Journal of Translation and Languages مجلة الترجمة واللغات
ISSN : 1112-3974

Invariance of speech percepts

Sebane Zoubida
University of Oran - Algeria
zsebane@yahoo.fr

Abstract: *A fundamental question is about the invariance of the ultimate, percepts or features. The present paper gives an overall picture of the issues relevant to this question, aiming at finding a solution. A discussion of the main elements contributing to invariance will be made explicit, namely coupling and interaction. The main question will be that of the principles guiding coupling and interaction during perceptual development. The conclusion of this research work is that perceptual invariance appears to be based on a radial representation of the vocal tract around a singular point at which boundaries are context-free, natural and coincide with the neutral vocoid.*
Keywords: *Invariance, models of perception, coupling, predispositions.*
Résumé : *Une question fondamentale porte sur l'invariance de l'ultime, des percepts ou des caractéristiques. Le présent article donne une vue d'ensemble des problèmes liés à cette question, dans le but de trouver une solution. Une discussion des principaux éléments contribuant à l'invariance sera rendue explicite, à savoir le couplage et l'interaction. La question principale sera celle des principes guidant le couplage et l'interaction au cours du développement perceptif. La conclusion de ce travail de recherche est que l'invariance perceptive semble reposer sur une représentation radiale du conduit vocal autour d'un point singulier où les frontières sont hors contexte, naturelles et coïncident avec la vocoïde neutre.*
Mots clés *: Invariance, modèles de perception, couplage, prédispositions.*

1. Introduction

In the study of speech acquisition, a fundamental question is about the invariance of the ultimate, percepts or features. Examination of various data on place and voicing 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 coupling and contextual interaction between features. Both coupling and interaction contribute to invariance. However, this occurs at the expense of profound qualitative changes in perceptual boundaries implying that features are neither independently nor invariantly perceived. The question is then to understand the principles which guide coupling and interaction during perceptual development. The answer might be 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 vocoid, 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 short, perceptual invariance seems to be grounded on a radial representation of the vocal tract around a singular point at which boundaries are context-free, natural and coincide with the neutral vocoid.

2. Definition of Invariance

What remains invariant in speech, when everything always changes? A fairly classical solution to the non-invariance problem is to look for constant those relationships. According to Everitt (1998), invariance is a property of a set of variables or a statistic that is left unchanged by a transformation.

Jacobson (1973) defines features as being the ultimate units of language. They are the best candidates as building blocks for speech perception (Jacobson, Fant&Halle, 1952). Features were first defined on phonological grounds, as a function of their distinctive feature in the language, hence 'distinctive' features. They were later defined on their 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 depends on the phonetic context, 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.

In a practical way, 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, for instance /b/ and /p/ separated by a single feature (e.g. voicing), the perceptual boundary is the point along some acoustic continuum at which categories are equally perceptible. Boundaries are usually measured by collecting labelling 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& Abrason,1967), and the boundary value corresponds to the point at which the two labelling responses are equi-probable.

Perceptual boundaries can then be matched with the distributions of the major cue in the production of the categories (fig 1). Results on voicing perception (in English: LiskeraAbramson,1976; in French: Sernicleas, 1987) show that both the perceptual boundary and the productive categories change from /ba-pa/ to /gi-ki/ I, fig 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 (Dormanet 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 it will be shown, 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.

This work 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 consider two basically different models of speech perception, the one based on auditory properties (Stevens,1989), the other on motor ones (Liberman&Mattingly,1985).

It will be argued that feature couplings are driven by a specific version of the speech-specific model, based on a radial representation of the vocal tract. Further, it will be argued that this representation is based on a central reference corresponding to the neutral vocoid (the schwa) and that the distinction between language-specific and auditory-like processing disappears around that central reference. The latter is not only central but also singular

3. Perceptual dependencies between features

There are numerous examples to suggest that the perception of a given feature is affected by the phonetic context. 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 that affect both the target feature as well as the contextual features. For instance, the duration of formant transitions affects both the perception of voicing and the 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 (fig 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 (Seniclaes, 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 kek for solving the non-invariance problem.

