Cross-linguistic approaches to speech processing
Jacques Mehler1, Emmanuel Dupoux2, Christophe Pallier1
and Ghislaine Dehaene-Lambertz1

CNRS-EHESS1 and France Telecom2, Paris, France

Recent advances in the field of speech processing indicate that speakers of differing languages process speech relying on units that are appropriate to the rhythmical properties of their maternal tongue. Studies with young infants suggest that the acquisition of these processing routines takes place before the end of the first year of life. Further evidence shows that the left hemisphere initially processes any language and gradually becomes specialized for the maternal language.

Current Opinion in Neurobiology 1994, 4:171–176

Introduction

Speech processing became an active area of research during the Second World War when it was necessary to figure out how to improve understanding under very noisy conditions (reviewed in [1]). Thereafter, a number of investigators have tried to discover the acoustic invariants underlying phonemes and distinctive features (e.g. see [2–4]). Today speech research is an active and thriving domain, but possibly one of the most interesting recent areas to be studied is speech production (see [5,6] and also [7]). In this short review, however, our focus will be on speech perception.

In this paper, we review the effect of language diversity on speech processing. We explore how the brain treats native and foreign utterances. We examine the different ways in which speakers of different languages process connected speech. Lastly, we review studies on very young infants that have uncovered how the processing routines become appropriate to the structure of one’s maternal language.

Currently, the biological specificity of spoken language and the ability of the human mind to cope with linguistic variability are being examined. The studies we describe below have begun to explore the impact of the phonological variations across languages on processing. This line of research also tries to understand how the mind copes with such variability by relating models of adult performance to the initial perceptual abilities of newborns.

Biology of spoken language

Our understanding of how language is related to its underlying anatomical structures has advanced through the use of positron emission tomography (PET) scanning, magnetoencephalography and new event-related brain potential (ERP) methods. A number of brain-imaging studies have explored how visual word presentation [8], word generation, listening to words [9], and listening to phonologically correct non-words [10] activates cortical surfaces. Zatorre et al. [11] confirmed that phonetic processing was located, by and large, in the left hemisphere and that Broca’s area appears to be the main center of these processes. Mazoyer et al. [12] carried out a study to compare the brain areas that are activated when subjects listen to stories spoken in their mother tongue and in a foreign language. In right-handed male subjects listening to their native language, the activity was distributed across the temporal and frontal areas of the left hemisphere. In contrast, when these subjects listened to a story in a foreign language, both hemispheres were activated to the same extent, that is, mostly in the superior temporal gyri. This suggests that the left hemisphere has become attuned to process utterances from one’s own language and not from a foreign language. A study by Hinke et al. [13] used functional magnetic resonance imaging to study activation while subjects silently generated words. They found that Broca’s area was more activated in the left than in the right hemisphere, licensing the hope that in a not too distant future, methods less invasive than PET will allow us to pursue investigations of the areas of the brain that mediate cognitive activities.

Studies using a behavioural measure — the habituation–dishabitation of the sucking-rate response — by Bertoncini et al. [14] showed that very young French infants displayed a right ear advantage for speech, but not for non-speech stimuli. Interestingly, these infants were tested with synthetic stimuli built with parameters of American speech that were, therefore, not

Abbreviations

ERP—event-related potential; PET—positron emission tomography.

© Current Biology Ltd ISSN 0959-4388
selected to sound French. Nonetheless, the right ear advantage was present, which suggests that although the left hemisphere may initially process any speech stimulus, regardless of language, it will, progressively, only process speech samples drawn from the maternal language — or, possibly, from any language that the subject understands. A similar result was obtained by Best [15] with slightly older American infants. The study of Mazoyer et al. [12°] also showed that when speech was impoverished, the level of activation of the left hemisphere decreased. Moreover, the prefrontal cortex in the right hemisphere was activated when subjects listened to possible but non-existent words. These results mesh well with the findings reported by Mills, Coffrey-Corina and Neville [16°], who observed right-hemisphere activation for unfamiliar words in twenty-month-old infants. All these different studies support the conjecture that the left-hemisphere asymmetry for language becomes progressively more and more specific to the maternal language. Further evidence supporting this conjecture is found in studies of the development of infant speech perception.

