

Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory

Jean-Luc Schwartz(1), Christian Abry(1), Louis-Jean Boë(1),
Lucie Ménard (2), Nathalie Vallée (1)

(1) Institut de la Communication Parlée,
CNRS-INPG-Université Stendhal, Grenoble, France

(2) Université du Québec à Montréal,
Département de linguistique et de didactique des langues, Montréal, Canada

Abstract

In a recent paper in this journal, Polka & Bohn (2003) report on a robust asymmetry effect in vowel discrimination, present in infants as well as adults. They interpret this effect as a preference for peripheral vowels, providing an anchor for comparison. We discuss their data in the framework of the Dispersion-Focalisation Theory of vowel systems. We show that focalisation, that is the convergence between two consecutive formants in a vowel spectrum, is likely to provide the ground for anchor vowels, by increasing their perceptual salience. This enables to explain why [y] is an anchor vowel, as well as [i], [a] or [u]. Furthermore, we relate the asymmetry data to an old experiment we had done on the discrimination of focal vs. non-focal vowels. Altogether, it appears that focal vowels, more salient in perception, provide both a stable percept and a reference for comparison and categorisation.

Keywords: vowel discrimination, peripheral vowels, focalisation, salience, stability, anchor.

1. Introduction: Global vs. local constraints in substance-based theories of phonological systems

In the search for substance-based principles shaping phonological systems, both global and local constraints have been considered in the literature. *Global* constraints are based on the relations between elements in a system, so that a given gesture/sound/percept is included in the system not because of its own properties, but because of its contribution to a global function in relation with the other elements of the system. A prototypical example is provided in Lindblom's Dispersion Theory DT (Liljencrants & Lindblom, 1972; Lindblom, 1986), later becoming the Theory of Adaptive Variability TAV (Lindblom, 1990; and a revised version in Diehl et al., 2003). In this framework, distinctiveness of the units within the system is the driving force, and hence units must be as different as possible one from another, in perceptual terms. Therefore, "a selected unit is highly valued, not because of its individual qualities, but depending on its contribution as a team player" (Lindblom, 2003). On the contrary, *local* constraints should result in focusing the selection towards specific regions in the articulatory-acoustic-perceptual space, preferred for intrinsic properties, independent of the properties and configurations of the other elements in the set. This provides the basic rationale for Stevens' Quantal Theory QT (Stevens, 1972, 1989) in which non-linearities in the articulatory-to-acoustic or acoustic-to-auditory transforms define natural boundaries that would be exploited by phonological contrasts.

If such local attraction regions do exist, and in some sense "pre-exist" to linguistic systems, then it should be possible to display their existence through adequate non-linguistic or pre-linguistic experimental paradigms. This line of reasoning has produced some striking successes, particularly concerning two consonantal contrasts, that is place of articulation and voicing for plosives. Just to mention the second one, the existence of a "natural" boundary between unvoiced and voiced plosives has received support from VOT (Voice Onset Time) categorical experiments. These experiments involved either animals (Kuhl & Miller, 1978) or prelingual infants (Eimas et al., 1971). Both experiments displayed categorical perception with increased discrimination around the boundary between voiced and unvoiced plosives, though language was not (for animals) or not yet (for infants) present. The same kind of results was obtained with a non-linguistic continuum "mirroring" the linguistic one (TOT, Tone Onset Time: Miller et al., 1976).

The situation is not so clear for vowels, for which categorical perception does not seem to apply (Repp, 1984). Discrimination experiments were used by Kuhl to introduce the "perceptual magnet effect", according to which some regions of the acoustico-perceptual space could provide anchor points for categorisation (called "magnets"), both for adults and 6-month-old infants ... though not for animals (Kuhl, 1991). But it appeared later that these regions were in fact the product of a *learning phase* from 0 to 6 months old. Indeed, different magnet regions were observed for 6-month-old infants of different languages, and these regions were related to adult prototypes in the corresponding language (Kuhl et al., 1992). Therefore, the magnet effect characterizes the tuning to a specific language under exposition, rather than universal local constraints on vowel systems.

A few years ago, we introduced a new theory for the prediction of vowel systems, integrating global and local peripheral constraints on the shaping of vowel inventories. This theory, called the Dispersion – Focalisation Theory (DFT), includes both the global dispersion ingredient exploited by Lindblom and colleagues, and an additional local property called focalisation, related to preferred regions in the perceptual space (Schwartz et al., 1997a; see 2.2). The DFT was shown to predict quite well the major characteristics of existing vowel inventories in the world languages (Schwartz et al., 1997b; Vallée et al., 1999).

In this context, we read with enthusiasm a recent paper in this journal by Polka & Bohn (2003) in which they summarise in a clear and striking way a series of experimental data from themselves and others, consistently showing the existence of asymmetries in vowel perception (see 2.1). The authors interpret these asymmetries as a predisposition for more peripheral vowels in the F1-F2 space that could provide a perceptual anchor for vowel systems. These regions could in fact, in our view, be better described in the framework of the DFT, and seem to provide an interesting argument in favour of the theory, and particularly its “focalisation” component. The purpose of the present paper is hence to propose a reinterpretation of the paper by Polka & Bohn within the DFT. Their data, together with the DFT sketch, will be briefly recalled in Section 2. Reinterpretation will be done in Section 3, before a conclusion section.

