Language-specific listening

Christophe Pallier, Anne Christophe and Jacques Mehler

Languages differ in their phonological structure and psycholinguists have begun to explore the consequences of this fact for speech perception. We review research documenting that listeners attune their perceptual processes finely to exploit the phonological regularities of their native language. As a consequence, these perceptual processes are ill-adapted to listening to languages that do not display such regularities. Thus, not only do late language-learners have trouble speaking a second language, also they do not hear it as native speakers do; worse, they apply their native language listening procedures which may actually interfere with successful processing of the non-native input. We also present data from studies on infants showing that the initial attuning occurs early in life; very young infants are sensitive to the relevant phonological regularities which distinguish different languages, and quickly distinguish the native language of their environment from languages with different regularities.

Speakers of American English, Japanese or Igbo begin life with identical language processing systems: were any of them to be adopted at birth into a different linguistic environment, they would grow up as perfect native speakers of the adoptive language, not of the parental language. All babies are born equipped with the same processing abilities. However, it is also obvious that, as monolingual adults, these speakers cannot understand each other's languages as they do not know the sounds, words or grammatical structures of languages other than their own. Thus, processing must, in part, be different for each language; a given language may use information that is not relevant in another – Igbo distinguishes between words just by changing the tone with which they are spoken, English allows a vowel to occur in a full or a reduced form, Japanese places verbs at the end of a sentence, and so on. To what extent does the understanding of spoken language involve universal characteristics, which are fundamental to the cognitive architecture of the human language system, and to what extent is processing dependent on particular features of specific languages? These are questions that recently have prompted considerable research efforts in the area of psycholinguistics. In this paper we focus on evidence that perception is dependent on the language learned by the listener.

Language-specificity in adult processing

That adult language processing is tailored to a particular language becomes apparent when adults attempt to learn a second language. The flexibility that human listeners are accustomed to displaying – immediately understanding speakers never heard previously, understanding speech against a noisy background, and so on – disappears[1]. Even for those who have attained a high level of competence in the lexical and grammatical structures of a second language, so that perhaps they can read it with relative ease, listening can remain difficult. For instance, there may be a striking deafness to contrasts that are obvious immediately to native speakers. This is easy to demonstrate with phonemic contrasts: thus English listeners have trouble distinguishing between dental and retroflex stop consonants of Hindi, or between uvular and velar stop consonants of the North American language Salish, although native speakers of those languages perceive the distinctions with ease[2,3]. The locus of this failure to discriminate non-native phonetic contrasts has been the target of years of research[4,5]. These efforts have revealed that adults have not necessarily lost the ability to discriminate between pairs of sounds that do not belong to their language: discrimination may occur without acoustic experience if the contrasting phonemes cannot be subsumed by any native categories – thus English listeners can discriminate Zulu clicks. If, however, the foreign sounds are both similar to one phoneme of the native language, then typically discrimination is difficult[4].

Not only do listeners of different languages use different phonemic categories to represent the speech signal, but their perceptual system also exploits knowledge about the constraints on the co-occurrence of these phonemes (i.e. the 'phonotactics'). In Japanese, a nasal consonant followed by a stop consonant will always have the same place of articulation as the stop; thus *tombo* and *kinko* are words, but *tonbo* and *kimko* could not be. In English and Dutch, the same constraint holds in general, but there are many exceptions, especially in prefixed and compound words (*unbearable* and *tomcat* in English; *renbaan* and *imker* in Dutch). Japanese listeners can exploit place of articulation match to speed detection of a stop consonant preceded by a nasal consonant,

C. Pallier, A. Christophe and J. Mehler are in the Laboratoire de Sciences Cognitives et Psycholinguistique, 54 Bd Raspail, 75270 Paris Cedex 06, France. C. Pallier is also in the Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands.

