Linguistica Silesiana 27, 2006
ISSN 0208-4228

ARKADIUSZ ROJCZYK
University of Silesia, Katowice

CATEGORISATION OF SPEECH SOUNDS IN AUDITORY PERCEPTION

The paper aims to present the phenomenon of categorization in speech perception. It commences with showing the diversity and distribution of speech sounds in the world's languages. Next, certain distributional tendencies, such as back vowel rounding and low vowel nasalization, are presented and discussed in the light of auditory and acoustic principles. Categorical perception, as the pivotal concept of the paper, is delineated from a procedural and methodological point of view. Last but not least, the article discusses the arguments against categorization is speech perception as well as various attempts to define a unit of perception other than a phonetic segment As a conclusion, proposals for the reconciliation between categorical and continuous modes of perception are shown.

Richness and variety of speech sounds

Even very cursory acquaintance with one or two foreign languages impresses on us how diverse speech sounds are. Polish learners of English must face the fact that there are no qualitatively same vowels in their native language and English. English learners of Polish, on the other hand, are forced to struggle with abundant fricative consonants, so common in Polish. It is apparent that languages exploit differently articulatory combinations to organize their phonemic systems; some sounds may be favoured by one language and at the same time absent in another.

Throughout over 100 years of study by phoneticians, the domain of speech sounds has been scrupulously catalogued (Kluender 1994). The results of their meticulous work have been collected in the UCLA Phonological Segment Inventory Database. It comprises a representative sample of the phonological categories of the world's 451 distinct languages (De Boer 1999). The phonological inventory for each language was determined by listing phonemically contrastive sounds in a language. As a result, each sound is not only phonemically contrastive but also not predictable from context

(Epstein 2000). Despite some voices of criticism, for the time being it is the largest and most reliable catalogue of speech sounds found in the world's languages[1].

What becomes apparent upon careful inspection of UPSID data is the sheer diversity of speech sound categories. Maddieson (1984) classified 869 sound categories occurring in languages: 558 consonants, 260 vowels and 51 diphthongs. An ample testament to the rich variety of sounds used in the world's languages is the fact that he needed as many as 58 phonetic attributes for the classification. Not only are the sound categories diverse themselves, but also there is significant heterogeneity in sound inventories across languages. The Rotoka and Mura languages need only 11 phonemes to organize their phonological systems, whereas the language !Xũ exploits no fewer than 141 phonemes (Maddieson 1984, Kluender 1994, Epstein 2000).

The majority of all 869 phonemes are relatively rare, while a handful are extremely common. One hundred and sixteen of the languages in UPSID have at least one sound that no other language in the database has or, assuming a different perspective, 47% of speech sounds classified in UPSID are unique, i.e. occur only in one language (Epstein 2000). From the data collected we can also learn that all catalogued languages have stop consonants. The majority of languages have three places of articulation for stop consonants – bilabial, alveolar, and velar. Moreover, over 80% of languages utilize the voiced – voiceless distinction at these three places of articulation (Maddieson 1984).

The UPSID database lists 30 different fricatives, most of them rare, but a few very common. In fact, over 90% of all languages use fricative consonants. Alveolar [s] is found in 80% of the classified languages with palato-alveolar [ʃ] and labiodental [f] being fairly frequent as well. In terms of the voiced – voiceless opposition only roughly 30% of fricatives used by the world's languages are accompanied by the vocal cord vibration, the remaining majority being voiceless (Maddieson 1984).

Vowel systems found across languages also abound in diverse categories. Phoneticians have found at least 44 different vowels in the world's languages (De Boer 1999). Some languages make do with only 3 vowels, whereas others use as many as 24 (Maddieson 1984), for example, Norwegian utilizes 15 different vowel qualities (De Boer 1999). 21.5% of the UPSID languages have five vowels and most of the five-vowel systems tend to use the same vowels (Maddieson 1984). Almost all languages contain [i], [a] and [u] – they appear in 87%, 87% and 82% of the languages respectively. Many languages also use [e] (65%) and [o] (69%) (De Boer 1999).

---

[1] Criticism is forwarded by phonologists who claim that UPSID applies phonological comparisons using arbitrary selections of the phonetics of the languages involved. Critics also indicate the fact that phonemic systems belonging to the phonological level of comparison are compared in phonetic terms thus misrepresenting the abstract nature of a phonological system, which must lead to gross oversimplification of phonetic patterns (see Simpson 1999). Kingston (1991) suggests, on the contrary, that some degree of imprecision is, in fact, a virtue for statistical evaluation of what is and is not a typical phonetic contrast.