Acoustic cue integration undoubtedly contributes to perceptual invariance. According to the 'phonetic' model (Carden et al.; 1981), contextual effects also truly arise from cross-dependencies in the perception of different features. Perception of a given might simply bias the phonetic categorization of another feature. 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.

4. Auditory Invariance: The Locus Model

The Locus Model of place of articulation is undoubtedly the most elaborated form of Auditory-acoustic model of feature perception. According to this model, developed by Delattre (Delattre, Liberman, &Cooper, 1995) 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 invariance of 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 (Equation1).

Equation 1 (F2) onset = I = B place (F2) offset

Where place is [labial, coronal, dorsal, veler]
And (F2) onset, (F2) 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 onto boundary invariants in order to cope with perceptual data (equation 2).

Equation 2 (F2) onset = I + B labial –coronal (F2) offset where B labial coronal is a linear transform of B labial and B coronal and (F2) onset, (F2) 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):

- Linear relationships are quite common in the acoustic environments of species which are able to operate complex auditory processes.
- Vertebrates are endowed with pre-adapted mechanisms for processing linear relationships.
- The human vocal system would result from an evolutive pressure leading to the production of stimuli which conform to these relationships.

With these linear conception, the Locus is not fixed for each place but depends on the vocalic contexts. However, the invariant remains acoustic in nature because conceptual adjustments operate through acoustic the 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.

5. Perceptual dependencies between features

5.1. 1The phonetic vs the 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 stimuli, 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 voice and place cues were fixed at ambiguous values, showed that fluctuations in voicing categorization depended on those in place categorisation. 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).

5.2. 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 categorisation of another feature. This is the 'additive' model (equation3). Alternatively, perception pf 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) onset = I vowel + B labial-coronal *(F2, F3) offset
Equation 4 (F2, F3) onset = I vowel + B labial-coronal * (F2, F3) offset* vowel, where vowel represents the perceived identity of the vowel.

Examination of previous data on the perception of English synthetic /si , Si , su, Su / syllables by Nearey (1990) led to the conclusion that effects of vowel on consonant identification were additive. Logistic Regression functions were used by Nearey for testing the additive vs the interactive perceptual model. However, in a more recent study, on Dutch fricative-vowel syllables, Smits (2001a) found evidence supporting perceptual interactions using a Hierarchical Categorisation model (HICAT°: Smits,2001b).

HITAC 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.

5.3. A specific phonetic interactive model: The Radial Model

A further test is provided of the perceptual dependencies between features in an experiment on the perception of synthetic / fricative+vowel/ syllables generated by fractional modification of formant transitions onset-offset, which 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).

Equation5 (F2, F3) onset= I= B (labial-coronal) *(F2, F3), offset*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, (fig 1). This means that there is a point in the perceptual space at which place perception is context free. Interestingly, the convergence point corresponds stimulus with a flat F2-F3 formant transitions with values corresponding to the neutral vocoid (1500 Hz F2-2500Hz F3), corresponds to the uniform vocal tract. Further, flat transitions and falling 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, place perception does not strongly depend on the perception of the vocalic context, the interaction between place and vowel perception generates speech specific boundaries which become significantly different with the distance from the neutral vocoid measured on directions which depend on the perceived identity of the vowel. This suggests that place of articulation is based on a 'radial' representation anchored on the neutral vocoid. 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 fig 2) to back vowel contexts (on the left-hand in fig 2). F2 onset (Hz)

6. Couplings between perceptual predispositions for speech

6.1. Models of speech development

Human infants are born with predispositions for perceiving all possible phonetic contrasts, which are then activated or not as function of the presence vs absence of the corresponding contrast in the linguistic environment. This fairly classical view on speech development (werker&Tees,1999) is grounded on considerable amount of empirical evidence. Neonates can already discriminate between a range of phonetic categories (Eimas, Siqueland, Jusczyk& Vigoroto,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,1984b).