tel: +33 1 49 54 22 76
fax: +33 1 45 44 98 35
e-mail: pallier@lscp.ehess.fr

 1364-6613/97/$17.00 PII: S1364-6613(97)01044-9

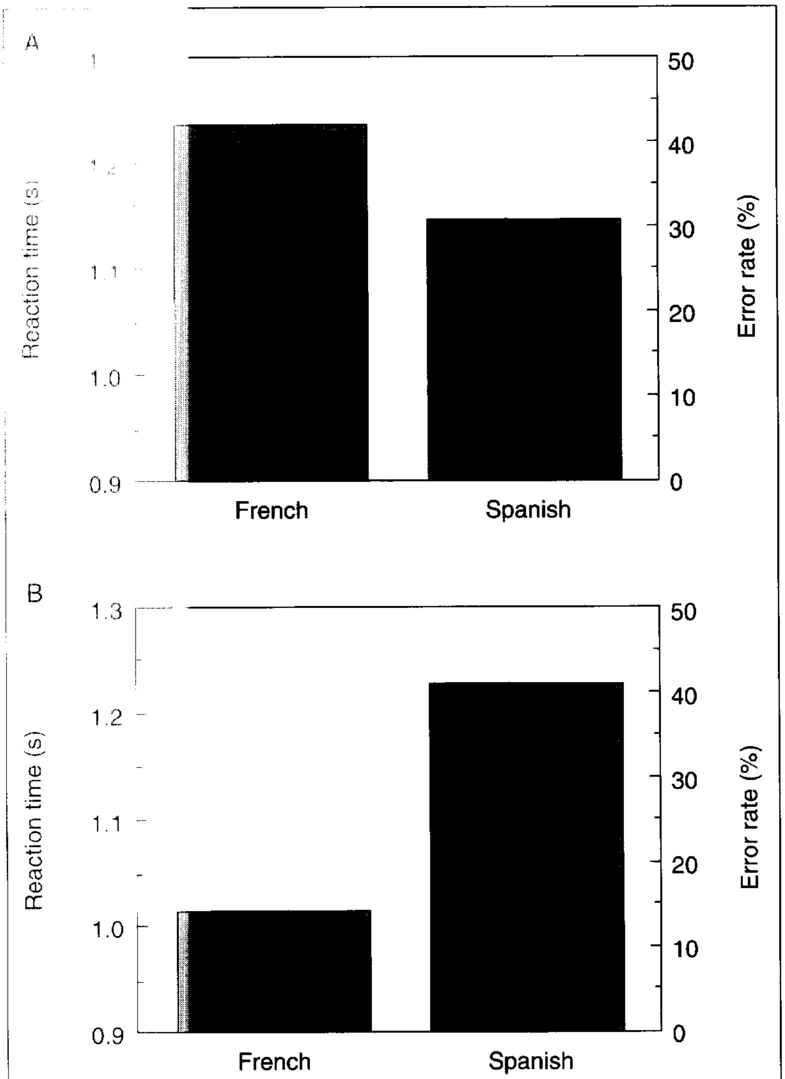

Fig. 1 Reaction times (in grey) and error rates (in black) in ABX discrimination judgments in French and Spanish subjects. (A) Discrimination based only on accent, phonemes fixed (e.g. 'VAsuma', 'vaSUma', 'VAsuma'; correct response: first item). **(B)** Discrimination based only on phonemes, with irrelevant variations in accent (e.g. 'VAsuma', 'faSUma', 'vaSUma': correct response: first item). These data show that Spanish subjects are more sensitive to accent variations than French subjects.

but Dutch listeners do not use this information[6]. Not only can speakers use phonotactics to predict the upcoming phoneme; in their attempts to build up a representation of the speech signal that follows the patterns of the native language, speakers can even insert 'illusory' phonemes when a stimulus does not conform to this pattern. Thus, speakers of Japenese (a language that does not allows word-internal obstruent clusters) have a lot of trouble discriminating between VCCV and VCVCV (V, vowel; C, consonant) sequences[7,8].

Some levels of phonological organization such as stress require longer stretches of speech in order to be extracted. In English, words such as *insight* and *incite*, or, in Spanish, *'bebe* and *be'be*, contrast only in stress. Native speakers of these languages have no difficulty telling such stress pairs apart; but Dupoux *et al.*[9] demonstrated that speakers of French, a language that does not have stress contrasts between words, may ignore stress contrasts entirely. In an ABX discrimination paradigm, in which subjects judged whether the last of three nonsense items (pronounced by native speakers of Dutch) most resembled the first or the second item, French listeners performed significantly worse than Spanish listeners when the decision was based on stress (*bope'lo, bo'pelo, bo'pelo*), but significantly better than Spanish listeners when the decision was based on segmental structure and required that stress variation be ignored (*sope'lo, bo'pelo, bope'lo*; see Fig. 1). French listeners' 'deafness' to stress is not due to unfamiliarity with the acoustic contrast per se, since inter-syllable differences in accent do occur in French. However, in French, such differences never distinguish one word from another; in consequence, listeners may ignore them.

Speakers of Spanish and French show similar sensitivity to the syllabic structure of utterances in various psycholinguistic tasks[10–13], but speakers of Japanese are sensitive to another unit: they automatically group phonemes into morae – subsyllabic units consisting of a vowel, a CV or a syllable-final consonant[6,14,15]. Importantly, these studies showed that listeners parse foreign language input using their native units. For example, French listeners segment Japanese in terms of syllables[15], while Japanese listeners impose a moraic structure on English, French and Spanish words[14,16].