Auditory perception and its effect upon the distribution of categories

The systematic analysis of languages collected in UPSID database reveals the tendency of systems to favour some categories over another. In other words, there are common sounds found in most languages, whereas some sounds appear to be rare or even unique. The first attempts to explain the irregular selection of phonetic inventories concentrated in the domain of articulation. Phoneticians have typically attempted to account for phonetic irregularities in terms of physical and physiological constraints on speech production, suggesting that some sounds are merely easier to produce than others (Kluender 1994). Keating, Linker, and Huffman (1983), for example, recommended that the preference of languages for voiceless rather than voiced stops is motivated by aerodynamic and articulatory efficiency. The role of articulatory ease was also evidenced in the study by Lindblom and Maddieson (1988). They classified consonants using the following labels: those requiring "basic articulation", "elaborated articulation", "complex articulation." Inspecting the phonemic inventories of languages, they found that languages have a strong tendency to exploit "basic" consonants before incorporating "elaborated" consonants and to include "elaborated" consonants before "complex" consonants. According to the authors, articulatory cost does play an influential role in the selection of consonants, therefore languages tend to use articulatorily less demanding consonants before incorporating more complex consonants.

Not all distributional tendencies, however, can be explained in terms of articulation itself. It is especially the case with vowels, since it is rather impossible to explain why certain vowels are less common than others in articulatory terms. Accordingly, the rationale for the selection and distribution of vowels across languages is sought in the domain of auditory perception. Sounds utilised by a system perform their communicative function only when they are perceived discretely. As a result, sound systems of languages have adapted to exploit general characteristics of auditory perception so as to optimise their phonetic distinctiveness both acoustically and auditorily (Kluender 1994). The above assumption led to the formulation of the **Auditory Enhancement Hypothesis,** which claims that spoken linguistic information is transmitted via an acoustic/auditory channel, and language users exploit the transmission characteristics of auditory systems in order to maintain the integrity of their linguistic message (Diehl and Kluender 1989). In the text below we will resort to Auditory Enhancement Hypothesis in order to explain two selectional phenomena; lip rounding in back vowels and nasalization in low vowels.

One dominant tendency among languages is that back, but not front, vowels are produced with the lips rounded. In the UPSID database there are 254 languages that have the high back rounded [u], but only 20 languages that use the high back unrounded vowel. Similarly, 271 languages collected in UPSID have the high front unrounded vowel [i] and only 21 have its high front rounded counterpart (Maddieson 1984). No apparent articulatory constraint provides a physiological link between the front position of the tongue and lip rounding. The explanation of this tendency can be, however, found in the acoustic properties of front and back vowels. The vowels [i] and [u] are primarily distinguished by the frequency of the second format (F2). Despite having relatively low frequency first formants (F1), in the production of [i] F2 rises (approximating the third formant, F3), whereas the production of [u] results in low F2 (near F1). It was discovered that the effect of lip rounding lowers F2, thus making back vowels even more acoustically backlike (Diehl and Kluender 1989). For [u] and other back vowels, lip rounding serves to make them acoustically and auditorily more distinct from front vowels. The tendency across languages for back vowels to be rounded is consistent with the view that systems distribute sounds with a view to making them maximally distinct.

Nasalization can be generally described as the lowering of the velum in order to allow air to escape through the nose. It is a yet another feature of vowels that exhibits strong distributional tendencies. Across languages, low vowels tend to be nasalized much more often than high vowels (Ohala 1974). Any articulatory or mechanical explanation of this phenomenon can be rejected since there is no evidence for any indirect coupling between the tongue and velum (Kluender 1994). What is more, electromyographic analysis shows that the velum is actively raised during the production of high vowels, thus preventing their nasalization (Lubker 1968). Again, the rationale for nasalizing back vowels seems to be found in their acoustic and articulatory characteristics. The distinction between high and low vowels depends on the frequency of F1 – low vowels have relatively high frequencies of F1 compared to high vowels. One of the acoustic consequences of slight nasalization is the raising of the frequency of F1 (House and Stevens 1956). According to the Auditory Enhancement Hypothesis, low vowels are more nasalized than high vowels by virtue of making their high F1 even higher, consequently strengthening the high-low distinction. Perceptual studies appear to bear out the acoustic explanation of the nasalization of low vowels. Listeners tend to identify nasalized vowels of varying height as lower in comparison to their nonnasalized counterparts (Krakow, Beddor et al. 1988).