Infants 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 depend 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 upper-stage phonological level. Alternatively, the adult perceptual space might not be straightforwardly related to the universal predispositions (Kuhl, 199-200-), What he considers as '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 creates new functional entities inside which features are integrated. The term coupling is common-place in the study of visual perception, e.g. for describing perceptuo-motor integration in depth perception (Hochberg,1981).

7. 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 fig 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 short vs long positive VOT categories (e.g. English, Fig 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, 1987), and the perceptual boundary is located around 0 ms (Seniclaes,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.

In support of the coupling hypothesis, examination of the studies on children raised in Spanish-speaking environments showed that the 0 ms VOT boundary is not predicted by the infant's predispositions (Lasky at &l.; 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 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, Lienard, 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 (Jacobson 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-F3 transitions is probably present in the newborn, although there is no direct evidence on this point. The predispositions for perceiving F2-F3 transitions might straightforwardly be used in four-category languages, two binary features allowing to discriminate four place categories.

However, the natural F2-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 two boundaries (FIG 4). These boundaries can only be obtained by trade –off between F2-F3 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 cued but instead precluded trade-off between F2 and F3 transitions means coupling between predispositions.

Evidence has been found recently in support of the conjecture by collecting both identification and discrimination responses to / stop+neural vocoid/ synthetic syllables generated by either factorial or combined modification of F2 and F3 transition onsets. Preliminary results (Serniclaes, Bogliotti& Carré,2003; see fig 4) showed that French adult speakers discriminated natural F2 and F3 boundaries reflected trade-offs between F2 AND f3. The fact that perceptual boundaries for place of articulation are built on trade-offs between the coupling hypothesis. These results have since been confirmed with a larger sample of subjects (Bogliotti,2005) two acoustic cues, which are endowed with natural boundaries, provides further supports.

References

[1] Anisfeld, M. (1979). Interpreting Imitating Responses in Early Infancy. *Science.*
[2] Abercrombie, D. (1967). *Elements of general Phonetics.* University press, Edimburg.
[3] Abry, C.& Boe, LJ. (1983). Plateaus, Catastrophes and the Structuring of Vowel Systems. *Journal of Phonetics 17,*45-54.
[4] Badin, P.&Boe, L.G (1987). The vocal Tract Vocalic Nomograms. Acoustic *Considerations.11th International congress of Phonetic Sciences,* Estonia, USSR,352-355.
[5] Badin, P.& Fant, G. (1984). *Vocal Tract Frequency Calculation Techniques*. J. Acoustic.Soc. Am.,87,1290-1300.
[6] Baldwin, A.L. (1967). *Theories of Child Development,* New York: Wiley.
[7] Bates, E. (1976). *Language and Context: The Acquisition of Pragmatics,* New York: Academic Press.
[8] Baudoin de Courtenay, J. (1894). *An attempt to a Theory of Phonetics Alter Nations.* (Article published in Polish in 1894).
[9] Bell, A. (1867). *Visible Speech.* Ed by Simpkin & Marshall, London.
[10] Bloom, L. (1970). *Language Development" Form and Function in Emerging Grammars*. Cambridge, MA: MIT, Press.
[11] Braine, MDS. (1963). *The Ontogeny of the English Phrase Structure.* The Forst Phase, "Language,39.01-13.
[12] Braine, MDS. (1971). The Acquisition of Language in Infant and Child, in *the Learning of Language.*pp. 7-95, C. Reed (ed) New York.
[13] Bronstein, N. (1965). *Vowel Acoustic Characteistic. Phonology Year Book,* Vol.3, Cambridge.
[14] Brown, R., Cazden., &Bellugi, U. (1973). *The Child Grammar form I to III*.
[15] Brownman, C.P.&Goldstein, L. (1987). *Towards an Articulatory Phonology.* Phonology Yearbook, Vol.3, Cambridge,219-252.
[16] Bruner, J.S. (1978). *The Social Context of Language.* New York: Wiley.
[17] Carlson, R., FANT, G.&GRANSTROM, B. (1975). *Two- Formant Models, Pitch and Vowel Perception,* Ed.by FantG.London.
[18] Cazden, C.B. (1972). *Child Language and Education.* NewYork: Holt, Rinehart&Wintson. Chomsky, Noam.
[19] Chomsky, N.& Halle, M. (1968). *The Sound Pattern of English.* Ed by Harper&Row, N.Y.
[20] Comrie, B. (1981). *The Languages of the Soviet Union.* Cambridge Univ.Press, N.Y.
[21] Crothers, J. (1978). *Typology and Vowel Systems in Phonology.* Standford Univ.
[22] Darwin, C.R. (1859). *The Origin of Species.* Massachussets. Ed in 1964.
[23] De Grolier, E. (1989). *Aux origins du language. Les origins,* l'Harmattant, Paris,189-251.