Another dimension in which languages differ concerns cues to word boundaries: speakers do not pause between words or otherwise provide definitive cues to word boundaries, yet listeners nevertheless have the impression of hearing speech as a sequence of individual words. The way that this efficient segmentation occurs also differs across languages. Thus, in English and Dutch, most words begin with strong syllables (syllables containing an unreduced vowel)[17,18] and, indeed, listeners treat strong syllables as likely to be word-initial and weak syllables as likely to be word-internal[19–21]. Such a strategy is simply not available in languages where the strong-weak distinction is not used. Similarly, Finnish listeners exploit vowel harmony in speech segmentation[22], but obviously this is not possible in languages without vowel harmony. We expect that whatever their language, listeners exploit the regularities in phonology and lexicon in order to help segmentation.

Finally, the evidence combines to show that listening itself is highly language-specific. Partly, this is inevitable, simply because languages differ in the type of information that they provide. But the effects are more far-reaching. With non-native languages, foreign accent in production appears to have a direct perceptual equivalent: listeners employ their native phonological processing procedures, and when these are inappropriate for the structure of the foreign language, listening is difficult.

The development of language specificity

How does it come about that adult language users, who started with the same infant processing abilities, end up with processing routines tailored so exquisitely to the requirements of the native language that they actually interfere with the processing of non-native tongues? Assuming that babies are born equipped with constraints on what a human language can be[23,24], one has to explain how they learn their mother tongue. The current view is that when listening to speech signals, be they native or foreign, infants

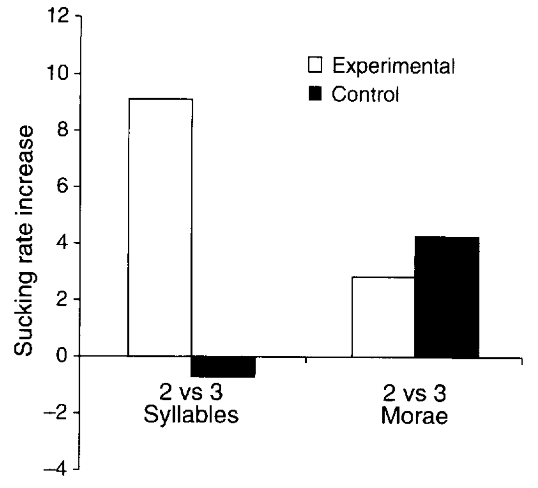

Fig. 2 Discrimination by French newborn babies of lists of di- versus trisyllabic items, and lists of di- versus trimoraic items. Discrimination is attested by a larger increase in sucking rate in the experimental group (change stimuli and syllable/morae number) than in the control group (change stimuli only). This is statistically significant in the case of stimuli differing in number of syllables, but not for stimuli differing in number of morae (the Japanese rhythmic unit, see text).

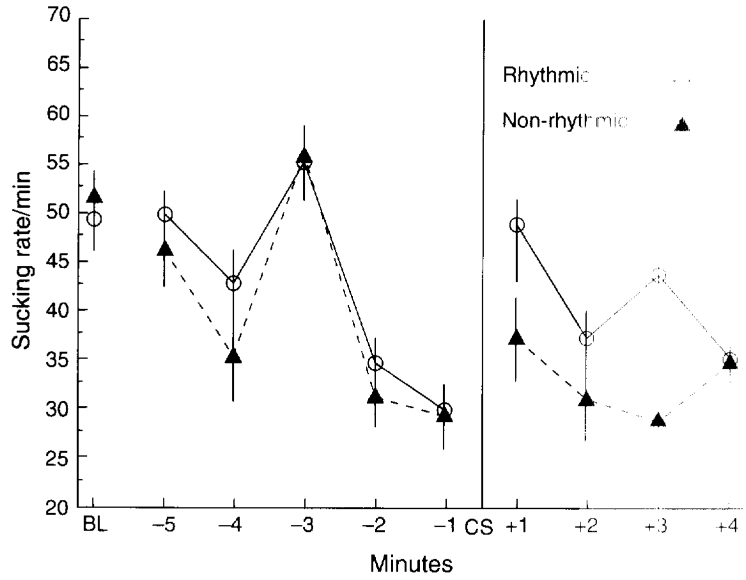

Fig. 3 Mean sucking rate in a non-nutritive sucking experiment with 32 French newborn babies. Measurements were made during the baseline period (BL), five minutes before, and four minutes after the change in stimulation (CS). The rhythmic group was switched from a mixture of sentences taken from two stress-timed languages (Dutch and English) to a mixture of sentences from two syllable-timed languages (Spanish and Italian), or vice versa. The non-rhythmic group also changed languages, but in each phase of the experiment there were sentences from one stress-timed and one syllable-timed language (e.g. Spanish and English, then Italian and Dutch). Infants from the rhythmic group reacted significantly more to the change of stimulation than infants from the non-rhythmic group.

represent all the features necessary to process any of the world's language (e.g. stress, vowel length, moraic structure, complex syllabic structure, tone and so on).

