

1995

The role of suprasegmentals in speech
perception and acquisition

Jacques Mehler Josiane Bertoncini

Emmanuel Dupoux

Christophe Pallier

LSCP CNRS-EHESS, 54 Bd Raspail, 75006 Paris FRANCE

The role of suprasegmentals in speech perception and acquisition

Jacques Mehler Josiane Bertoncini

Emmanuel Dupoux

Christophe Pallier

LSCP CNRS-EHESS, 54 Bd Raspail, 75006 Paris FRANCE

During the last world war scientists tried to improve the intelligibility of speech perceived under very noisy conditions. To this end they studied the process of language comprehension (Miller, 1951) and gave birth to the field of psycholinguistics. Later, during the early sixties, this area became an important chapter of the fashionable cognitive revolution. Psychologists became aware of the usefulness of formal linguistic approaches (Miller 1962). Work in acoustic-phonetics had uncovered some acoustic correlates of putative speech units (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967); (Stevens & House, 1972; Stevens & Blumstein, 1981); and (Klatt, 1979, 1989); among others. Today, psycholinguists still study speech comprehension to which they have added the study of speech production (Levelt, Schriefers, Vorberg, Meyer, Pechman, & Havinga, 1991) and also (Dell, 1988). In this paper we review some recent cross-linguistic investigations of speech comprehension. Developmental and neuropsychological considerations complete this research.

Traditionally, there has been a phonemistic or phonemo-centric bias to the study of speech perception. Linguists, phoneticians, psycholinguists and most students of language have assumed that if one understands how phonemes (or the distinctive features that make-up the phonemes) are perceived then one understands automatically how speech is perceived. This view, which is possibly misguided, derives

from our graphemic-phonemic writing system. In such systems, letters correspond, by and large, to the segments of the speech signal. Thus, if one can read a text by discovering the sequence of letters why can't homologous processes explain speech perception as the sequential discovery of the phonemes in the signal? This naive but straightforward tack has largely become central to investigations of speech perception over the best part of the last thirty years.

Different languages use a subset of phonemes. Speakers of these languages always use the same procedures to process speech. The only significant differences that can be observed between speakers of the various languages relates to the phonemes each has to discover. Speakers of some languages sometimes have to deal with two phonemes in a dimension where speakers of other languages just have to deal with one. How do English and say French differ? In the repertoire of the phonemes they use. So, for the speakers of English one /u/ exists in the vowel space where the French hear two. Likewise, the French do not hear a clear difference among /t/ and /th/. However, except for such minor classificatory contrasts speakers of both languages behave exactly alike.

Today, we can no longer adhere to such a phonemo-centric position. Linguists, but also psycholinguists have discovered that structures that organize phonemes into syllables, syllables into prosodic wholes, etc. also play a predominant role in the understanding of speech. So, for instance, if French is a language, which like English that allows for syllables with onsets and codas that can be rather complex, e. g., *spleen*, *tract*, *orchestre*, the Japanese language only allows for syllables without onset clusters and without codas. Only a nasal N can appear after a vowel and before a consonant, as in the word *van*. Basically, however, the language has a CV structure. Today, we know that such a structural contrast has processing consequences, see Otake, Hatano, Cutler, and Mehler (1993). So the nature of the molar units that are implemented in

a language determine the behavior of Ss.

Languages differ not only because of the phoneme repertoire that they use but also in the way they use suprasegmental information. In Italian a pair of words can often only be distinguished thanks to the syllable that bears the accent, e. g., *ancOra vs Ancora, mEta vs metA, etc.* In contrast, in French, an oxytonic language, all the words bear an accent on their last syllable. Does this difference between Italian and French affect the way in which Ss process speech? If one believes recent results by Sebastian et al. (in preparation) this is the case. If so, one should not neglect the rhythmic and prosodic aspects of speech when striving towards a better understanding of speech perception. The only way to secure this is to liberate oneself from the above mentioned phonemo-centrism.

In this paper, we begin by reviewing some recent discoveries about the biology of spoken language that show a specialization of the left hemisphere for the native language. It is our contention that even this basic specialization is made possible by emphasizing the prosodic properties of surrounding language.