Categorical perception

The mode of auditory perception referred to as categorical perception was first observed and characterised at Haskins Laboratories in the study initiated by Liberman et al. (1957). Categorical perception occurs when changes across some dimension of the speech signal are not perceived continuously but in a discrete manner. Listeners are limited in their ability to discriminate between different sounds belonging to the same phoneme category. In other words, the sounds within a category are identified in an absolute manner, hence the discrimination between the sounds is only possible when they belong to different categories (Massaro 1994). In a typical experiment, a series of synthetic consonant-vowel (CV) syllables, which vary in an acoustic parameter of F2 and range perceptually across initial consonant [bV], [dV], and [gV], are presented to listeners for identification (Diehl et al. 2004). Two consistent patterns are evident in the results. First, it transpires that there are abrupt boundaries between categories, i.e. listeners, who are presented with a number of synthetic consonants ranging from e.g. [b] to [d], perceive the same precise boundary when the category for [b] changes into the category for [d]. Second, discrimination accuracy was very poor for stimulus pairs within a phoneme category, but nearly perfect for the stimuli that belonged to two categories. This leads to the conclusion that the ability to discriminate between sounds is closely related to the presence or absence of functional (phonemic) contrasts between sounds.

The ability to categorise sounds in auditory perception has aroused interest in the question whether we are in any way biologically predisposed for learning phonetic categories. Eimas (1991), assuming a nativist approach, asserts that infants are born with a specialised biological predisposition to discriminate a universal set of phonetic contrasts. Accordingly, the process of learning a language involves either a decline or a reorganisation of this universal sensitivity. Kluender (1994: 200), however, expresses his scepticism about this hypothesis claiming that "evolution does not give rise to structures and processes *de novo.* Nature will either make do with existing structures or processes, or it will adapt existing processes, in which case there must be ample selective pressure to do so." It is therefore doubtful that languages favour certain sound inventories owing to special biological predispositions. There is simply too much sound diversity in languages and innately specified processes should instead be conducive to greater conformity in phonetic inventories.

Another more plausible attempt to explain the ability to categorise sounds in speech perception stresses the role of experience. Experience seems to play a critical role in categorising speech sounds. Different languages partition the domain of possible speech sounds differently, which implies that the categorisation depends on linguistic experience (Kluender 1994). Two concepts appear to be inextricably connected with experience in the speech categorisation; the lack of invariance and prototype.

Wittgenstein (1953, reviewed in Kluender 1994) noted that the necessary and sufficient conditions rarely, if ever, exist for establishing membership in natural categories. It is also widely accepted in speech perception studies that phonemes do not have invariant properties, which is referred to as the **Lack of Invariance** (Cutler and Clifton 1999). Due to many contextual dependencies, acoustic and articulatory properties of speech sounds vary and it seems to be impossible to indicate what the best representative of each category is.

The spectrogram in Fig. 3 displays the contextual dependencies of a plosive consonant [k]. The articulation of [k] before the following vowel [y] is more front, whereas when the back [u] follows, [k] has more back articulation.

Another ubiquitous finding in categorisation is the fact that members of categories are not equally good members. The categories are said to have a graded structure and are often described as being centred around an ideal instance of the category – **a prototype** (Rosch 1975). In general, categories exhibit an internal structure that not all instances of a category constitute equally good members. Kuhl (1991) conceptualised a **Perceptual Magnet Effect,** which states that the generalisation of two variants of the same category is greater when one of the stimuli is judged to be near ideal. It was also demonstrated in the same study that instances nearer to the prototype are better examples of the category.