2. Recalling Polka & Bohn (1994-2003) and the DFT (1997)

2.1. *Asymmetries in vowel perception*

In a study of vowel discrimination in infants, Polka & Werker (1994) discovered a robust effect of order of presentation, in which a given direction of change (e.g. [y] before [u]) was better discriminated by English infants than the reversed order, for both 6-8 and 10-12 months olds. First assuming that this was a consequence of the magnet effect, in which the familiar [u] (for English listeners) would act as an anchor point, Polka & Bohn (1996) later discovered that a vowel in a pair could play the role of an anchor point independently of the status of the vowels in the listener’s phonological system. In their recent paper, Polka & Bohn (2003) summarise all occurrences of such asymmetries in vowel discrimination in infants, among various studies in the literature. They conclude that these asymmetries could “reveal a language-universal perceptual bias that infants bring to the task of vowel discrimination”, and suggest that “the more peripheral vowel within a contrast serves as a reference or perceptual anchor” (p. 221).

2.2. *The Dispersion-Focalisation Theory of vowel systems*

2.2.1 *Dispersion of (F1, F’2) perceptual patterns in a 3D formant space (F1, F2, F3)*

In the dispersion theory, it is classical to describe vowels by two acoustico-perceptual parameters. The first parameter is the first formant F1. The second one is F’2, an “effective second formant” describing in a non-linear way the combined effect of F2 and higher formants. In a series of experimental and modelling works (Escudier et al., 1985; Schwartz and Escudier, 1987), we showed that F’2 was likely to be the product of a 3.5 Bark “Large-scale Spectral Integration” (LSI) mechanism, earlier proposed by Chistovich and colleagues (Chistovich & Lublinskaya, 1979; Chistovich et al., 1979), grouping F2, F3 and F4 in a major peak above F1, well correlated with experimental F’2 values. In the DFT, vowels are described as four-formant spectra with F4 fixed at a value typical of F4 for male speakers (3560 Hz), and F1, F2 and F3 varying within the available “maximal vowel space” providing a 3D extension of the F1-F2 “vowel triangle” (Boë et al., 1989). (F1, F2, F3) triplets are converted into (F1, F’2) pairs through a simplified LSI model computing F’2 as a function of F2, F3 and F4. Dispersion is then provided by summed weighted distances between (F1, F’2) values of vowel pairs in the considered system.

2.2.2. *Focalisation as a perceptual goal*

The convergence between consecutive formants (F1, F2), (F2, F3) or (F3, F4) is a basic component of articulatori-acoustic-perceptual stability in Stevens' quantal theory (Stevens, 1972, 1989). This is what we called "focalisation" (Boë & Abry, 1986), and we showed that the LSI mechanism was at the basis of the stability of focal vowels, decreasing spectral variability and increasing acoustic salience (Abry et al., 1989). This led us to propose that focalisation, that is formant convergence, was a property of spectral configurations that could provide a benefit for speech perception. It could make the corresponding spectrum easier to memorise and to process, more focal spectral configurations being preferred to less focal ones in vowel systems: we shall come back to this in Section 3.1.2. Focalisation in the DFT is a function of individual spectra, summing the inverse of squared distances $(F2-F1)^2$, $(F3-F2)^2$ and $(F4-F3)^2$, for each vowel in the system.

Altogether, the DFT principle consists in minimising, for a given number of vowels in a vowel system, an energy function summing two terms, that is a structural dispersion term based on inter-vowel perceptual distances, and a local focalisation term based on intra-vowel perceptual salience, which aims at providing perceptual preferences to vowels showing a convergence between two neighbouring formants.

3. Reinterpreting vowel perceptual asymmetries within the DFT

Our reinterpretation of the Polka & Bohn's work within the DFT is focussed on two basic questions that remain open in their paper. Firstly, we discuss the mechanism that could enable an infant to characterise the peripheral nature of a given vowel. Secondly, we address the case of [y], which raises a problem to their interpretation. In both cases, the concept of focalisation seems to provide a rather attractive and efficient solution.

3.1. *Focalisation grounding anchor vowels in Polka & Bohn*

Polka & Bohn (2003) summarise their finding by noting that "vowel discrimination is easier for infants when they were presented a change from a less peripheral to a more peripheral vowel", and define more peripheral by "closer to the limits or corner of the vowel space" (p. 224): this is displayed by their Fig. 1a, reproduced here in Fig. 1. Now the question is: what enables infants (and adults) to determine that a vowel is more peripheral than another one? Polka & Bohn mention that this ability seems uniquely human, according to the comparative data on animals they provide in their Fig. 1b. They suggest that it could be innate as well as learnt from the exposure to speech, and particularly its motherese version increasing hyperarticulation and peripherality. But the question remains: *what* characterises a vowel as more peripheral – and, by the way, *why* does it provide an anchor for discrimination?

