On the invariance of speech percepts

Willy Serniclaes
CNRS & Université René Descartes, Paris 5

A fundamental question in the study of speech is about the invariance of the ultimate percepts, or features. The present paper gives an overview of the non-invariance problem and offers some hints towards a solution. Examination of various data on place and voicing perception suggests the following points. Features correspond to natural boundaries between sounds, which are included in the infant's predispositions for speech perception. Adult percepts arise from couplings and contextual interactions between features. Both couplings and interactions contribute to invariance. But this is at the expense of profound qualitative changes in perceptual boundaries implying that features are neither independently nor invariantly perceived. The question then is to understand the principles which guide feature couplings and interactions during perceptual development. The answer might reside in the fact that: (1) adult boundaries converge to a single point of the perceptual space, suggesting a context-free central reference; (2) this point corresponds to the neutral vocoïd, suggesting the reference is related to production; (3) at this point perceptual boundaries correspond to the natural ones, suggesting the reference is anchored in predispositions for feature perception. In sum, perceptual invariance seems to be grounded on a radial representation of the vocal tract around a singular point at which boundaries are context-fee, natural and coincide with the neutral vocoïd.

1. Introduction

You never bath twice into the same river (Heraclitus). While everything always changes what remains invariant? A fairly classical solution to the non-invariance problem is to look for constant relationship. According to Everitt (1998), invariance is "A property of a set of variables or a statistic that is left unchanged by a transformation" (p. 168). The purpose of this paper is to give some hints for handling the non-invariance problem in speech communication.

Features, the ultimate units of language (Jakobson, 1973), are the best candidates as building blocks for speech perception (Jakobson, Fant & Halle, 1952). Features were first defined on phonological grounds, as a function of their distinctive function in the language, hence "distinctive" features. They were later defined on articulatory grounds in the framework of Generative phonology; hence "phonetic" features (Chomsky & Halle, 1968). Though features are key concepts in empirical investigations, their perceptual invariance has been repeatedly questioned (Fromkin, 1979). How can we pretend that features are perceptually constant when there is massive evidence (Repp, 1982) to show that the *perception* of a given feature (e.g. stop place of articulation) depends on the phonetic context (e.g.: the following vowel, Schatz, 1953)? Simply by looking at contextual variations in feature *production.* Features are invariant to the extent that perceptual variations parallel those in production. Whenever this is true, the relationship between perception and production does not change across contextual transformations, conforming to the very definition of invariance.

Practically, invariance can be tested by comparing perceptual boundaries with productive categories, i.e. those present in speech production and which can be specified with acoustic measurements. With two different categories (e.g. /b/ and /p/) separated by a single feature (e.g. voicing), the perceptual boundary is the point along some acoustic continuum at which the categories are equally perceptible. Boundaries are usually measured by collecting labeling responses to stimuli generated by modifying an acoustic cue known to play a major role in the perception of the feature (e.g. for voicing: Voice Onset Time, VOT; Lisker & Abramson, 1964), and the boundary value corresponds to the point at which the two labeling responses are equi-probable (e.g. 50 % /b/ and /p/ labeling). Perceptual boundaries can then be matched with the distributions of the major cue in the production of the categories (Figure 1). Results on voicing perception (in English: Lisker & Abramson, 1976; in French: Serniclaes, 1987) show that both the perceptual boundary and the productive categories change with the context (e.g. voicing boundary and related productive categories change from /ba-pa/ to /gi-ki/ in Figure 1). However, as the relationship between voicing boundaries and categories remains fairly constant across contexts (as in Fig.1), feature perception can be considered to be nearly invariant. Studies on place of articulation also suggest parallel contextual shifts in perception and production (Dorman et al., 1977).

The fact that contextual variations do not grossly affect the relationship between perceptual boundaries and productive categories suggests that featural percepts are invariant. However, as we will see, this is at the expense of cross-dependencies in the perception of different phonetic features: the perception of a given feature (e.g. voicing) depends on other features (e.g. place or vowel), and vice-versa.