In the light of the scarcity of clear-cut divisions between boundaries, and variance of different instances of the same category, the studies in speech perception endeavour to explain how infants can learn phonetic categories. Rosch (1978) suggests that categories are learned as clusters of features that tend to co-occur. In other words, infants learn categories through experience with co-occurrence of features that are correlated with the category and with each other. An infant learning the phonemes of his or her own language must learn the category boundaries, which means they must learn how acoustic and auditory attributes co-occur and how they define phonetic categories. Kluender (1994) discusses interesting phenomena of how discriminable acoustic differences that have the potential to define phonetic categories in some languages, but do not define the categories in an infant's language, are assimilated into the structure of categories relevant to the infant's language. First, acoustic and auditory attributes of two contrasting nonnative categories are well correlated with attributes of a single native category. For example, dental stops are acoustically quite similar to alveolar stops. Accordingly, only two of the world's languages utilise the dental – alveolar distinction, namely Aranda and Nunggubuyu (Maddieson 1984). Second, we can observe a process of **single-category assimilation** (Best et al. 1988). American English alveolar stops may have contextual allophonic realisations as retroflex when followed by an approximant [r], e.g. in the word *adroit*, or dental when followed by dental fricatives, e.g. *width*. It seems that retroflex and dental attributes are well correlated within the category of alveolar stops. Werker and Lalonde's (1988) study appears to confirm this hypothesis – stimuli that are identified by native Hindi listeners as dental or retroflex are all assimilated into the category of alveolar stops by native English learners. Third, it has been noted that attributes of one category of a two-category nonnative contrast can be well correlated with attributes of a single native category while attributes of the other category of the nonnative contrast are less correlated with attributes of the native category. In the literature the phenomenon is referred to as **category-goodness assimilation** (Best et al. 1988). One example of this assimilation is a Farsi distinction between velar and uvular stops. It transpires that native English listeners do not lose the ability to discriminate Farsi velars from uvulars. Instead, Polka (1992) notes that they perceive the Farsi velar and uvular stops as being good and poor instances of English [g]. The study demonstrates that English listeners assimilate uvular stops into the category of English velar plosives, however, since they share fewer attributes with English [g], they are perceived as poor instances of the prototype. Finally, Best et al. (1988) report **two-category assimilation,** which describes the situation when the native language does not share a contrast with a nonnative language but the native language has an analogous contrast that facilitates the perception of the nonnative contrast. Jamieson and Morosan (1986) show that even though French does not utilise the voiced – voiceless distinction for dental fricatives, native French listeners can differentiate between English voiced and voiceless dental fricatives, perceiving them as versions of French dental stops. One can conjecture that French listeners perceive the English dental fricatives as version of French dental stops because the acoustic and auditory attributes of the dental fricatives are correlated with attributes of the French dental stops.

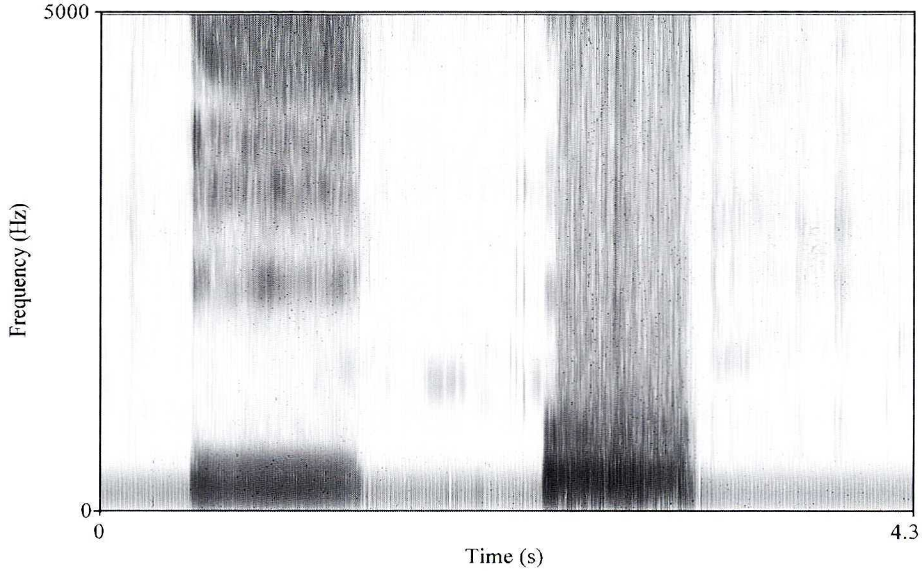

Fig. 1. Spectrogram of author's Polish vowels *III* (on the left) and [u] (on the right). One can see the rise of F2 reaching F3 in [i] and low F2 near F1 in [u].