3.1.1. *What characterizes a vowel as more peripheral?*

The "*what*" question can first be considered as a problem of geometry of the vowel space. All vowels are inside a triangle in the (F1, F2) plane. A triangle is specified by its three sides, which are geometrical segments that can be defined by affine equations in the plane. On Fig. 2, we have plotted typical (F1, F2) values for point vowels [i], [a] and [u] in Barks. The three triangle sides in Fig. 2 are defined by Equations (in Bark) " $F1=constant$ " (segment joining [i] and [u]), " $F2-F1=constant$ " (segment joining [a] and [u]) and " $F1+F2=constant$ " (segment joining [a] and [i]). This provides three spectral features, namely F1, associated to the high-

low dimension and minimum for the [i]-[u] boundary, $F2-F1$, called “spread” feature in Fant (1983) and minimum for the [a]-[u] boundary, and $F1+F2$ (or $-(F1+F2)$), called “flat” feature in Fant, 1983), maximum for the [i]-[a] boundary. These boundaries are superimposed to the experimental data from Polka & Bohn in Fig. 1 (though in an approximate way, since frequencies were expressed in Hz and not Bark by the authors in their original Fig. 1a). If we consider the corners as “more peripheral” than the other points in the sides, we just need to combine two criteria. For example, [i] would optimise both the $max(F2+F1)$ and the $min(F1)$ criteria.

Then, we may notice that, because of intrinsic constraints on the repartition of $F1$, $F2$, and $F3$, front vowels, which have a high $F2$ and hence a high $(F2+F1)$, also have a small $(F3-F2)$: thus, the second criterion could as well be expressed as $(F3-F2)$ minimum. Actually, the $(F3-F2)$ distance seems to play an important role in the characterisation of vowel sounds (Syrdal, 1985).

In the DFT, we consider the $(F1, F2, F3)$ vowel space, determined from an articulatory model of the vocal tract (Boë et al., 1989). On Fig. 3, we plot this 3D vowel space, together with its $(F1, F2)$ and $(F2, F3)$ projections. It appears that this space has four natural corners, that is [i], [u], [a] and [y] ... which are precisely the major “winners” in the asymmetry report by Polka and Bohn (2003) in their Table 1. In our first work on the articulatori-acoustic characterisation of focalisation, we studied various articulatory nomograms in the framework of Fant’s four-tube model of the vocal tract (Badin et al., 1991). The conclusion was that four vowels could be pure focal vowels, able to display an almost perfect formant convergence. These were:

- [u], with possibly equal values of the Helmholtz resonances of the back and front cavities ($F1=F2$)
- [a], with possibly equal values of the Helmholtz resonance of the back cavity, and of the quarter-wavelength resonance of the front cavity ($F1=F2$)
- [y], with possibly equal values of the half- wavelength resonance of the back cavity, and of the Helmholtz resonance of the front cavity ($F2=F3$)
- [i], with possibly equal values of the quarter-wavelength resonance of the front cavity, and of the resonance of the constriction ($F3=F4$).

Other systems of resonances and affiliations have been proposed, with two- or three-tube models (see e.g. Stevens, 1989), with basically the same kinds of pivot configurations. On Fig. 3, we have superimposed 3 planes respectively of equations $(F1=F2)$, $(F2=F3)$ and $(F3=F4)$. Focalisation in the DFT is expressed by attraction of vowels towards these planes, which results in minimising either $(F2-F1)$, or $(F3-F2)$, or $(F4-F3)$. This attracts vowels towards the peripheral sides, and particularly the corners [i], [u], [a] and [y]. Therefore, focalisation could provide the cue to peripherality responsible for asymmetries observed by Polka & Bohn: the closer two neighbouring formants, the more attractive the vowel as a reference for discrimination.

3.1.2. Why should peripheral vowels provide an anchor for discrimination?

Focalisation could also enable us to answer the “why” question. To make this clear, let us briefly summarize an experiment published in Schwartz & Escudier (1989) in this journal, which provided for the first time evidence for the possible role of focalisation in vowel perception.

This experiment dealt with synthetic stationary stimuli with fixed $F0$ (100 Hz), $F1$ (450 Hz), $F2$ (2000 Hz), and $F4$ (3350 Hz), and various $F3$ values from a position close to $F2$ ($F3=2300$ Hz) to a position close to $F4$ (3000 Hz). Firstly these stimuli were identified by a set of French listeners, and it was checked that they were all consistently perceived as a mid-high front unrounded [e] vowel. Hence this corpus was “phonemically homogeneous”, with

no category boundary within it. Then, a discrimination experiment was performed on pairs of stimuli with different F3 values. It appeared that patterns with the greatest formant convergence (that is, F3 close to either F2 or F4, at a distance of 300 Hz or typically 1 Bark) were more stable in short-term memory. They produced a lower level of false alarms, while patterns with less convergence, namely with F3 at an equal distance from both F2 and F4, were more difficult to memorise, with a significantly higher number of false alarms. These differences in short-term memory could not be due to differences in phonological encoding, since the F3 continuum was phonemically homogeneous. The interpretation was hence perceptual rather than phonemic. Focalisation was considered to be the main determinant of the subjects' behaviour, more focal stimuli with F3 close to either F2 or F4 being supposed to be more salient, and therefore more stable in short-term memory.

This result is reminiscent of the work by Rosch-Heider in the 70s, about universals in colour naming and memory. In a series of experiments (Rosch-Heider, 1972), the author showed that there are specific areas of the colour space (defined in terms of hue, value and saturation), which are more accurately remembered (both in short-term and long-term memory) independently of the language and its corresponding categories. These areas would form "the focal points of basic colour naming across languages" (p. 11). Her conclusion is a wonderful definition of what we previously called prelinguistic local constraints on language: "far from being a domain well suited to the study of the effects of language on thought, the colour space would seem to be a prime example of the influence of underlying perceptual-cognitive factors on the formation and reference of linguistic categories" (p. 20). Formant convergence results in prominent spectral peaks that would make it easier to perceive the sounds, just as, in the domain of colour vision, saturated colours are easier to differentiate than muted ones. Our proposal is that more *peripheral* vowels in Polka & Bohn's experiments are in fact more *focal* vowels, and that more focal vowels within a contrast would serve as a reference or perceptual anchor for discrimination, thanks to their increased auditory salience.