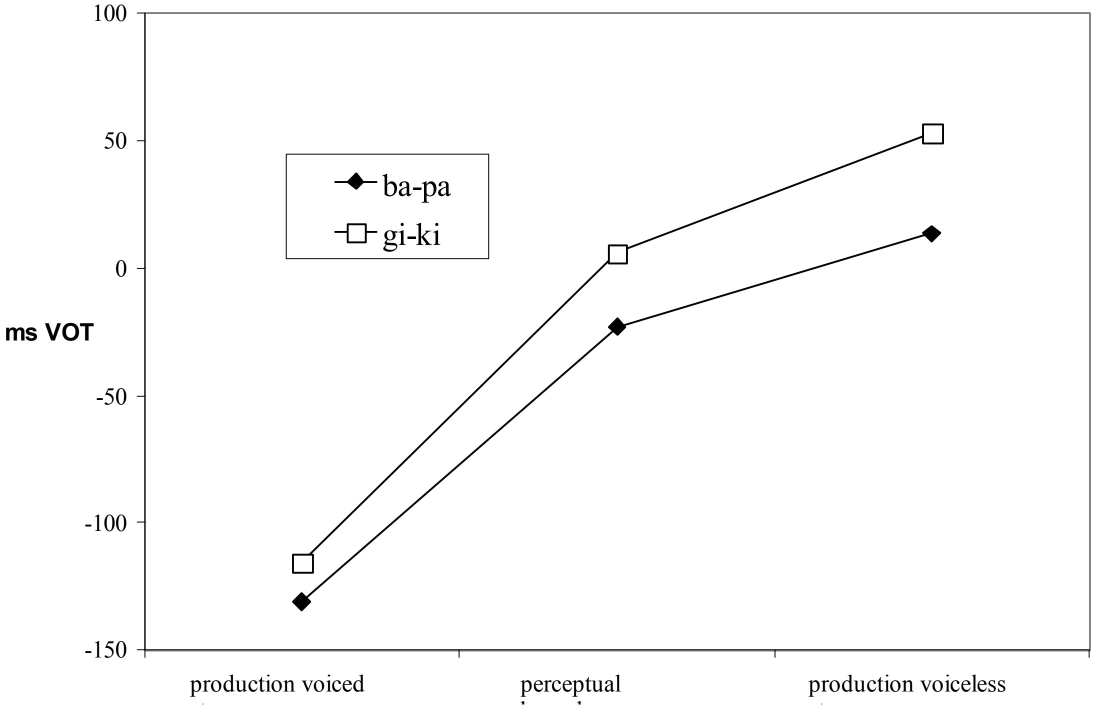

Figure 1. Relationship between voicing perception and production in French (adapted from Serniclaes, 1987). Mean acoustic measurements of VOT in voiced and voiceless stops as well as the mean perceptual boundaries along a synthetic VOT continuum are given in two phonetic contexts, i.e. /labial stop + a/ (/ba-pa/) and velar /stop + i/ (/gi-ki/). The contextual shift in perception (29 ms VOT) is about half-way between those in production (15 and 39 ms VOT for voiced and voiceless stops respectively; geometric mean = 25 ms). Perceptual boundaries follow the productive variations, resulting in a fairly stable relationship across contexts.

The present paper gives an overview of the non-invariance problem and offers some hints towards a solution. First, the empirical evidence for cross-dependencies in feature perception is reviewed. Then, data which suggest that adult percepts arise from couplings between perceptual predispositions for the perception of phonetic features will be presented. Perceptual couplings are combinations between phonetic features giving rise to language-specific features, hence "phonological" in nature. A further question will be to understand the nature of the representation which guides the development of feature couplings during language acquisition. At this point I will consider two basically different models of speech perception, the one based on auditory properties (Stevens, 1989), the other on motor ones (Liberman & Mattingly, 1985). I will argue that feature couplings are driven by a specific version of the speech-specific model, based on a radial representation of the vocal tract. Further, I will argue that this representation is based on a central reference corresponding to the neutral vocoïd (the "schwa") and that the distinction between language-specific and auditory-like processing disappears around that central reference. The latter is therefore not only central but also singular.

2. Perceptual dependencies between features

There are numerous examples to suggest that the perception of a given feature is affected by the phonetic context (for a review: see Repp, 1982). As a rule, contextual variations in feature production are paralleled by contextual adjustments in feature perception. Several models of contextual adjustment are possible. According to the "Auditory-acoustic" model, contextual effects in perception are due to simultaneous changes of acoustic cues which affect both the target feature and the contextual features. For example, the duration of formant transitions affects both the perception of voicing and place of articulation in stop consonants: longer transitions indicate both back vs. front place of articulation (/g-k/ vs. /b-p/) and voiced vs. voiceless category (/k-p/ vs. /g-b/). The inclusion of transition duration in the repertoire of voicing cues therefore contributes to the shift of the VOT boundary towards longer values (i.e. more voiceless) in a /b-p/ vs. /g-k/ context (Figure 1), transitions being longer (i.e. more voiced) in the latter. More generally, the multiple cueing of phonetic features might open the way for solving the non-invariance problem since the acoustic cues contributing to the perception of the same feature vary in a complementary way across contexts (Serniclaes, 1975; Dorman et al., 1977): when one cue is weaker (e.g. VOT is short in /p/, long in /k/), another is stronger (e.g. transitions are short in /p/, long in /k/). As the contextual variations of the cues compensate for each other, cue integration might give the key for solving the non-invariance problem.

While acoustic cue integration undoubtly contributes to perceptual invariance, this is not the whole story. According to the "Phonetic" model – to take back the classical Haskins' terminology (Carden et al., 1981) - contextual effects also truly arise from cross-dependencies in the perception of different features and, as we will see, this model is supported by the results of fairly sophisticated experiments. Two different "Phonetic" models are in turn possible. Perception of a given feature might simply bias the phonetic categorization of another feature. This is the "Additive" model. Alternatively, perception of a given feature might affect the processing of the acoustic cues involved in the perception of another feature. This is the "Interactive" model.

2.1. Auditory invariance: Locus model

The Locus model of place perception is undoubtedly the most elaborated form of Auditory-acoustic model of feature perception. According to this model, first settled by Delattre (Delattre, Liberman, & Cooper, 1955) the perceptual invariance of stop place of articulation in CV syllables is based on the virtual onset of F2 transition, extrapolated from its acoustic onset and offset. According to Delattre, the invariant for each place category is the frequency *value*, or Locus, towards which F2 transitions point in different vocalic contexts. As further research demonstrated that the Locus was not constant across vocalic contexts, the model has since been reformulated by Sussman (Sussman, McCaffrey, & Matthews, 1991; Sussman, Fruchter, Hilbert & Sirosh, 1998). Instead of a single value, now it is the *linear relationship* between the onset and offset of F2 transition which is supposed to be invariant for each place category (Equation 1).

