F1-F2 for [a] and [u], F2-F3 for [y], and F3-F4 for [i] (with F4 constant for simplicity). Anyway as a first step, the gray floor on Fig. 1, can be considered as the basic acoustic vowel position space (VPS).

It is noticeable that available Haskins' patterns, for their Play-back system, allowed to represent – long before locus equations – a triangle of CV transitions. When moving typically toward [a] – in a F2-F3 plane – we can take as a starting point the "hub" locus at about [g] release, where F2 equals F3 (Fig. 2). From there, into [ga], F3 is rising while F2 is falling. The pattern for [b] shows F2-F3 rising together; while they are both falling for [d]. Plotting these starting points (C-Release values) in Fig. 3, on the F2-F3 plane, it is possible to set in this rectangle triangle of consonantal audio movements: [g] at the 90° angle, moving to [a], close to the hypotenuse, while [d] and [b] are also converging to [a] along this hypothenuse.

Voiced consonants have in common their F1 motion, which is basically of no use to parse their place, but only their manner, i.e. their opening audio movement, from wall vibrations (180 Hz) to F1 vowel target value. This is schematically shown on Fig. 3 by projections on the F1-F2 plane, below the lowest F1 values for high vowels.

In speech development, F1 is the audio movement corresponding to the carrier of speech, the mandible, with typical labial [bababa] or coronal [dadada] "frames" (less [gVgV]), following MacNeilage's proposal. When independence of carried articulators (lip and tongue) from the jaw becomes to settle, the F2-F3 stream carries the information on places of contact. Thus the F2-F3 plane is developmentally orthogonal to F1. When coarticulation emerges, after one year, the [a] vowel can now be produced during the closure phase, before opening: i.e. vocalic F1 is *coproduced*. Up to 4 years the mastering of the control of the whole vocal tract for [i] and [u] vowels will be in progress, if, like almost all languages, the mother tongue gets them. The differentiation process is thus comparable within the two streams (the Where-When-What of scene analysis via the common destiny imposed by vocal-tract cavities). They will be finally overlapping (intertwined): F1-F2 recruited mainly for universal [i]-[a]-[u] vowels, and F2-F3 for universal places of consonants, say [b]-[d]-[g]. So the Dispersion Theory proves efficient for structuring both acoustic planes. This is an answer to the challenge launched by Ohala, as reported above. Moreover this answer does not call *ab initio* for compositionality, like the Maximum Use of Available Features principle. Of course, these two orthogonal triangles in the F1-F2-F3 space adjoin each other, since they are generated by the acoustic theory of the vocal-tract tube with one or two pinches. Our F2-F3 triangle with the maximum transitions (Fig. 4) can be obtained once [a] coarticulation with [b]-[d]-[g] is emerging, i.e. as early as the beginning of the second year. An additional dimension may be used, F3 for vowels (e.g. to contrast [i] *vs.* [y], like in French), or F1 for consonants (contrasting pharyngealized *vs.* plain segments, as in Arabic). In these cases our Dispersion-Focalization Theory is better at issue.

The connectivity of the two planes mirrors the coordination of the two streams: i.e. coarticulation, as a result of coproduction, will end in substantial modifications of these extreme vowel and consonant positions, and finally in some current consonant-to-vowel changes. This is a conclusion to escape the invariance issue which has yet produced beautiful diagrams like Lindblom's from Öhman, Sussman's locus equations (improved for dorsals, by Bailly in our lab, with formant affiliations), Nearey's logit maps from Haskins, etc. Everybody knows that bayesian learning of density probabilities became the cleanest state-of-the-art statistical solution for HMM and, probably, Japanese quails.

Aknowledgements: to Louis-Jean Boë and Jean-Luc Schwartz who supported this solo (I hate signing alone) benefiting from a lot of work from people at ICP. Apologies to all scientists whom I mentioned with admiration, but without references… And to all to whom I simply alluded, with the same admiration (for people, not for HMM and Japanese quails!).

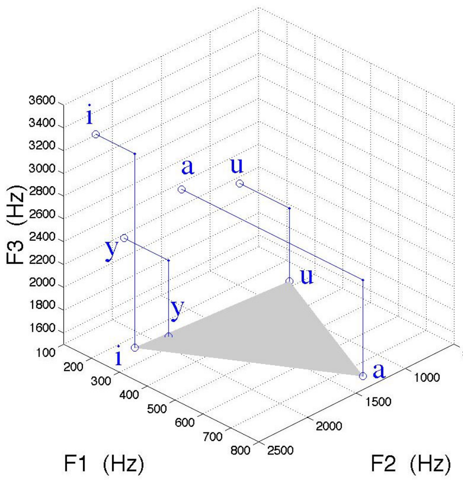

Figure 1: Three formants (F) vowel position space (VPS). In gray the classical acoustic vocalic triangle F1-F2.