Moreover, Schwartz & Escudier (1989) found exactly the same kind of asymmetries as in Polka & Bohn (2003), with a better discrimination of pairs contrasting a less focal vowel and a more focal one, when the less focal vowel was presented first. In Fig. 4, we summarise the pattern of results displayed in their work, with both high levels of false alarms in the middle of the F3 continuum, and significant asymmetries with better discrimination in the sense displayed by the arrows: discrimination was better from II to I than from I to II, and from II to III than from III to II. Actually, while Schwartz & Escudier (1989) were able to explain false alarms in the paper (by assuming a better stability of focal vowels), they had admitted their inability to explain asymmetries: "consistent asymmetries occur ... At present, we have no strong enough explanation for this interesting fact, which also occurred in Repp et al., 1979, but within a phonetically non-homogeneous corpus" (Schwartz & Escudier, 1989, Table 5 and p. 250). This is a very important point, since these data cannot be explained by a criterion such as "more vs. less peripheral", while focalisation is here a direct cue, likely to increase the perceptual salience of spectral patterns I or III vs. II in Fig. 4.

Altogether, a focal vowel would provide an anchor for discrimination experiments, including a stable percept (displaying a low level of false alarms) and a magnet percept playing the role of a reference for comparing sounds.

3.2. *The case of [y]*

In Polka & Bohn's framework, [y] is considered as a peripheral vowel (thanks to its minimum F1 value, together with [i] and [u]), hence its role of anchor vowel in the [y]-[ʊ] pair, displayed on Fig. 1. But it should *not* play the role of an anchor in the [i]-[y] pair. This is

however what happened in the only study considered by Polka & Bohn (2003) as contradictory with their assumptions, that is a study of English infants at 3-5 months of age, tested on a Norwegian [i]-[y] pair by Best & Faber (2000). They found that infants could discriminate the contrast in the [i] to [y] direction, but not in the inverse direction. This is particularly striking if one considers that it is one of the rare cases, in the asymmetries reported by Polka & Bohn (2003) in their Table 1, in which the anchor [y] *does not* belong to the listener's system, while the non-anchor [i] *does*.

The vowel [y] naturally emerges from the DFT as one of the four favoured ones (together with [i], [a] and [u]) owing to its strong (F2-F3) convergence: typically, the F3–F2 distance is about 1 Bark for [y] in Fig. 3, and it is around 1 to 1.5 Bark in various languages (Schwartz et al., 1993). Hence, the anchor vowel here seems to be selected neither as a learnt prototype, nor as a more peripheral sound, but because of the perceptual salience provided by its focal nature. Of course, it remains to be understood why the F2-F3 focalisation for [y] may overcome the F3-F4 focalisation for [i] in the Best & Faber's study.

The vowel [y] has provided a pivot of our reasoning in the DFT since the beginning (see Schwartz et al., 1997a, p. 261). As a matter of fact, it raises a serious problem for the DT, (even in its most recent version in Diehl et al., 2003), since the DT cannot likely predict the existence of an [i] vs. [y] contrast in vowel systems, while [y] is stabilised by focalisation in the DFT. In this context, it is important to mention the recent data obtained by Ménard et al. (submitted) on the acquisition of the French vowel system by French children. Indeed, the differences in anatomy between children and adults, with the smaller ratio of the back cavity length reported to the front cavity length in children, should result in a “defocalisation” of [y], separating F2 (affiliated to the front cavity) and F3 (affiliated to the back one). However, Ménard et al. observe that all speakers from 4 to 39 years old produce a focal [y] with close F2 and F3 (around 1 to 1.5 Bark). This means that children compensate for their smaller back cavity size by slightly fronting the constriction. This results in a higher F2 value (closer to F3), and hence a higher F'2 value, which might compromise the perception of the vowel as rounded. Actually, this is what happens, and for the youngest 4 years-old child in the study, [y] is perceived as [i] by a relatively large number of adult listeners in a perceptual test. Therefore, this study shows that the (F2-F3) focalisation for [y] is part of the speech production goal, in spite of the risk to produce a too acute sound perceived as unrounded.

In summary, [y] appears as a focal anchor vowel both in production (Ménard et al., submitted) and in perception (Polka & Bohn, 2003). Of course, its proximity with [i] in (F1, F'2) terms explains why it remains a relatively unfavoured vowel in vowel systems (present in only 8% of the systems in the UPSID database, see Schwartz et al., 1997b), because of the other pressure applied on sound systems: the systemic global dispersion cost which makes the [i]-[y] pair rather poor in terms of auditory distance and perceptual distinctivity (Robert-Ribes et al., 1998).