Equation 1. $(F2)_{\text{onset}} = I + B_{\text{Place}}*(F2)_{\text{offset}}$

where Place $\in \{$labial, coronal, dorsal, velar$\}$
and $(F2)_{\text{onset}}$, $(F2)_{\text{offset}}$ correspond to acoustic measurements of the second formant in CV syllables
where I is the intercept

The invariants were originally formulated in terms of categories because they were primarily intended to be tested with production data, but they can be easily transposed into boundary invariants in order to cope with perceptual data (Equation 2).

Equation 2. $(F2)_{\text{onset}} = I + B_{\text{labial-coronal}}*(F2)_{\text{offset}}$

where $B_{\text{labial-coronal}}$ is a linear transform of B_{labial} and B_{coronal}
and $(F2)_{\text{onset}}$, $(F2)_{\text{offset}}$ correspond to the acoustic values of the second formant at the perceptual boundary.

This model is motivated by both ecological and phylogenetic considerations. According to Sussman et al. (1998): (1) linear relationships are quite common in the acoustic environments of species which are able to operate complex auditory processes; (2) vertebrates are endowed with pre-adapted mechanisms for processing linear relationships; (3) the human vocal system would result from an evolutive pressure leading to the production of stimuli which conform to these relationships. With this linear conception, the Locus is not fixed for each place but depends on the vocalic context. However, the invariant remains acoustic in nature because contextual adjustments operate through acoustic cue integration and do not depend on the perception of the adjacent vowel. According to the Locus equations, the percept does not depend on variations in the vocalic percept as long as the acoustic stimulus remains unchanged. This implies that fluctuations in vowel perception occurring with ambiguous stimuli should not affect consonant perception.

Among the widespread criticisms which have been addressed to the Linear model (cf. the comments to Sussman et al., 1998), the most important ones for our concern here are those related to the non-invariance problem. To sum up these criticisms, invariance has to rely on several different acoustic cues – not only F2 transition but also F3 transition and the burst- and the relative weightings of these cues should depend on the speaker and context (in the comments to Sussman et al., 1998: Carré p.262; Blumstein, p.260; Diehl, pp.264; Nearey, p.277). While this meshes neatly with the abundant evidence on cue multiplicity (Delattre, 1968) and contextual changes in the contribution of the different cues (such as those of formant transitions and burst: Dorman et al., 1977), the question is to know whether the contextual effects are indeed entirely acoustic in nature.

2.2. Perceptual dependencies between features

2.2.1. The phonetic vs. acoustic model

Although there is an acoustic component in contextual adjustments, the acoustic model cannot account for different data which suggest that identification of a given feature depends on the perceived identity of the surrounding features. These data show that in conditions where all the possible effects of acoustic cues were controlled, including those arising from random fluctuations in cue extraction with the same stimulus, contextual effects were still present and could then only arise from perceptual dependencies. Carden et al. (1981) demonstrated that place perception in consonants depended on whether exactly the same stimuli were presented either as stops or as fricatives. Similarly, using /Stop+ Vowel/ stimuli in which both voicing and place cues were fixed at ambiguous values, we showed that fluctuations in voicing categorisation depended on those in place categorization (Serniclaes & Wajskop, 1992). Further, the inclusion of vowel identification responses is necessary to account for consonant place identification as evidenced by the analysis of perceptual data with Logistic Regression models (Nearey, 1990).

2.2.2. The phonetic interactive vs. additive model

While these experiments suggest that the Auditory-acoustic model is too simple, different speech specific models are in turn possible. Perception of a given feature might simply bias the phonetic categorization of another feature. This is the “additive” model (Equation 3). Alternatively, perception of a given feature might affect the processing of the acoustic cues involved in the perception of another feature. This is the “interactive” model (Equation 4).

Equation 3. $(F2, F3)_{\text{onset}} = I + \text{Vowel} + B_{\text{labial-coronal}}{*}(F2, F3)_{\text{offset}}$

Equation 4. $(F2, F3)_{\text{onset}} =$
$I + \text{Vowel} + B_{(\text{labial-coronal})}{*}(F2, F3)_{\text{offset}}{*}\text{Vowel}$

where ‘Vowel’ represent the perceived identity of the vowel.

Examination of previous data on the perception of English synthetic /si, ʃi, su, ʃu/ syllables by Nearey (1990) led to the conclusion that effects of vowels on consonant identification were additive. Logistic Regression functions were used by Nearey for testing the additive vs. interactive perceptual models. Vowel and consonant bias terms were significant but interactive terms were not significant, which supported the additive model. However, in a more recent study on Dutch fricative-vowel syllables, Smits (2001a) found evidence supporting perceptual interactions using a Hierarchical Categorization model (HICAT: Smits, 2001b). HICAT allows to separate tests of the effects of vowel on consonant perception from those of consonant on vowel perception, a distinction which was not addressed in Nearey’s work.

2.2.3. A specific phonetic interactive model: the Radial Model

We provided a further test of the perceptual dependencies between features in an experiment on the perception of synthetic /fricative+vowel/ syllables generated by factorial modification of formant transitions onset-offset, with F2 and F3 covarying (Serniclaes & Carré, 2002). The data also supported an interactive model of phoneme perception and further showed that the additive component was not necessary (Equation 5).