The biology of spoken language

Classically, the search for the cortical loci that sustain language competence is, by and large, undertaken with an outlook similar to the one referred to above. Most neuropsychologists hypothesize areas for each of the components of linguistic analysis: a phonological area, a lexical area, a syntactic area and possibly an area where utterances are understood. The area for phonology is conceived of as confined to the identification and production of phonemes. As we shall illustrate below, the picture has become considerably more complex than this.

The advent of powerful new brain-imaging techniques has improved our understanding of the anatomical structures that allow for language to develop in the human brain. Among the most informative, we must

include positron emission tomography (PET) scans, functional magnetic-resonance-imagery (MRI), magneto-encephalography and the high-density brain evoked response potential (ERP). Several PET based studies have explored the cortical activation that arises after visual word presentation, Petersen, Fox, Snyder, & Raichle (1990), word generation, listening to words Frith, Friston, Liddle, & Frackowiak (1991), and listening to phonological sequences in non-words Demonet, Chollet, Ramsay, Cardebat, Nespoulous, Wise, Rascol, & Frackowiak (1992). Zatorre, Evans, Meyer, & Gjedde (1992) showed that phonetic processing is, by and large, located in Broca's area. In these studies, only isolated items were presented so it is not surprising that the areas uncovered were restricted to Broca's area.

Mazoyer, Dehaene, Tzourio, Frak, Murayama, Cohen, Levrier, Salamon, Syrota, & Mehler (1993) carried out a study using connected speech and examined which areas of the cortex are activated when Ss pay attention to simple narratives. They discovered that when right-handed French males listen to a simple story in French, a left hemisphere neural network shows increased activation; the network includes large parts of the temporal and frontal areas of the cortex. In contrast, when Ss listen to a story in an unknown language (Tamil), both the left and the right superior temporal gyrii are symmetrically activated. This result suggests that the left hemisphere is specialized in the processing of French utterances rather than foreign language ones. Moreover, in other studies using sentences with pseudo-words, sentences with correct syntax which make no sense, and lists of content words, it was shown that the anterior temporal poles were activated only when Ss heard sentences and not when they listened to lists of words. This finding can be interpreted in one of two ways: temporal poles play a role in syntactic processing or, much more probably, temporal poles intervene in the processing of the familiar prosodic properties of the native language.

The above findings raise many questions. How does the brain become specialized in the processing of utterances of the native language? Shortly after birth, the brain has no information about which language it will have to master. Thus, the specialization of the left-hemisphere network for the native language cannot yet obtain in infants. Thus, at birth, either all utterances, regardless of language, are processed by the left-hemisphere, or, alternatively, the infants' brain processes utterances drawn from any language symmetrically. Studies of functional lateralization at birth, are unfortunately incomplete at present. Bertoncini, Morais, Bijeljac-Babic, McAdams, Peretz, & Mehler (1989) and Best (1990) have shown that very young infants display a right ear advantage (REA) for syllables and a left ear advantage for other acoustic stimuli. Interestingly, the French infants, in this study, were tested with stimuli which were synthesized to sound like good English syllables. Yet a REA was observed suggesting that the baby's left hemisphere is prepared to process speech stimuli, regardless of the source language. Progressively, however, the left-hemisphere becomes specialized and processes speech samples only if they are drawn from the maternal language. Mazoyer *et al.* also show that when speech is syntactically or semantically impoverished, the left hemisphere's activation becomes increasingly less prominent. A caveat is necessary at this point. In our presentations we always talk about native versus foreign language. However, it is unclear at this point whether the brain distinguishes between native and foreign languages rather than between languages that are understood and languages that are not. Future studies are currently being carried out to address this issue. For the time being we can only state that while adults have specialized areas for processing the prosodic structures of their language, infants have a generalized left-hemisphere superiority for any language-like stimulus.

Though the above hypothesis meshes well with the findings reported

by Mills, Coffrey-Corina & Neville (1993) and also with the studies reported by Molfese (1990), and Vargha-Khadem and Corballis (1979) and others, they have not found a REA for speech. Dehaene-Lambertz and Dehaene (1994) also incite us to be cautious. Indeed, in a study in which they measured high-density ERPs in two-month-olds, Dehaene-Lambertz and Dehaene report a two-component response to the presentation of simple linguistic stimuli. Although one of the components can be assimilated to a discriminative response, no clear superiority of the left hemisphere was observed for this last response though a trend in that direction is apparent from their figures.