4. Concluding remarks

Polka & Bohn (2003) provide a perspective in which there would be a “default structure” of vowel systems, language being formed against this default structure. This corresponds nicely to the two components of the DFT, focalisation for the default structure, and dispersion for the patterning of each individual system in the default structure. Our interpretation of the Polka & Bohn's study within the DFT hence provides a possible interpretation framework for their data – which seem to give in return some evidence for the theory. Of course, much remains to be done to better understand and more precisely specify the exact computation of focalisation in the listener's brain. Moreover, independently of our analysis, all the questions

that Polka & Bohn formulate in their discussion remain open, and particularly the innate vs. acquired one. Focalisation itself could as well be an innate or acquired ability, and no work has been done on this topic, or on possible preference for focal sounds in animals. However, we believe that focalisation, as well as asymmetries and variations in levels of false alarm in discrimination tasks, should be included in further analyses and theories of vowel systems, their ontogeny and their phylogeny, in the search for the default structure of vowel production/perception.

Bibliography

- Abry, C., Boë, L.J., & Schwartz, J.L. (1989). Plateaus, catastrophes and the structuring of vowel systems. *Journal of Phonetics* 17, 47-54.
- Badin, P., Perrier, P., Boë, L.J., & Abry, C. (1991). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *J. Acoust. Soc. Am.* 87, 1290-1300.
- Best, C.T., & Faber, A. (2000). Developmental increase in infants' discrimination of nonnative vowels that adults assimilate to a single native vowel. *International Conference on Infant Studies*, Brighton, UK.
- Boë, L.J., & Abry, C. (1986). Nomogrammes et systèmes vocaliques. *Actes des 15èmes journées d'étude sur la parole, Société Française d'Acoustique*, 303-306.
- Boë, L.J., Perrier, P., Guérin, B., & Schwartz, J.L. (1989). Maximal vowel space. *Proc. of Eurospeech* 89, 281-284.
- Chistovich, L.A., & Lublinskaya, V.V. (1979). The center of gravity effect in vowel spectra and critical distance between the formants. *Hearing Research* 1, 185-195.
- Chistovich, L.A., Sheikin, R.L., & Lublinskaya, V.V. (1979). 'Centers of gravity' and the spectral peaks as the determinants of vowel quality. In *Frontiers of Speech Communication Research* (B. Lindblom & S. Ohman, eds.), pp. 143-158. London: Academic Press.
- Diehl, R., Lindblom, B., & Creeger, C. (2003). Increasing realism of auditory representations yields further insights into vowel phonetics. *Proc. 15th ICPhS*, poster.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303-306.
- Escudier, P., Schwartz, J.L., & Boulogne, M. (1985). Perception of stationary vowels: internal representation of the formants in the auditory system and two-formant models. *Franco-Swedish Seminar, Société Française d'Acoustique, Grenoble*, pp. 143-174
- Kuhl, P.K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93-107.
- Kuhl, P.K., & Miller, J.D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *J. Acoust. Soc. Am.* 63, 905-917.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception by infants by 6 months of age. *Science* 255, 606-608.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language* 48, 839-862.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In *Experimental Phonology* (J.J. Ohala and J.J. Jaeger, eds.), pp. 13-44. New-York: Academic Press.
- Lindblom, B. (1990). On the notion of possible speech sound. *Journal of Phonetics* 18, 135-152.
- Lindblom, B. (2003). Pattern,s of phonetic contrast: towards a unified explanatory framework. *Proc. 15th ICPhS*, 39-42.
- Ménard, L., Schwartz, J.L., & Boë, L.J. (submitted). Production-perception relationships during vocal tract growth for French vowels: analysis of real data and simulations with an articulatory model. *Journal of Phonetics*.
- Miller, J.D., Wier, C.C., Pastore, R.E., Kelly, W.J., & Dooling, R.J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *J. Acoust. Soc. Am.* 60, 410-417.
- Polka, L., & Bohn, O-S (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *J. Acoust. Soc. Am.* 95, 1286-1296.

- Polka, L., & Bohn, O-S (2003). Asymmetries in vowel perception. *Speech Communication* 41, 221-231.
- Polka, L., & Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *J. Exp. Psychol. : Human Percept. Perform.* 20, 421-435.
- Repp, B.H., Healy, A.F., & Crowder, R.G. (1979). Categories and context in the perception of isolated steady-state vowels. *J. Exp. Psychol. : Human Percept. Perform.* 5, 129-145.
- Repp, B.H. (1984). Categorical perception : issues, methods, findings. In N.J. Lass (ed.) *Speech and Language 10. Advances in basic research and practice* (pp. 243-335). New York : Academic Press.
- Robert-Ribes, J., Schwartz, J.L., Lallouache, T., & Escudier, P. (1998). Complementarity and synergy in bimodal speech: auditory, visual and audiovisual identification of French oral vowels in noise. *J.I Acoust. Soc. Am.* 103, 3677-3689.
- Schwartz J.L., & Escudier, P. (1987). Does the human auditory system include large scale spectral integration?. In M.E.H. Schouten (ed.) *The Psychophysics of Speech Perception* (pp. 284-292). *Nato Asi Series*, Dordrecht : Martinus Nijhoff Publishers.
- Schwartz, J.L., & Escudier, P. (1989). A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. *Speech Communication* 8, 235-259.
- Schwartz, J.L., Beautemps, D., Abry, C., & Escudier, P. (1993). Interindividual and cross-linguistic strategies for the production of the [i] vs [y] contrast. *Journal of Phonetics* 21, 411-425.
- Schwartz, J.L., Boë, L.J., Vallée, N., & Abry , C. (1997a). The dispersion-focalization theory of vowel systems. *Journal of Phonetics* 25, 255-286.
- Schwartz, J.L., Boë, L.J., Vallée, N., & Abry , C. (1997b). Major trends in vowel system inventories. *Journal of Phonetics* 25, 233-254.
- Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human Communication: A Unified View* (E.E.Davis Jr. & P.B.Denes, eds.), pp. 51-66. New-York: Mc Graw-Hill.
- Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics* 17, 3-45.
- Syrdal, A. (1985). Aspects of a model of the auditory representation of American English vowels. *Speech Communication* 4, 121-135.
- Vallée, N., Schwartz, J.L., & Escudier, P. (1999). Phase spaces of vowel systems : A typology in the light of the Dispersion-Focalisation Theory (DFT). *Proc. of the XIVth International Congress of Phonetic Sciences*, 1, 333-336.