Equation 5. $(F2, F3)_{\text{onset}} = I + B_{(\text{labial-coronal})}{*}(F2, F3)_{\text{offset}}{*}\text{Vowel}$

Geometrically, the absence of an additive component means that the boundaries converge to a single point in the space of formant transitions onset-offset, (Figure 1). This means there is a point in the perceptual space at which place perception is context free. Interestingly, the convergence point corresponds to a stimulus with flat F2-F3 formant transitions with values corresponding to the neutral vocoïd (1500 Hz F2-2500 Hz F3), corresponding to the uniform vocal tract. Further, flat transitions constitute a natural auditory boundary between rising transitions and falling transitions (Cutting & Rosner, 1974). It thus seems that place perception is organized around a central reference characterized by both natural and context free boundaries, and corresponding to the neutral productive category. With the vocal tract in a fairly neutral position, place perception does not strongly depend on the perception of the vocalic context and is derived from natural auditory sensitivities. However, outside the neutral context, the interaction between place and vowel perception generates speech specific boundaries which become increasingly different with the distance from the neutral vocoïd measured on directions which depend on the perceived identity of the vowel. This suggests that place perception is based on a "radial" representation anchored on the neutral vocoïd. This representation is suggested by the fact that perceptual boundary for place of articulation executes a radial movement from the front vowel contexts (on the right-hand in Figure 2) to back vowel contexts (on the left-hand in Figure 2).

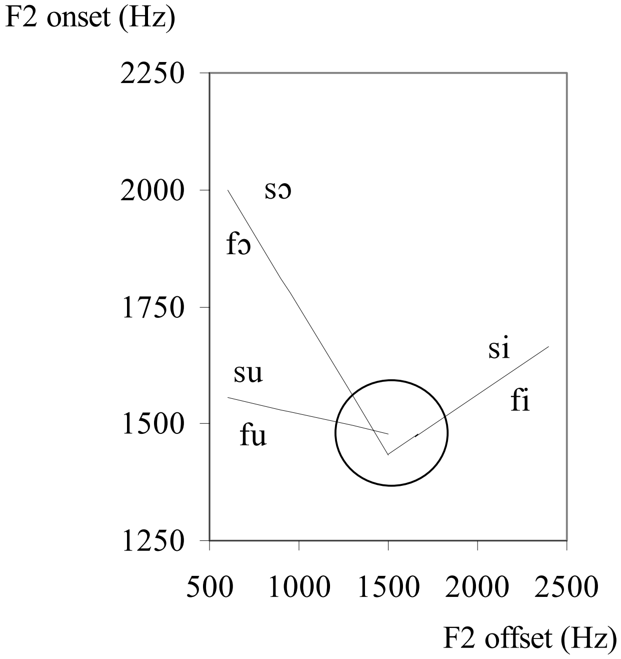

Figure 2. (adapted from Serniclaes et al., 2002): Perceptual boundaries in the F2 onset-offest plane. F3 onset-offset values covaried with those of F2 in this experiment and F3 was close to 2500 HZ for F2 of 1500 Hz, which corresponds to the neutral vocoïd values. In agreement with the radial model, the obtained boundary lines converge to the F2 flat transition when the offset value is close to 1500 Hz. For stimuli with offset values close to 1500 Hz F2 (circled region), place perception is fairly independent of the vocalic context.

3. Couplings between perceptual predispositions for speech

3.1 Models of speech development

Human infants are born with predispositions for perceiving all the possible phonetic contrasts, which are then activated or not as a function of the presence versus absence of the corresponding contrast in the linguistic environment. This fairly classical view on speech development (Werker & Tees, 1999) is grounded on a considerable amount of empirical evidence. Neonates can already discriminate between a range of phonetic categories (Eimas, Siqueland, Jusczyk & Vigorito, 1971), even between those which are not present in their ambient language (e.g. Lasky, Syrdal-Lasky & Klein, 1975). The initial ability to discriminate the universal set of phonetic contrasts however declines within the first year of life (Werker & Tees, 1984a) but the decline involves a change in processing strategies rather than a sensorineural loss (Werker & Tees, 1984b).

Infant studies not only show that the discrimination between phonetic categories is already present at birth, they also indicate that the location of phonetic boundaries already depends on several acoustic cues. Thus, the discrimination between voiced and voiceless stops by infants below six months of age depends both on voice onset time (VOT) and F1 transition duration, just as for adult speakers of English (Miller & Eimas, 1983). Innate mechanisms might thus also explain the integration of multiple cues for the perception of the same phonetic feature.

Perceptual development would be fairly simple if it was restricted to selecting the percepts in a stock of innate predispositions, as in Werker's model. Phillips (2001) calls this a ”structure-adding” approach, all features being processed at a universal “phonetic” level processing and only those specific to the language at an upper-stage “phonological” level. Alternatively, the adult perceptual space might not be straightforwardly related to the universal predispositions (Kuhl, 1994; 2000), what Phillips considers as a “structure-changing” approach. A third possibility is that language specific features are generated by couplings between phonetic features (Serniclaes, 1987; 2000), which implies both structure-adding and structure-changing.

Couplings are combinations between features. Couplings create new functional entities inside which features are integrated. The term “coupling” is commonplace in the study of visual perception, e.g. for describing perceptuo-motor integration in depth perception (Hochberg, 1981).