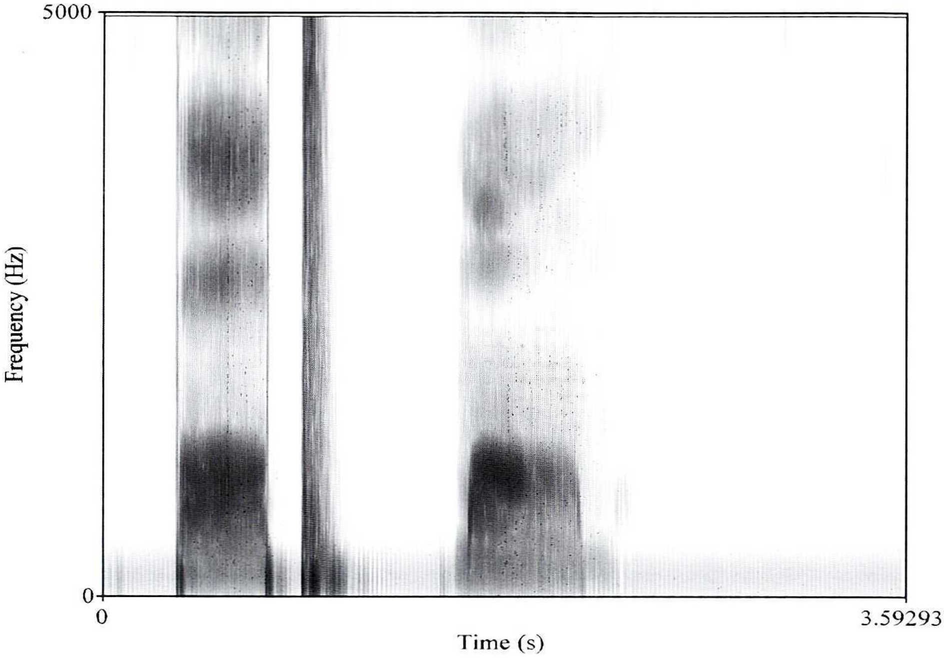

Fig. 2. Spectrogram of the author's Polish syllables [pap] (on the left) and [mam] (on the right). The mean frequency of Fl for [a] in [pap] is 871 Hz. The mean frequency of Fl for [a] in [mam], due to stronger nasalization resulting from the contextual effect of the preceding and following nasals, is 950 Hz.

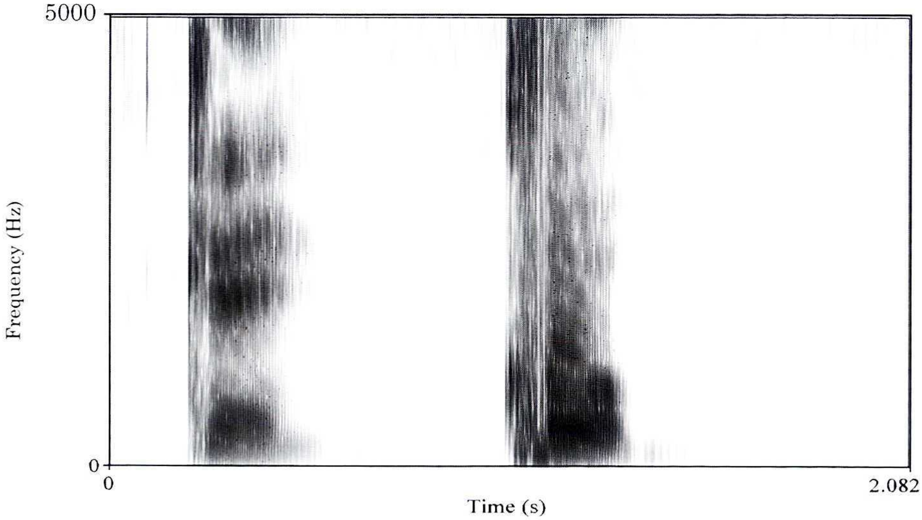

Fig. 3. Spectrogram of the author's Polish syllables [ky] (on the left) and [ku] (on the right). One can discern contextual influence on the articulation of [k] by a following mid-front vowel [y] (fronting) and a back vowel [u] (retracting).

The demise of categorical perception?