Fig. 1. Significantly activated brain regions in five experimental conditions: (a) listening to a story in Tamil (n = 5), (b) listening to a list of French words (n = 5), (c) listening to sentences with pseudowords (n = 6), (d) listening to semantically anomalous sentences (n = 6), and (e) listening to a story in French (n = 10). The anterior commissure vertical plane (VAC) and the bicommissural plane (AC–PC) were used to limit the projection of the temporal pole region (TP; light grey shading). The inferior frontal gyrus region (IFG; grey shading) includes the pars opercularis, triangularis, and orbitaris of the third frontal gyrus. A superior prefrontal area corresponding to Brodmann’s area 8 (area 8) was defined on individual magnetic resonance imaging (MRI) using a stereotactic atlas. STG—superior temporal gyrus region; MTG—middle temporal gyrus region. Reprinted with permission from [12°].
Development of speech perception

Eimas et al. [17] established that infants have the ability to discriminate phonemic contrasts of their own languages, as well as contrasts that belong to other languages. In contrast, to take some well-known examples, adult speakers of Japanese have great difficulty distinguishing /l/ from /r/, and Spanish speakers fail to distinguish the French vowels /e/ and /æ/. In cross-sectional and longitudinal studies, Werker and Tees [18] found that infants' ability to discriminate the phonemic contrasts that do not belong to their maternal language (i.e. the language surrounding them), decreases between eight and twelve months of age. Best, McRoberts and Nomathemba [19] have conjectured that this apparent impoverishment occurs only when the foreign contrasts can be assimilated to a category that exists in the child's language. If the contrast has no close alternative in the native language, the ability to discriminate will not fade. Thus, the discrimination of Zulu clicks remains very good for American adults as well as for 12–14 month-old and 8–10 month-old infants. What remains unclear is whether these preserved abilities are similar to the ones that the infant uses when he or she establishes the categories for the maternal language. Finally, Kuhl et al. [20] found that the convergence towards vowel categories starts at an even earlier age. They found evidence that American and Swedish infants have established a prototype for the vowels in their language by six months of age. Thus, the reorganization of speech perception does not result from the learning of a lexicon. Indeed, reorganizations like the ones described above are observed much before infants begin to acquire words.

Currently, an active area of research explores the filtering properties that humans must have to be able to segregate some signals from noise, or from other signals (see [21] and also [22]). The infants' ability to focus on speech rather than noises was explored by Colombo and Bundy [23]. Inputs can be more than just 'noisy', however, they can also include utterances from diverse languages that the learning infant apparently manages to sort out. Just imagine what it would mean to learn English and Japanese without realizing that the utterances correspond to two rather than to one language. Mehler et al. [24] demonstrated that French four-day-old newborns could discriminate French and Russian utterances. This ability remained when these utterances were low-pass filtered (i.e. erasing information above 400 Hz) to remove all phonetic cues and only the prosodic structure (namely global properties of the utterance like rhythm and melody) remained. Moreover, these newborns could discriminate between English and Spanish, two unfamiliar languages. By two months of age, a clear behavioral evolution was noticed: whereas two-month-old infants could discriminate a novel language from their maternal language, they no longer could discriminate two novel languages from each other. This suggests that two-month-old infants have established a template for their own language against which they evaluate all linguistic input.

In more recent experiments, Dehaene-Lambertz (G Dehaene-Lambertz, unpublished data) has replicated these results with two-month-old American infants using short sentences. She also found that when the prosody was disrupted (in lists obtained by scrambling words excised from the original sentences), the infants no longer show a faster orientation to the familiar language stimuli. Although the number of syllables and the length was similar in the clauses and the scrambled word utterances, the destruction of the phrasal prosodies hindered the emergence of a discriminative response. Moon, Panneton-Cooper and Fifer [25] also observed an ability to distinguish native from foreign language sentences in American neonates born from Spanish or English families. So far, all the evidence indicates that the prosodic properties of languages — e.g. durations of phonemes and syllables, and variations in pitch and energy — are extremely important for the early stages of acquisition. A study by Friederici and Wessels [26] (see also [27]) uncovered that infants could not detect familiar phonotactic structures (legal sequences of phonemes versus illegal ones) before the age of nine months.