[24] Edwards, M (1973). *Speech Intelligibity and Children Verbal Behaviour.* Child Development 47.452-468.
[25] Edwxards, M. (1971). One Child's Acquisition of English Liquids. *Papers and Reports on Child Language Development.3*.101-9.
[26] Firth, J.R. (1948). Sounds and Prosodies.*Transactions of the Philological Society.* 127-152.
[27] Fromkin et al. (1974). An Example of Linguistic Consciousness in the Child. In *Studies of Child Language Development,* pp 155-8, C. A Ferguson and D.I. Slobin (eds)., New York:
[28] Cutler, A. (1990). Sound Patterns of Men's and Women's name. *Journal of Linguistics. 26*, 471-482.
[29] Greenberg, J.H. (1958). A theory to Language Development. *International Journal of American Linguistics*. P 94.
[30] Harris, Z. (1955). *Transformation in Linguistic Structure.* Stranford University Press. 4, vol. Stranford California.
[31] Hadege, C. (1982). *La structure des langues.* Collection "Que sais-je". P.U.F.
[32] Halliday, M.A.K. (1975). *Learning how to Mean.* New York: Elvesier.
[33] Hayes, L.A& Clark, E. (1970). What's in a Word? On the Child's Acquisition of Semantics in his First Language. In *Cognitive Development and the Acquisition of Language*.pp65-110, T.E Moore ed. New York. Academy Press.
[34] Ingram, D. (1976). Phonological Disability in Children. *Journal of Verbal Learning and Verbal behaviour 13 448*-456.
[35] Jakowitz, E.R. (1980). Development in early Play. *Psychology.*
[36] Jakobson, R., Fant, C.G.M., &Halle, M. (1963). *The Distinctive Features and Their Correlates.* Cambridge, Mass.
[37] Jakobson, R. (1941). Child Language, Aphasia and Phonological Universals *(English translation published by Mouton in 1968).*
[38] Jakobson, R. (1949). *On the Identification of Phoneme Entities.* Travaux du Cercle Linguistique de Copenhague 5 :205-213. Reprinted in Jakobson 1962:418-425.
[39] Janson, H.R. (1991). The Phonemic System in Khoisan, Clicks. *Journal of American Linguistcs, vol2.*
[40] Kent, L. (1981). An Event Approach to the Study of Speech Perception from a Direct-Realist Approach, *Journal of Phonetics, 14:3*-28.
[41] Kuhl, P.K(1994). Learning and Representation in Speech and Language. Vol4. 812-822.
[42] Kuhl, P.K. &Andrusky, J.E. (1979). Early Phonetic Experience and Phonetic Perception. Summer Institute of Linguistics.p.14.
[43] Trubetskoy, N.S. (1939). *Grandzuge des phonologie.* Travaux du cercle de Prague,7.272p.
[44] Vygotsky, R. (1962). T*eaching and Learning Children Sounds.* Press. Moscou.