3.2 Test of a mixed model: coupling between predispositions

In support of the coupling model, previous research already suggested that voicing perception in several languages is based on a VOT boundary which is not precluded in the infant's predispositions. Up to about 6 months of age, infants discriminate three voicing categories, separated by two VOT boundaries (see Figure 2; Lasky, Syrdal-Lasky, & Klein, 1975; Aslin, Pisoni, Hennessy, & Perrey, 1981). After 6 months of age, only the positive VOT boundary remains active in languages with a single distinction between short vs. long positive VOT categories (e.g. English; Figure 3; Eilers, Wilson & Moore, 1979). Languages such as Spanish and French use a single distinction between negative VOT and moderately long positive VOT categories (Caramazza & Yeni-Komshian, 1974; Williams, 1977), and the perceptual boundary is located around 0 ms (Serniclaes, 1987). The fact that the boundary is located around 0 ms means that negative and positive VOT are equally important for voicing identification and hence that the categorical predispositions for the perception of negative and positive VOT are both activated and coupled in the course of perceptual development. It might be argued that the 0 ms VOT boundary simply emerges in the course of development, while the positive and negative boundaries are deactivated. While this is of course possible, the inclusion of predispositions combinations in the predispositions would seriously entail the parsimony of the model.

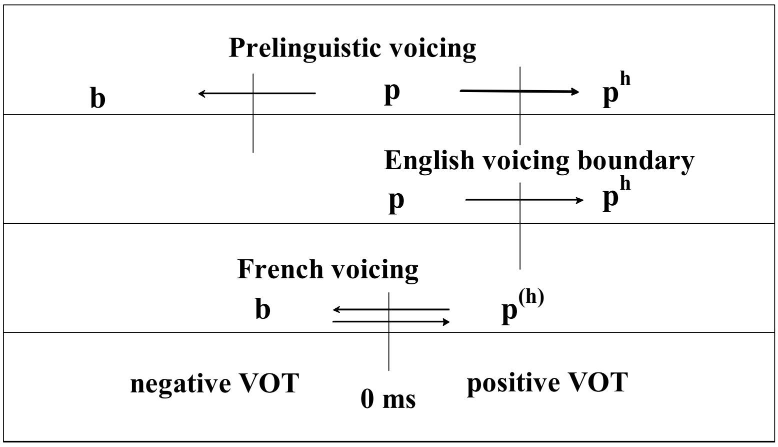

Figure 3 (from Serniclaes et al., 2004): Perceptual boundaries between voicing categories in prelinguistic children, in English and in French for stops in syllable-initial position . Prelinguistic boundaries correspond to predispositions (indicated by arrows) for the perception of all potential categories (voiced as for /b/, voiceless as for /p/ and voiceless aspirated as for /p^h/) in the world's languages. In English, a single predisposition is activated for performing the distinction between voiceless unaspirated and voiceless aspirated stops. In French, two predispositions are coupled in order to perform the distinction between voiced and slightly aspirated voiceless stops.

In support of the coupling hypothesis, examination of studies on children raised in Spanish-speaking environments showed that the 0 ms VOT boundary is not predicted by the infant's predispositions (Lasky et al., 1975), although it appears fairly early in the course of language development (Eilers et al., 1979). Recently, data collected on children raised in French-speaking environments suggested those around 4 months of age discriminated the negative and positive boundaries (located at -30 and +30 ms VOT respectively) whereas those around 8 months of age discriminated the 0 ms VOT boundary (Hoonhorst, 2004). Further evidence on couplings between predispositions has been obtained for the perception of place of articulation. F2 and F3 transitions allow separating the three place categories usually found in languages, i.e. labial, coronal and velar. In the neutral vocalic context, stimuli with raising F2-F3 transitions correspond to /b/ percepts, those with falling F2-F3 transitions correspond to /d/ percepts and those with falling F2 and rising F3 transitions to /g/ percepts (Carré, Liénard, Marsico & Serniclaes, 2002). However, a fourth category characterized by raising F2 and falling F3 transitions is also possible and it might correspond to the palatal consonants found in Czech (Jakobson et al., 1952) and also in Hungarian (Geng et al., 2005). As the perception of rising vs. falling transitions is grounded on natural boundaries –flat transitions, see above- the discrimination of F2 and F3 transitions is probably present in the newborn, although there is no direct evidence on this point. The predispositions for perceiving F2 and F3 transitions might straightforwardly be used in four-category languages, two binary features allowing to discriminate four place categories.

However, the natural F2 and F3 boundaries are not optimal for perceiving consonants in three-category languages. The F2-F3 perceptual space should be divided into three equally sized regions for optimal use, which would require new boundaries (Figure 4). These boundaries can only be obtained by trade-off between F2 and F2 transitions, e.g. a strongly falling F3 compensating for a slightly raising F2 for perceiving /d/ instead of /b/. Notice that if F2 and F3 transitions are not simply two different acoustic cues but are instead precluded into different perceptual predispositions, the very existence of a perceptual trade-off between F2 and F3 transitions means coupling between predispositions.

We have recently found evidence in support of this conjecture by collecting both identification and discrimination responses to /stop + neutral vocoïd/ synthetic syllables generated by either factorial or combined modification of F2 and F3 transition onsets. Preliminary results (Serniclaes, Bogliotti & Carré, 2003; see Figure 4) showed that French adult speakers discriminated natural F2 and F3 boundaries -i.e. those corresponding to flat transitions- though their labelling boundaries reflected trade-offs between F2 and F3. The fact that perceptual boundaries for place of articulation are built on trade-offs between of the coupling hypothesis. These results have since been confirmed with a larger sample of subjects (Bogliotti, 2005) two acoustic cues, which are each endowed with natural boundaries, provides further support.