Prosodic structure is the main property that we invoke to explain the early discrimination abilities in young infants. As suggested by Gleitman and Wanner [28], isolating the relevant prosodic units can help bootstrap lexical and grammatical learning. This prosodic bootstrapping hypothesis was studied by Jusczyk et al. [29], who showed that nine-month-old infants are sensitive to markers of phrasal boundaries: they prefer listening to sentences with pauses artificially inserted between phrasal units, rather than within phrasal units. Six-month-old infants did not seem sensitive to phrase boundaries, but could detect clause boundaries. This experiment involved whole sentences. In studies with much younger infants, Christophe et al. [30] found that newborns could discriminate between two lists of bisyllabic items; in the first list, all the items had been spliced from the middle of long words, whereas in the second list, the items were made of syllables that straddled two words (e.g. [mati] extracted from 'mathématikien' versus from 'pyjama tigré').

Speech processing in the adult

The suggestion we draw from the above studies is that the speech processing system adapts itself to become well suited to the phonology of the natural language one learns as a child. Apparently, speakers of different languages not only master different inventories of phonemes, but also segment the speech stream in a language-specific way. Mehler and colleagues [31,32] and Segui [33] proposed that the syllable plays an central role in the segmentation of speech, and that it is specific to the language. Indeed, French speakers are faster to detect speech fragments that match
the first syllable of words [31,32]. In contrast, English speakers are not affected by the syllabic structure of words [34–36], but are sensitive to the distribution of strong and weak syllables in their language [37]. In addition, Cutler et al. [34,35] found that monolingual speakers of French or English applied their native-language segmentation procedure to a foreign language. The rhythmic structure of English and French was invoked to explain these differences in behavior, that is, whereas French has a syllabic rhythm, English is characterized by a stress rhythm, involving the alternation of strong and weak syllables. Sebastian-Galles et al. [38] have also shown that the syllable plays a role in the segmentation of both Spanish and Catalan. Subsequently, this series of studies was extended to Japanese, a language for which the mora (a subsyllabic unit) had been put forward as the basic rhythmic unit, on phonological grounds [39]; but see [40,41]). Port, Dalby and O'Dell [42] studied speech production in Japanese speakers and found that Japanese speech has a rhythm that is close to mora timing. On the perceptual side, Otake et al. [43] have found that the mora is to Japanese as the syllable is to French.

Let us mention, however, that the above studies only used one task. To broaden the empirical basis of these studies it is necessary to use a greater variety of methods. Pallier et al. [44**], inspired by a method used by Pitt and Samuel [45], found that French and Spanish subjects performing a phoneme-detection task could focus their attention toward a syllabically defined phonemic position inside words. In a task that does not require explicit manipulation of syllables, subjects relied on the syllable to generate their responses. This new tool will also be used to investigate other languages. Another quite different method has involved the use of artificially compressed speech. Using an algorithm to speed up speech without distorting its pitch, a team of researchers investigated how the listener adapted to fast speaking rates [46†]: when subjects listen to very fast sentences (accelerated to more than twice the rate of natural utterances), their comprehension increases steadily during the presentation of 10 sentences. This adaptation takes place even when the person speaking the sentences changes (E Dupoux, K Green, unpublished data), and seems to involve a rather abstract phonological processing stage. One might then expect adaptation in one language to transfer to another language, to the extent that they share common rhythmic properties. Results obtained so far with English, French, Japanese, Spanish and Catalan support this contention: only inside the Romance languages was transfer of adaptation observed.