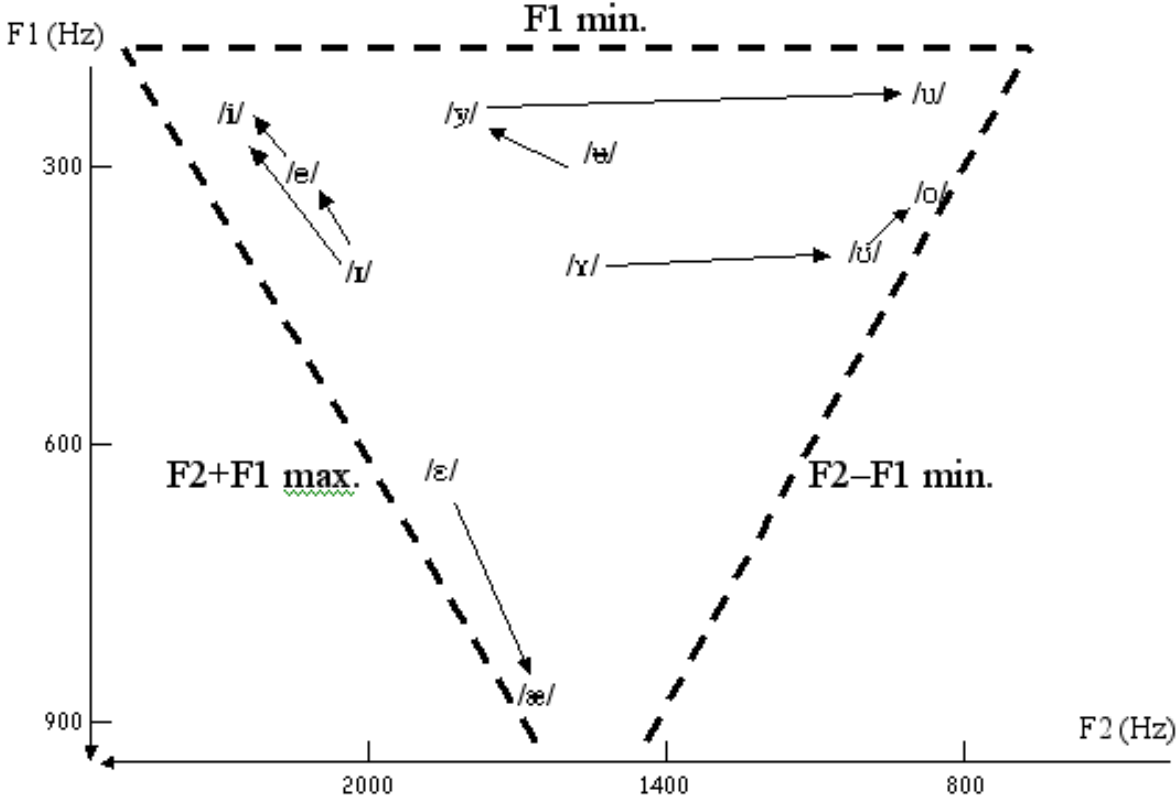


Figure 1 – Plot of F1/F2 frequencies for contrasts showing asymmetries in vowel discrimination with arrows pointing to the reference for the contrast: vowel changes in this direction were easier to discriminate by infants (from Polka & Bohn, 2003, Fig. 1a). The three criteria characterising the boundaries of the vowel triangle are also displayed on the figure.