It is fair enough to say that the study of categorisation of speech sounds has dominated to a large extent the investigation of auditory perception. Massaro (1987: 90) writes, "The study of speech perception has been almost synonymous with the study of categorical perception." Along the development of categorical perception the evidence has been gathered that casts some doubts on the mode of perception that parses sounds into categories. This led Massaro (1994: 225) to write,

> categorical perception is a belief that will not die or fade away easily. Many textbooks and tutorial articles also state that speech is perceived categorically... However, I have argued in too many places that previous results and more recent studies are better described in terms of continuous perception – a relatively continuous relationship between changes in a stimulus and changes in perception.

One of the most severe objections to categorical perception concerns the methodological approach, namely the relation between identification and discrimination. In a typical experiment two alternative categories are chosen and a set of stimuli between them is synthesised. Listeners are asked to identify each of the stimuli as one of the categories. They are also asked to discriminate between those stimuli. According to Massaro (1994) the results have been misinterpreted as showing categorical perception because discrimination performance was predicted by identification performance. He claims that this relation between identification and discrimination provides no support for categorical perception for two reasons. First, the categorical model provides an inadequate description of the relation between identification and discrimination and has not been proved to provide a better description than continuous models. Second, even if the results provided undeniable support for the categorical perception model, explanations other than categorical are also possible.

Another argument undermining the tenets of categorical models is the rate of speech processing and the exact unit of perception. Certainly, there are findings indicating that it is a phonetic segment, which is a unit of perception. For example, segment-size errors are common in both production and perception, orthographies attempt to approximate segmental structure of sounds, diachronic studies suggest that the segment is a critical unit of language change (Remez 1994). However, there is also ample evidence against a traditional category (phoneme) as a unit of perception. It transpires that the transmission rate of the speech exceeds our perceptual capacity. Phonetic segments occur at a rate of 10 to 20 per second, whereas the studies show that humans cannot identify nonspeech signals at even half this rate (Massaro 1994). Moreover, it has been evidenced that words can be recognised without necessarily identifying the phonetic categories that make it up (Warren 1982). In L1 acquisition studies it has been observed that a child acquires speech segments in terms of variety of sizes – syllables, words, or even phrases. For instance, the child learns to identify the word *through* not as a sequence of three segments [θru:], but as a whole syllable (Peters 1983).

The strongest candidate to supersede a phonetic category as a unit of perception is a syllable. Psychological studies using the targeting paradigm show that syllables such as for example [ba], [pi], or [de] are recognised more rapidly than the stop consonants [b], [p], or [d] themselves (Segui et al. 1981). Other models propound units consisting of a vowel plus a preceding syllabic onset or a following coda, which are called demisyllables (Samuel 1988). Klatt (1979), on the other hand, puts forward a diphone as unit of perception. A diphone is conceived of as the stretch of speech from the midpoint of one phoneme to the midpoint of the next, which comprises all the information relevant to the contextual effects on phonemic realisation (Cutler and Clifton 1999).

Concluding remarks

At the moment it is far too early to declare whether the speech perception is categorical or not. We have seen that categorisation indeed is present in the description of speech sounds of languages. It has also been shown that there are universal tendencies in the distribution of categories due to acoustic and auditory principles. There is evidence as well that listeners are in a position to categorise synthetic stimuli. On the other hand, a phonetic category as a perceptual unit has been questioned on different grounds. Not only does the perception of speech categories exceed human capacity for perceiving sounds, but also different studies attest to units larger than phonetic categories as units of perception. Recently there have been attempts to reconcile categorical and continuous approaches to speech perception using models based on neural networks (see Damper and Harnad 2000 for a review). It is suggested that speech perception may be continuous on a sensory level and next it is categorised as a mode of decision making. There are also endeavours to explain the relation between continuous and categorical perception using the phenomenon of filling whereby we perceive two different speech events as the same category. Categorical perception supposedly occurs because the sensory system blurs any stimulus differences within a category, and perhaps sharpens stimulus differences between categories.(Massaro 1994).

References

Best, C.T., McRoberts, G.W., Sithole, N.M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14, 345–360.

Cutler, Anne, Cliffton, Charles Jr. (1999). Comprehending spoken language: a Blueprint of the listener. Colin, M. Brown and Peter Hagoort (eds.) *The Neurocognition of language*. Oxford: Oxford University Press.