All in all, these different studies, inconclusive as they are, allow us to conjecture that the left hemisphere asymmetry for language, becomes progressively more specific for the maternal language. If so, the reported brain imaging studies reflect developmental changes in the speech processing systems. Conceivably, when learning their maternal language, language users develop specific optimal routines. Further evidence supporting this conjecture is found in studies on the development of speech perception in infants and of speech processing in adults.

Development of speech perception

In this section, we evaluate results in three areas of speech development. The first, are about the emergence of phonological knowledge in young children and its relation to lexical acquisition. The second, are about how speakers discover the discrete units into which continuous speech is segmented. Finally, we review results about language acquisition in multilingual environments. In each section, we argue that attention to suprasegmental information, e.g., prosodic and rhythmic information, is essential to our understanding of the observations and that it is because we consider suprasegmental information that we were able to explore new and exciting perspectives.

Development of phonological knowledge

One of the main aims of speech development is to explain the stabilization of the adult speech processing system. As we will show in greater detail below, some models distinguish the process of lexical recognition and the process of phonological encoding while others do not. The extent to which these two processes are interrelated in the adult is still a matter of controversy. However, since both lexical and phonological representations incorporate language-specific information, psycholinguists have to furnish an account of how these information-types are acquired by the young infant. Two broad views can be distinguished.

The first claims that the acquisition of phonological information depends on the prior acquisition of words. For instance, in order to specify the phonemic inventory of a language, linguists classically refer to minimal pairs of words: /p/ and /b/ are different phonemes because *path* and *bath* are different words. If children learn the phonological system of their language thanks to such distributional analysis, one should expect that lexical acquisition precedes the emergence of phonological knowledge in the young infant.

The second view states that, on the contrary, lexical items cannot be acquired unless one disposes of a prelexical representation to code lexical candidates. Evidence stemming from research with adults suggests that prelexical representations are language dependent. If true, this suggests that lexical items can only be acquired after stable phonological/prelexical representation has been acquired (Mehler, Dupoux, & Segui, 1990). In this light, one should expect to find evidence showing language-specific behavior in infants who are under one year old.

In the following discussion, we review how the phonetic inventory emerges in young children. The data show that many phonological contrasts are mastered within the first year of life. However, one cannot rule out the possibility that this acquisition co-occurs with the

elaboration of a lexicon. Nevertheless, recent studies on vowel perception support the view of a precedence of phonological acquisition over lexical acquisition. Based on even more recent studies, we claim that supra-segmental information is acquired before other parts of the language.

The seminal study by Eimas, Siqueland, Jusczyk, & Vigorito (1971), triggered many others that have contributed to establishing that infants are born with a disposition to discriminate phonetic contrasts whether they are instantiated in the language that surrounds them or not. This means that infants often perform discriminations that their parents can no longer make. This is to be expected since many other observations are compatible with it. For instance, we know that language users often fail to discriminate contrasts that do not exist in their own language. Thus, speakers of Japanese have great difficulty distinguishing /l/ and /r/, Spanish speakers fail to distinguish the French vowels /e/ and /E/, and so forth. Moreover, we know that whenever youngsters are given experience with a foreign language they do not display the same problems adults. In particular, Yamada and Tohkura (1992) carried out a detailed study of the acquisition of the /r/ vs /l/ distinction by native speakers of Japanese and found a dramatic decrease in the ability to learn after the age of seven. Why is this? We know that the presence of foreign accent when speaking a foreign language is closely related to age of acquisition of the second language. However, less is known about these difficulties with the perception of speech sounds from foreign languages. Werker & Tees (1984) explored these issues in some detail and corroborated that infants initially have the capacity to discriminate any contrast that is part of a natural language and that this capacity diminishes during the first year of life if the relevant contrast is not experienced by the infant. In a series of cross-sectional and longitudinal studies, Werker and her collaborators established that the ability to discriminate contrasts that infants do not

encounter in the surrounding language decreases between the age of eight and twelve months. Best (1988) conjecture that this apparent impoverishment occurs only when the foreign contrasts can be assimilated to a category that exists in the child's language. Otherwise, the ability to discriminate does not fade. Zulu click discrimination remains very good for American adults and for 12-14 month old and 8-10 month old infants. Are these abilities to discriminate clicks similar to those used to encode other parts of speech? The issue is whether clicks are a special case with special psycho-acoustic properties or not.