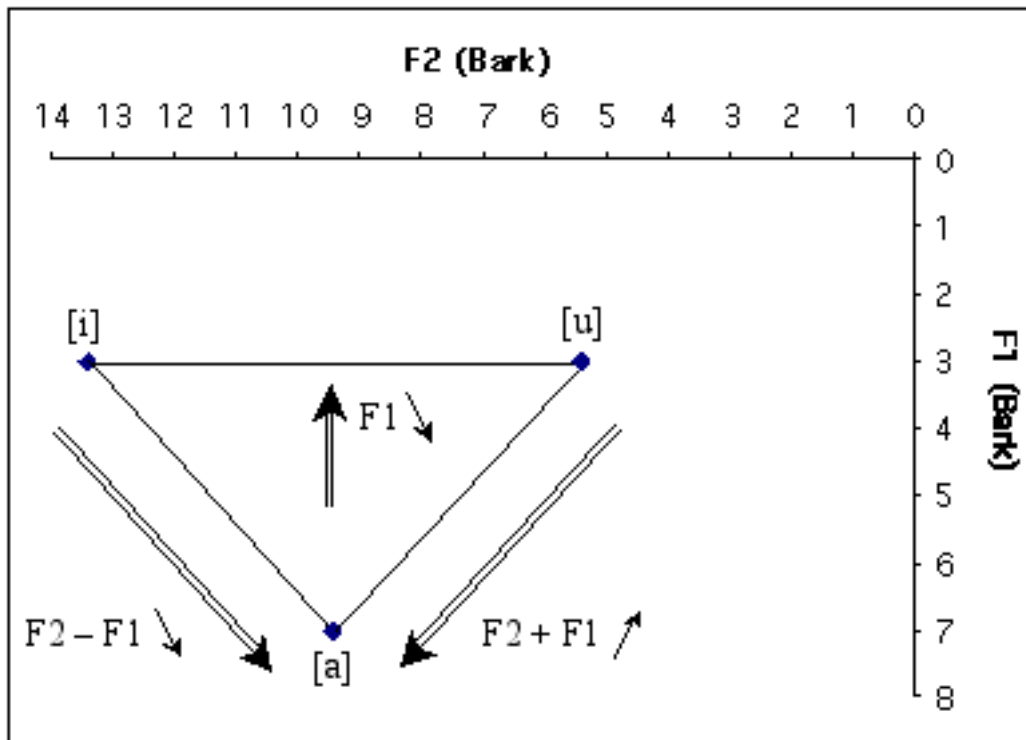


Figure 2 – The geometry of vowel triangle in the (F1, F2) plane. Typical values are provided for [i] (F1=3 Bark or 300 Hz, F2=13.5 Bark or 2200 Hz), [a] (F1=7 Bark or 750 Hz, F2=9.5 Bark or 1200 Hz) and [u] (F1= 3 Bark or 300 Hz, F2=5.5 Bark or 600 Hz). The triangle can be defined by the specification of its three sides: in this typical example, (F1=3 Bark) (segment [i]-[u]), (F2-F1=2.5 Bark) (segment [u]-[a]), and (F2+F1=16.5 Bark) (segment ([i]-[a])).