These new developments in the study of speech-specific processing abilities raise new and interesting questions with respect to bilingualism. Can a child raised with two languages master two independent sets of routines for speech segmentation, or is one language dominant? Cutler et al. [47] found evidence for the latter. They found that even highly proficient bilinguals still have a subjective preference for one language over another, and that this preference correlates with, for example, the propensity of displaying on-line syllabic effects in French, and effects of strong versus weak syllables in English.

Conclusions

The work reported so far suggests that the human brain comes equipped with specific systems to process speech. Moreover, the research on the contrasting fashion in which languages cope with signals is compatible with the idea that languages use a number of units that are hierarchically organized. It complements phonological investigation in that it shows that different languages give a more prominent role to some structures, for example, syllables in French, moras in Japanese and possibly some stress-based unit in English.

This summary of psycholinguistic research illustrates the utility of carrying parallel studies with infants and adults, looking at different languages. In the above studies, we focused on the sound pattern of natural languages. If Pettito and Marentette [48†] are right, however, in saying that sign languages have structural and developmental characteristics very similar to spoken languages, then the properties that are relevant to processing and acquisition are to be viewed as even more abstract than they are being portrayed here.

Acknowledgements

This research was supported by a grant from the Human Frontiers Science Program. C. Pallier was supported by a grant from the Direction General des recherches et Techniques.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest


These authors measured cerebral activation with PET using pairs of noise bursts and pairs of consonant-vowel-consonant syllables. Noise bursts activated the primary auditory cortex, whereas acoustically matched speech syllables activated secondary auditory cortices bilaterally. With syllable stimuli, instructions to discriminate phonetic structure led to increased activity in a part of Broca's area in the left hemisphere. In contrast, pitch discrimination instructions produced activation of the right prefrontal cortex, consistent with the importance of right-hemisphere mechanisms in pitch perception.


Using PET, the authors measured cortical activation and observed a network of activation in the left hemisphere of subjects as they listened to a story in French (their maternal language) but not when listening to a story in Tamil, a language they did not understand.


The authors used ERPs to explore the changes in cortical activation that result from language acquisition. They measured cortical activations for known and unknown words in twenty-month-old infants. They found that in addition to the left hemisphere superiority for linguistic material, there was also an anterior posterior asymmetry.


Using the head-turning technique, the authors showed that discrimination of vowels is already influenced by experience by six months of age: American and Swedish infants noticed a switch toward a prototypical vowel of their language, but not toward a prototypical vowel of the other language, after they had been trained on non-prototypical vowels.


Using the head-turning technique, the authors showed that discrimination of vowels is already influenced by experience by six months of age: American and Swedish infants noticed a switch toward a prototypical vowel of their language, but not toward a prototypical vowel of the other language, after they had been trained on non-prototypical vowels.


The authors report that neonates modify their sucking behavior to trigger more native language reinforcers than foreign language ones. For half of the infants, their maternal language was Spanish and for the other half, it was English.


The age at which an infant can use distributional regularities signaling word boundaries was studied with nine-month-old infants. It was found that young infants are sensitive to these regularities, even when prosodic cues have been removed.


The authors report that six-month-old American infants show a preference for lists of unfamiliar English words as compared to Norwegian words uttered by a single bilingual speaker. However, these infants show no preference for lists of English as opposed to Dutch words. Furthermore, similar behaviors are observed when the lists of words are low-pass filtered, which suggests that the infants are using some properties related to the prosodic structure of the words. The authors stress the fact that the prosodic properties of English and Norwegian words are radically different, whereas English and Dutch words have similar prosodic properties. This study indicates that six-month-old infants have a notion of what is a legal prosodic unit in their language, even though these prosodic units only range over fragments that are two syllables long.


Neonates are able to discriminate lists of bisyllabic items that contain a word boundary from similar lists of bisyllabic items that do not. Moreover, adults learn how to classify these items into distinct categories. These results show that connected speech has cues to prosodic boundaries that are perceptible even to infants and may be used during language acquisition.

Three populations of subjects were asked to detect speech segments in Japanese words. Japanese subjects were more accurate at detecting segments corresponding to the mora, French subjects were faster to detect segments corresponding to syllables, and English subjects were insensitive to both units.