Another controversial issue is the extent to which the elaboration of the phonemic inventory is linked or not to the emergence of the lexicon. Indeed, it is at around 12 months that sensitivity to some frequently used words has been claimed to manifest itself. It is thus possible that the phonemic inventory instead of being a prerequisite for lexical acquisition is derived in part from an analysis of lexical distributions. In fact, Jusczyk and Krumhansl (1993) claims that the acquisition of allophonic contrasts (the fact that a single phoneme can be realized in a context dependent way) is acquired very late, i. e. after the acquisition of a large lexicon, or even after the acquisition of orthography. However, more recent experiments suggest that in fact, segmental information is acquired before lexical information.

Jusczyk, Luce and Luce (in preparation) show that six-month-olds have not yet acquired a preference for words with segments that are arranged in serial order that appear frequently in their native language rather than for words whose segments have arrangements that appear less frequently. A preference for the more frequent forms appears by nine months. Thus, the phonotactic structure of the native language begins to be acquired around the age when babies begin to learn the consonantal repertoire that is relevant to their native language.

Kuhl, Williams, Lacerda, Stevens & Lindblom (1992), extended the above studies to vowels. They found that infants learn the vowel

categories corresponding to their native language before they begin to unlearn the phonetic contrasts that have no function in their language. Indeed, Kuhl and her collaborators reported that American and Swedish six-month-old infants have already extracted the prototype for the vowel categories in their language.

Regardless of the details, which will have to be filled in by future research, all these results can be taken to suggest that the reorganization of speech processing antedates learning the lexicon rather than being the result of such learning. Indeed, reorganizations like the ones described are observed before infants begin to acquire words, at least as estimated by either production or by research on the recognition of words heard previously, see Jusczyk and Kemler-Nelson (1993). If so, the infant is already processing the incoming signals to establish the relevant properties that are essential to characterize their native language. This would be sensible since the child has to represent the speech signal before ever being able to use lexical information to process it (Mehler et al., 1990).

Of course, further evidence is badly needed, and the evaluation of lexical knowledge in very young infants is hard to ascertain. Mehler, et al found that French infants, only a few days old, were sensitive to the syllabic structure of utterances. That is, infants distinguish a synthetic syllable /pat/ from a synthetic /tap/, but cannot do so when the vocalic portion is replaced by a voiceless fricative (/pst/ versus /tsp/)¹ The question that arises is whether such behavior is language dependent (since certain languages do allow syllables with consonantal nuclei), and if so whether convergence for the language-specific syllables could arise at an even earlier age. When does the Japanese infant learn that the only legal syllables in Japanese have a

¹ Interestingly, performance goes back to normal when vowels are appended and pre-appended to the stimuli (upstu/ versus /utspu/).

CV structure? When does the Spanish infant learn that there are no syllables in its language that begin with a /sp/onset? These and other questions have to remain without an answer for the time being.

When one moves to supra-segmental information, however, the picture seems to be quite unequivocal: acquisition takes place at a very early age. Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, (1993) show that six-month olds prefer English to Norwegian words. However, these infants have no preference for English over Dutch words. The results are the same for intact or low-pass filtered words. This supports the view that the preference for English words has to do with their prosodic properties. Indeed, while English and Dutch have similar periodic structures English and Norwegian are different in this respect. This study indicates that it is the periodic structure of words that is extracted before the age of six-months.

Moreover, as we show below infants under 2 months of age have already extracted some properties of the prosodic representation characteristic of their native language.

Prosodic constituents

A major difficulty that the infant is confronted with is that of segmenting continuous speech into discrete chunks. This problem could be crucial if it has not been solved when the moment comes to construct lexical entries. Recent simulation studies have shown that it is easier to chunk continuous speech when one has suprasegmental information available than with segmental information only, see Brent (in preparation).

Prosodic structure helps us elucidate early discrimination abilities in young infants. As suggested by Gleitman & Wanner (1982) isolating relevant prosodic units may help bootstrap lexical and grammatical learning. This variant of the prosodic bootstrapping hypothesis was explored by Jusczyk, Hirsh-Pasek, Kemler-Nelson, Kennedy,