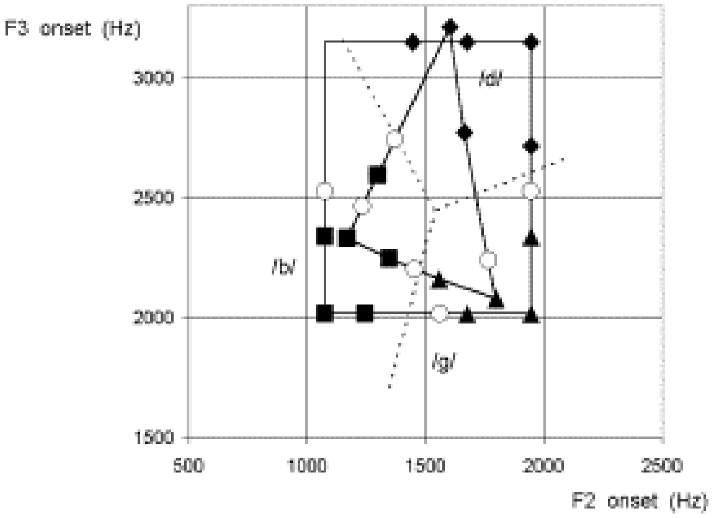

Figure 4. (from Serniclaes et al., 2003): Labeling and discrimination results for two places of articulation continua. Stimuli collecting at least 75% /b/, /d/ or /g/ responses are indicated by squares, diamonds and triangles respectively. Broken lines represent category boundaries. White circles indicate above chance discrimination peaks.

3.3 Speech perception in dyslexic children: a coupling deficit?

Phonological couplings between features imply considerable qualitative changes. It would therefore not be surprising to find coupling deficits in some part of the population. Our research on speech perception dyslexia lends support to the existence of coupling deficits and this paved the way to a new explanation of dyslexia. Our work in this domain is in the framework of the phonological explanation of reading deficits (for an overview see: Sprenger-Charolles, 2003). Previous investigations showed that dyslexics are affected by deficient grapheme-phoneme correspondences, deficits in phonological awareness, phonological short-term memory, phoneme discrimination and in categorical perception. But where is the core deficit? A first investigation showed that the categorical perception deficit in dyslexia is characterized by a better discrimination of within-category differences (Serniclaes, Sprenger-Charolles, Carré & Démonet, 2001). This result suggested a new hypothesis as to the origin of dyslexia, namely that it comes from a deficit in the coupling between predispositions in the course of perceptual development which gives rise to an allophonic, rather than phonemic, mode of speech perception. Allophonic perception offers a possible explanation to dyslexia. The child who perceives speech in allophones has an evident handicap for discovering the relationship between the speech sounds and alphabetic symbols, knowing that the opacity of the writing system adds a further difficulty, of cultural order. Allophonic perception has several testable consequences as to the difference between dyslexics and controls. The main one is that dyslexics should be less categorical than controls for phonemic distinctions and should be more categorical for the allophonic ones. This prediction was recently confirmed by the results of several different investigations (Bogliotti, 2003; Serniclaes, Van Heghe, Mousty, Carré & Sprenger-Charolles, 2004; Burnham, 2003).

4. Discussion and conclusions

To summarize the evidence contemplated in the previous sections, two basic findings emerge. Firstly, while phonetic features were initially conceived as autonomous dimensions of speech, it now appears that they are not independently perceived. Secondly, while phonetic features are language-independent dimensions of speech, they are not always directly used for speech perception in a given language. Rather, speech perception is also based on language-specific couplings between phonetic features.

Phonetic features were conceived as language-independent autonomous dimensions of speech production. The discovery of infant's predispositions for feature discrimination and the traces they leave in the adults make it clear that features remain the best candidates as building blocks of speech perception. However, couplings between predispositions show that phonetic features do not constitute independent units for language perception in the adult. If features are independent units at the start, why are they interactively processed? Presumably because when features are put into action in a linguistic frame, they do not have invariant acoustic correlates. Couplings between features constitute an obvious remedy to contextual effects in production. For example, voicing contrasts are easier to produce in labial stops before open vowels (e.g. /ba/-/pa/), while aspiration contrasts are easier to produce in velars stops before closed vowels (e.g. /gi/-/ki/). Coupling voicing and aspiration then gives rise to a more stable and possibly invariant compound. However, the fact that voicing and place perception are not independent indicates that couplings are not sufficient for attaining invariance and that contextual interactions further contribute to it.

An important question is to understand the principles which guide the development of feature couplings and interactions during language acquisition. How is the search for invariance implemented in development? Invariance requires that perceptual boundaries fit into productive categories. There are basically two different ways by which feature compounds might be invariant: invariance might either be driven by motor representations in perception or by auditory representations in production. Speech perception theories can be subdivided into four classes, depending on whether invariance is conceived with or without major contribution from learning and whether it is based on auditory or speech specific representations (Serniclaes, 2000). Both the Quantal Theory (Stevens, 1989), and the one based on "natural" psychoacoustic boundaries (Kuhl & Padden, 1983) are basically inneist as they consider that invariance is the by-product of auditory integration and that learning only plays a marginal role. While there are predispositions for feature perception, we have seen that acquisition plays a crucial role in speech perception. No wonder then if other auditory theories are centered on acquisition. Among them the 'Perceptual Magnet' theory (Kuhl, 1994; 2000) considers that adult percepts are shaped by linguistic experience. Though the perceptual magnets are not clearly related to innate dimensions, they might easily be accommodated with couplings between predispositions. However, while magnets are quite interesting concepts for understanding the genesis of linguistic categories, they should be conceived in speech specific rather than auditory terms.