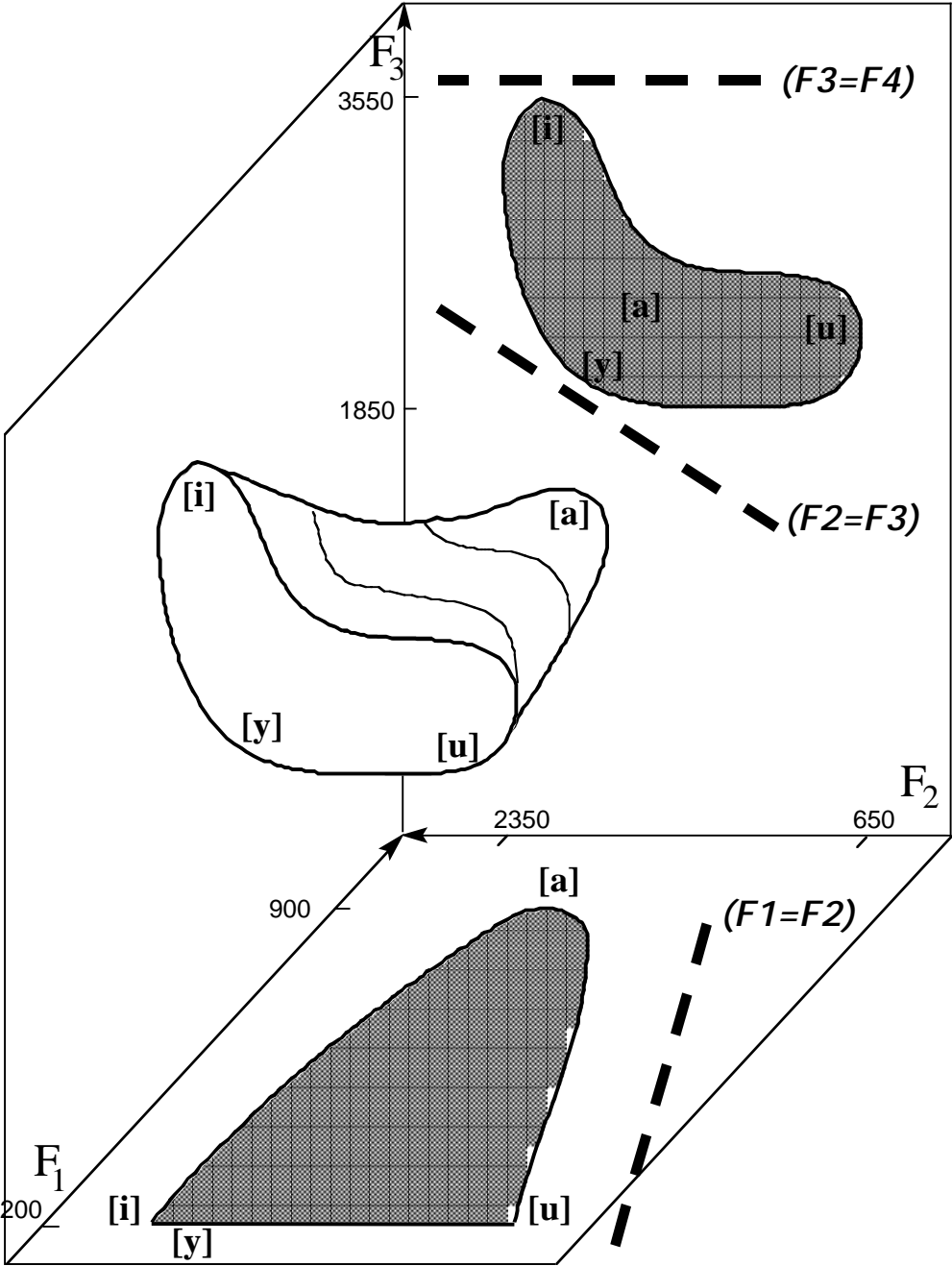


Figure 3 – The 3D maximal space of vowel spectra, and its (F1, F2) and (F2, F3) projections, together with the three focalisation planes (F1=F2), (F2=F3) and (F3=F4).

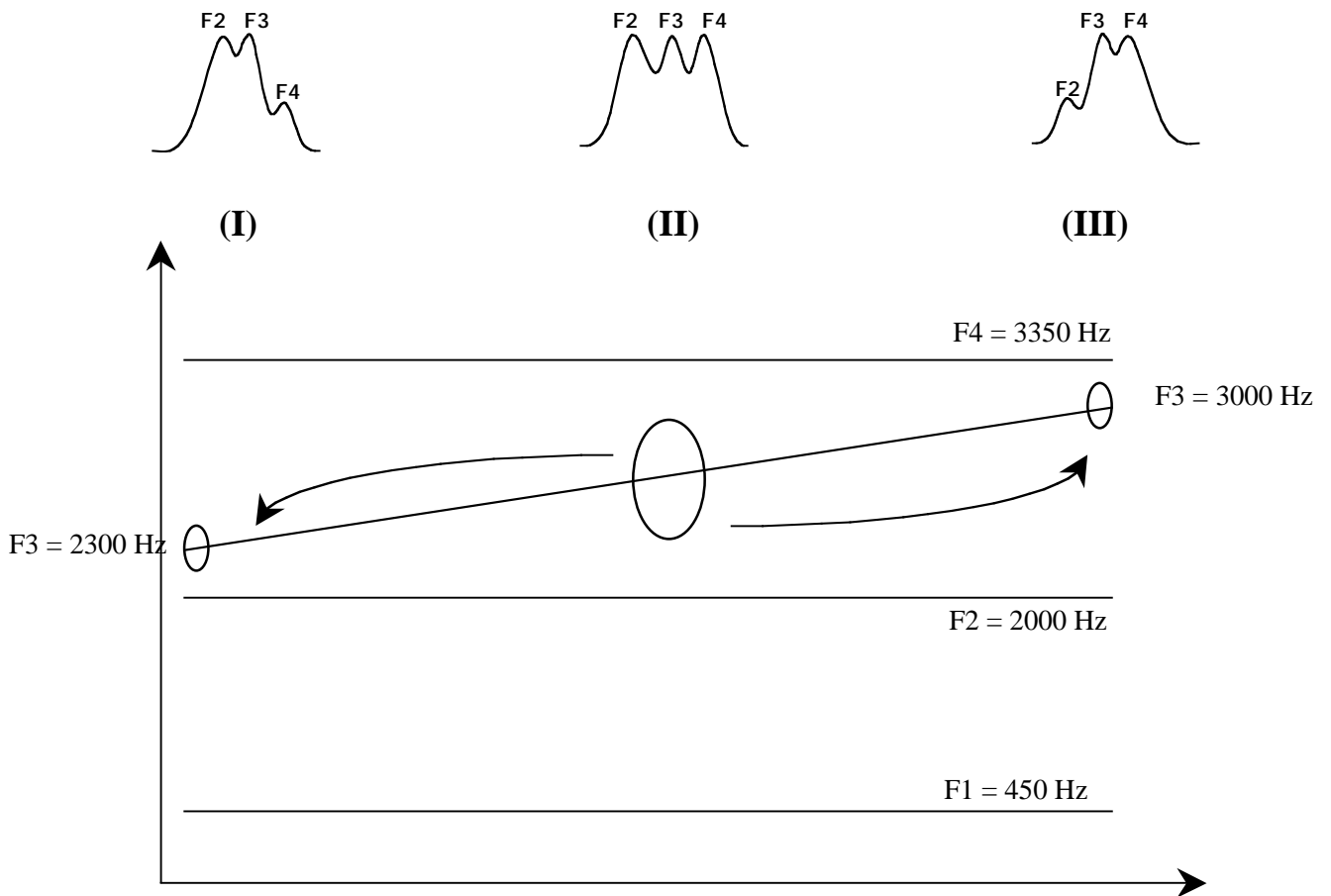


Figure 4 – A sketch of the main results in Schwartz & Escudier (1989). The studied continuum consisted of four-formant synthetic stimuli, with fixed F_1 , F_2 and F_4 , and varying F_3 . Discrimination experiments involving pairs of stimuli in regions (I) and (II) or (II) and (III) provide two main results: increase of false alarms for stimuli of type II (displayed by large vs. small ellipses), and asymmetries in discrimination with increased discrimination from (II) to (I) or (II) to (III) compared with inverse pairs (displayed by arrows). The (F_2 - F_3 - F_4) region of vowel spectra I, II and III is displayed in the upper part of the figure, showing the consequence of formant convergence in I and III.