Damper, R.I., Harnad, S.R. (2000). Neural network models of categorical perception. *Perception and Psychophysics* 62, 843–867.

De Boer (1999). Investigating the emergence of speech sounds. Dean, T. (ed.) *IJCAI*199, 364–369. San Francisco: Morgan Kaufmann Publishers.

Diehl, R.L., Kluender, K.R. (1989). On the objects of speech perception. *Ecological Psychology* 1, 121–144.

Diehl, Randy L. Lotto, Andrew J., Holt, Lori L. (2004). Speech perception. *Annual Review of Psychology* 55, 149–179.

Eimas, P.D. (1991). Comment: Some effects of language acquisition on speech perception. I.G. Mattingly and M. Studdert-Kennedy (eds.) *Modularity and the Motor Theory of Speech Perception*, 111–116. New York: Erlbaum.

Epstein, Melisa A. (2000). All the sounds of all the world's languages. *UCLA Working Papers in Phonetics.*

House, A.S., Stevens, K.N. (1956). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders* 21, 218–232.

Jamieson, D.G., Morosan, D.E. (1986). Training non-native speech contrasts in adults: Acquisition of the English [Θ] – [ð] contrast by francophones. *Perception and Psychophysics* 40, 205–215.

Keating, P., Linker, W., Hoffman, M. (1983). Patterns of allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11, 277–290.

Kingston, J. (1991). Phonetic underresolution in UPSID. *XIIth International Congress of Phonetic Sciences*, 359–362. Aix-en-Provence: Universite de Provence.

Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics* 7, 279–312.

Kluender, Keith R. (1994). Speech perception as a tractable problem in cognitive science. Gernsbacher, M.A. (ed.). *Handbook of Psycholinguistics*, 173–217. San Diego: Academic Press.

Krakow, R.A., Beddor, P.S., Goldstein, L.M., Fowler, C.A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America* 83, 1146–1158.

Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect". For the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50, 93–107.

Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C. (1957). The discrimination of speech sounds within and across, phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368.

Lindblom, B., Maddieson, I. (1988). Phonetic universals in consonant systems. L.M. Hyman i C.N.Li (ed.). *Language, speech and mind*, 62–78. New York: Routledge.

Lubker, J.F (1968). An electromyographic-cinefluographic investigation of velar function during normal speech production. *Cleft Palate Journal* 5, 1–18.

Maddieson, I. (1984). *Patterns of sound.* Cambridge: Cambridge University Press.

Massaro, Dominic W. (1994). Psychological Aspects of Speech Perception. Gernsbacher, M.A. (ed.) *Handbook of Psycholinguistics*, 173–217. San Diego: Academic Press.

Massaro, Dominic W. (1987). Categorical partition: A fuzzy logical model of categorical behavior. S. Harnad (ed.) *Categorical Perception: The Groundwork of Cognition*, 254–283. Cambridge: Cambridge University Press.

Ohala, J.J. (1974). Experimental historical phonology. J.M. Anderson i C. Jones (ed.) *Historical linguistics: II. Theory and description in phonology*, 353–389 Amsterdam: North-Holland.

Peters, A.M. (1983). *The units of language acquisition.* Cambridge: Cambridge University Press.
Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics* 52, 37–52.
Remez, Robert E. (1994). A guide to research on the perception of speech. Gernsbacher, M.A. (ed.). *Handbook of Psycholinguistics*, 173–217. San Diego: Academic Press.
Rosch, E.H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 3, 193–233.
Rosch, E.H. (1978). Principles of categorization. E. Rosch and B. Lloyd (eds.) *Cognition and Categorization.* New York: Erlbaum.
Samuel, A.G. (1989). Insights from a failure of selective adaptation: Syllable initial and syllable final consonants are different. *Perception and Psychophysics* 45, 485–493.
Segui, J., Frauenfelder, U.H., Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology* 72, 471–477.
Simpson, Adrian P. (1999). Fundamental problems in comparative phonetics and phonology. Does UPSID help to solve them? *Proc. XIVth ICPhS* 1, 349–352.
Warren, R.M. (1982). *Auditory perception: A new synthesis.* New York: Pergamon.
Werker, J.F., Lalonde, C.E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology* 24, 672–683.
Wittgenstein, L. (1953). *Philosophical investigations.* Trans. by G.E.M. Anscombe. Oxford. Blackwell.