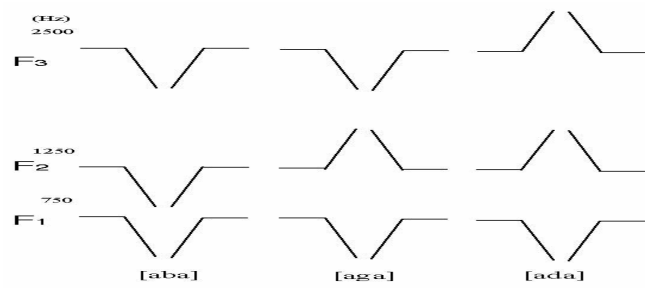

Figure 2: Schema of VCV transitions for [aba], [aga] [ada]. All F1 transitions are the same for the common voiced manner. At about consonant release, F2-F3 are changing together with the same slope, rising for [b], falling for [d]. At the center, the presentation enhances the "hub" locus effect for [g] where F2-F3 come close together.

ISBN 1-876346-48-5 © 2003 UAB

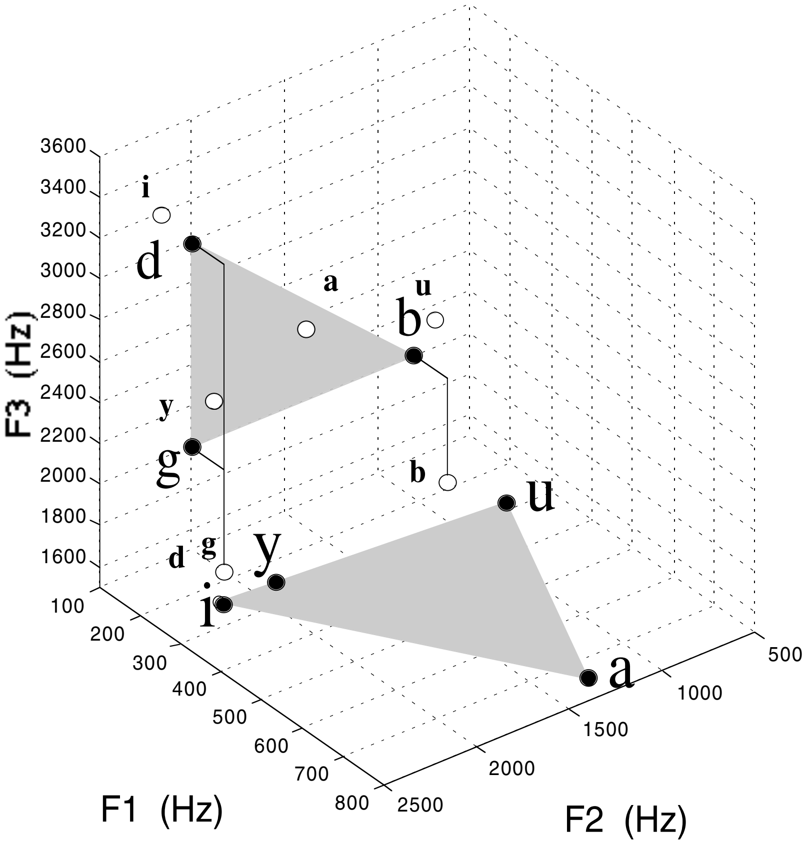

Figure 3: Three formants (F) consonant place space (CPS). In gray, the two triangles: F1-F2 for vowels and F2-F3 for mainstream consonants. Filled circles are focal vowels and maximally distant consonants (smaller phonetic symbols near empty circles are for projections of vowels in the F2-F3 plane and for wall vibration limit F1 frequency for consonants).

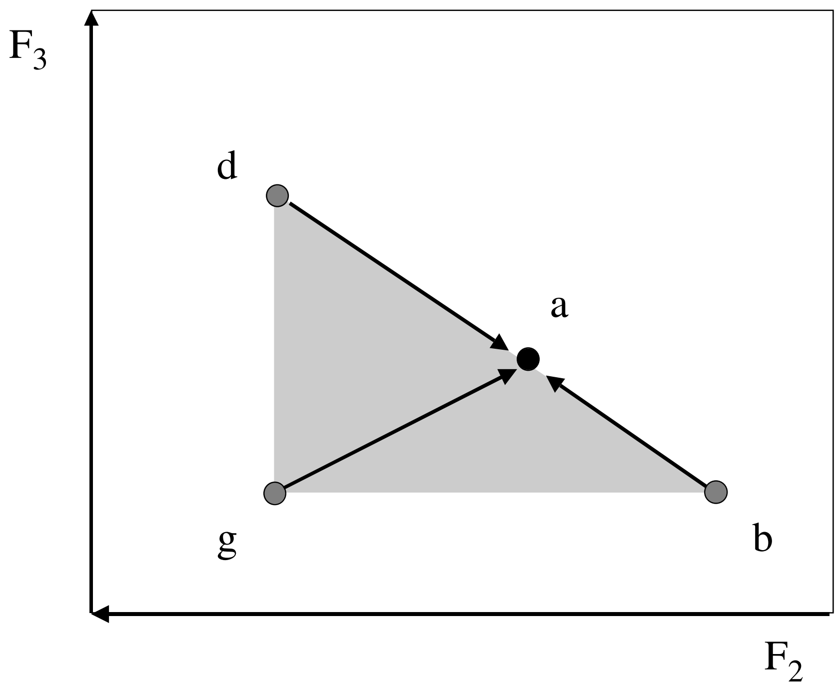

Figure 4: Schema of the typical in-triangle trajectories, i.e. audio movements of [b], [d], [g] toward [a], in the F2-F3 consonant place space (CPS).