Motor theories suppose that direct links exist between perception and motor commands (Liberman & Mattingly, 1985), a contention which received recent support by the existence of mirror neurons (Fadiga, Fogassi, Paresi & Rizzolatti, 1995; Rizzolatti, Fadiga, Gallese & Fogassi, 1996; Studdert-Kennedy, in press). In further support of this conception, fRMI results show that, with exactly the same acoustical stimuli, a change in perceptual mode from nonspeech to speech affects the localization of the brain activity (Dehaene-Lambertz, Pallier, Serniclaes, Sprenger-Charolles, Jobert & Dehaene, 2005). These results strongly suggest that speech perception is achieved through specific pathways different from those used in auditory perception because the change in the neural site of processing in the brain was obtained with exactly the same stimuli, thereby excluding possible confounding effects arising from differences in stimulus complexity.

While there is recent strong neuro-imagery evidence in support to the Motor theory, the latter is basically inneist, a view which is difficult to conciliate with couplings between predispositions. Articulatory theories, notably the “direct-realist” one (Studdert-Kennedy, 1985; Fowler, 1986), rely on the learning of invariants from environmental regularities and are therefore better suited for explaining the complexities of perceptual development.

While it seems fairly clear that speech perception is related to articulatory representations, the precise nature of these representations remains unknown. However, some hints might be found in our results on place of articulation perception (see above, Serniclaes et al., 2002; 2003). Place perception seems to be built up around a singularity of the perceptual space characterized by boundaries which are both context-free and natural. Further, this singularity coincides with the neutral vocoïd, which corresponds to the uniform vocal tract. This provides a straightforward link between perception and production (Carré, Liénard, Marsico & Serniclaes, 2002). As contextual adjustments in perception correspond to radial movements of boundary lines around the neutral point, it would seem that perception occurs in a spatial representation of the vocal tract with radial lines as contextual variants of a central, and context-free, reference.

This "radial" model of speech perception needs to be refined and tested with appropriate means. But it can already find some support by the fact that the only neural site with is specifically dedicated to the categorization of speech features is located in the left supra-marginal gyrus (Dehaene-Lambertz et al., 2005), a region which is linked to the sensory representation of the mouth and might correspond to part of the auditory cortex devoted to the processing of spatial information.

Acknowlegments

I would like to thank Bernd Pompino-Marschall and Noel Nguyen for his insightful comments on a previous version of this manuscript.

References

Aslin, R.N., Pisoni, D.B., Hennessy, B.L., & Perrey, A.V. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development, 52,* 1135-1145.

Bogliotti, C. (2003). Relation between categorical perception of speech and reading acquisition. In M.J.Solé, D.Recaesens & J.Romero (Eds.). *Proc. 15th International Congress on Phonetic Sciences*, 885-888.

Bogliotti, C. (2005). Perception categorielle et perception allophonique: Incidences de l'âge, du niveau de lecture, et des couplages entre predispositions phonétiques. *PhD. Thesis, Université Paris 7 - Denis Diderot.*

Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading and Writing, 16,* 573-609.

Caramazza,A. and Yeni-Komshian,G.H. (1974). Voice onset time in two French dialects. *Journal of Phonetics, 2,* 239-245.

Carden, G., Levitt, A., Jusczyk, P.W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception and Psychophysics, 29,* 26-36.

Carré, R., Liénard, S., Marsico, E., & Serniclaes, W. (2002). On the role of the "schwa" in the perception of plosive consonants. In J.L.H. Hansen & B. Pellom (Eds.). *7th International Conference on Spoken Language Processing*, 1681-1684.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York Harper and Row.

Cutting, J.E., & Rosner, B.S. (1974). Categories and boundaries in speech and in music. *Perception and Psychophysics*, *16,* 564-570.

Dehaene-Lambertz, G., Pallier, Chr., Serniclaes, W., Sprenger-Charolles, L., Jobert, & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24,* 21-33.

Delattre, P.C., Liberman, A.M., & Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America, 27,* 769-773.

Delattre P. (1968). "From acoustic cues to distinctive features" *Phonetica* 18, 198-230.

Dorman, M.F., Studdert-Kennedy, M., & Raphaël, L.S. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics, 22,* 109-122.

Eilers, R., Wilson, W. and Moore, J. (1979). Speech discrimination in the language-innocent and the language-wise: a study in the perception of voice onset time. *Journal of Child Language, 6,* 1-18.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Everitt, B.S. (1998). *The Cambridge dictionary of statistics*. Cambridge University Press.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608-2611.

Fowler, C.A. (1986). An event approach to the study of speech perception. *Journal of Phonetics, 14,* 2-28.

Fromkin, V.A. (1979). Persistent questions concerning distinctive features. In B.Lindblom and S.Öhman (Eds.) *Frontiers of Speech Communication Research.* London: Academic Press. 323-334.

Geng, C., Mády, K., Bogliotti, C., Messaoud-Galusi, S., Medina,V. & Serniclaes, W. (2005). Do palatal consonants correspond to the fourth category in the perceptual F2-F3 space? *ISCA Workshop on Plasticity in Speech Perception, London, June 15-17 2005:* 219-222.

Hochberg J. (1981). On cognition in perception: perceptual coupling and unconscious inference. *Cognition, 10,* 127-34.

Hoonhorst (2004). L'évolution de la discrimination phonologique des jeunes enfants entre 4 et 8 mois et ses implications sur la compréhension de l'étiologie de la dyslexie. Mémoire de Licence Spéciale en Logopédie, ULB-UCL.

Jakobson, R. (1973). *Essais de Linguistique Générale*. Paris: Editions de Minuit.

Jakobson, R., Fant, G. and Halle, M. (1952). *Preliminaries to speech analysis. The distinctive features and their correlates*. Cambridge Mass.: M.I.T. Press.

Kuhl, P.K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4,* 812-822.

Kuhl, P.K. (2000). Language, mind, and brain: Experience alters perception. In M. Gazzaniga (Ed.), *The cognitive neurosciences 2nd ed.* (pp. 99-115). Cambridge, MA: MIT Press.

Kuhl, P.K., & Padden, D.M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003-1010.

Lasky, R.E., Syrdal-Lasky, A., & Klein, R.E. (1975). VOT discrimination by four to six and a half months old infants from Spanish environments. *Journal of Experimental Child Psychology, 20,* 215-225.

Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Miller, J.L. and Eimas, P.D. (1983). Studies on the categorisation of speech by infants. *Cognition, 13,* 135-165.

Nearey, T.M. (1990). The segment as a unit of speech perception. *Journal of Phonetics 18*, 347-373.

Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science, 25,* 711-731.

Repp, B.H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81-110.

Rizzolatti, G., Fadiga, L., Gallese,V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131-141.

Serniclaes, W. (1975). Perceptual processing of acoustic correlates of the voicing feature. In G.Fant ed. *Speech Communication, Proc. of the SCS, Stockholm 1974* (pp. 87-94). New York: J.Wiley.

Serniclaes, W. (1987). *Etude expérimentale de la perception du trait de voisement des occlusives du français.* Unpublished doctoral thesis. Université Libre de Bruxelles. Téléchargeable à http://www.vjf.cnrs.fr/umr8606

Serniclaes, W. (2000). La perception de la parole. In *La parole, des modèles cognitifs aux machines communicantes*. P. Escudier, G. Feng, P. Perrier, J.-L. Schwartz, Eds.; Paris: Hermès, 159-190.

Serniclaes, W., Bogliotti, C. & Carré, R. (2003). Perception of consonant place of articulation: phonological categories meet natural boundaries. In M.J. Solé, D. Recaesens & J. Romero (Eds.). *Proc. 15th International Congress on Phonetic Sciences*, 391-394.

Serniclaes, W. & Carré, R. (2002). Contextual effects in the perception of place of articulation: a rotational hypothesis. In J.L.H. Hansen & B. Pellom (Eds.). *7th International Conference on Spoken language Processing*, 1673-1676.

Serniclaes, W. & Sprenger-Charolles, L. (2003). Categorical perception of speech sounds and dyslexia. *Current Psychology Letters: Behaviour, Brain & Cognition, 10,* http://cpl.revues.org/documents379.html

Serniclaes , W., Sprenger-Charolles , L., Carré, R., & Démonet, J.-F. (2001). Perceptual discrimination of speech sounds in dyslexics. *Journal of Speech Language and Hearing Research, 44,* 384- 399.

Serniclaes, W., Van Heghe, S., Mousty, Ph., Carré, R. & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology, 87,* 336-361.

Serniclaes, W., & Wajskop, M. (1992). Phonetic versus acoustic account of feature interaction in speech perception. in *Analytic Approaches to Human Cognition, Proc. of the Conference in Honour of Paul Bertelson*, Brussels June 1991. J. Alegria, D. Holender, J. Junça de Morais, and M. Radeau eds.(pp.77-91). Amsterdam: North-Holland.

Smits, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1145-1162.

Smits, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics, 63,* 1109-1139.

Sprenger-Charolles, L. (2003). Linguistic processes in reading and spelling: The case of alphabetic writing systems: English, French, German and Spanish. In T. Nunes and P. Bryant (Eds.). *Handbook of children's literacy (pp.43-65).* Dordrecht: Kluwer Academic Publishers.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3-45.

Studdert-Kennedy, M. (1985). Perceiving phonetic events. In W.H. Warren, Jr. and R.E. Shaw (Eds.), *Persistence and Change : Proceedings of the First International Conference on Event Perception* (pp. 139-156). Hillsdale, NJ : Erlbaum.

Studdert-Kennedy, M. (in press). How did language go discrete? In *Evolutionary Prerequisites of Language* (provisional title). M. Tallerman ed. Oxford: Oxford University Press.

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

Sussman, H.M., MCCaffrey, H.A., & Matthews, S.A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, *90,* 1309-1325.

Werker, J.F., & Tees, R.C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Werker, J.F., & Tees, R.C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Werker, J.F. & Tees, R.C. (1999). Influences on infant speech processing: Towards a new synthesis. *Annual Review of Psychology, 50,* 509-535.

Williams, L. (1977).The voicing contrast in Spanish. *Journal of Phonetics, 5,* 169-184.