During the first year, when infants are exposed to their mother tongue, they will stop using features that are not relevant to this language. This has been amply documented for the perception of phonemes: babies start off with a universal phonetic inventory that allows them to perceive any phonetic contrast from any of the world's language so far tested[4,5]. Between six and 12 months, their phonetic perception increasingly becomes similar to that of the adults from their linguistic environment[5,25,26].

Bertoncini and her colleagues[27] have explored another aspect of speech perception in newborns: they compared newborn babies' perception of items that varied in number of syllables and/or morae. They have shown that French newborn babies can discriminate between lists of phonetically varied words on the basis of the number of syllables[28] (two versus three; see Fig. 2), but not on the basis of the number of morae (either two as in *iga*, or three as in *iNga* or *iiga*). As mentioned above, French adults rely on the syllable while the mora is more salient for the Japanese. Two interpretations are available: the first one is that babies learn this characteristic of their mother tongue very rapidly (had we tested Japanese infants the results might have been different); the second possibility is that syllable-like units are universally more salient at birth. This is an important empirical question for future research.

On the basis of this result, among others, Mehler *et al.*[29] have proposed that initially babies pay attention mostly to the sequence of vowels in the speech stream. Languages with different rhythmic properties can be distinguished on such a representation. Thus, languages that share rhythmic properties may more readily be confused by babies than languages differing in this dimension. Recent work by Nazzi and his colleagues[30] provides initial confirmation for this hypothesis: they have shown that newborn infants tend to neglect the difference between two languages with similar rhythmic properties. Thus, French newborns fail to discriminate low-pass filtered English sentences from low-pass filtered Dutch sentences while they are perfectly able to discriminate between English and Japanese filtered sentences. Furthermore, when newborns are habituated with a set of sentences drawn from rhythmically close languages (e.g. Dutch and English), they notice the change to new sentences drawn from another rhythmic family (e.g. Italian and Spanish sentences). In contrast, they do not react to a change from, say, a mixture of Dutch and Italian sentences (two languages with different rhythms) to a mixture of English and Spanish sentences (see Fig. 3).

Once babies have established what features are relevant to represent speech, they can start using this representation to discover regularities about their native language. In particular, we have seen that adults exploit language-specific strategies to segment continuous speech into words. The literature suggests that by the age of nine months, babies have already discovered at least some of the regularities that form the basis of these strategies. Thus, nine-month-old American babies were shown to listen longer to lists of strong-weak words such as *beaver* (the most frequent pattern in English) than to lists of weak-strong words such as *abeam*[31]. This implies that from the age of nine months, English-speaking babies may, just like English-speaking adults, use this regularity of English to hypothesize word boundaries in the continuous speech stream.

Similarly, knowledge about the co-occurrence of phonemes (phonotactics) may provide powerful cues to the presence of word boundaries (e.g. there has to be a word boundary between the *d* and the *s* in *dstr*, as in *bad string*). Nine-month-old Dutch babies prefer to listen to lists of Dutch syllables that respect the phonotactics of Dutch (e.g. *bref, murt*) rather than to lists of impossible syllables in Dutch (e.g. *febr, rtum*)[32]. When Dutch and American nine-month-old babies are played lists of Dutch and American words that differ only in their phonotactics, they prefer to listen to the words from their native language (e.g. Dutch but not English allows 'vl' word-initial clusters such as in *vlammend*; English but not Dutch allows a word-final voiced consonant such as in *hubbub*[33]). American nine-month-old babies also prefer to listen to lists of English monosyllables that contain frequent rather than infrequent phonetic patterns[34]. Most of these findings do not hold true when six-month-old babies are tested, indicating that this learning occurred at some point between six and nine months of age.

Conclusions

So far, we have only been able to present studies relevant to sound patterns. We anticipate that similar studies will appear on other aspects of language processing such as morphology, syntax and possibly even semantics. We have reviewed a number of studies that illustrate the importance of language-specific procedures and representations. We have also shown when some of these language-specific devices are acquired.

A number of important issues remain to be explored by future research. For instance, currently we are investigating whether bilinguals can master equally well the specific processing routines that correspond to the two languages. Earlier research[35] indicates that bilinguals have a dominant processing routine (corresponding to one of the languages). We do not know whether early and equivalent exposure to two languages can produce two routines (one for each language mastered), each similar to that used by monolinguals. Also, we are exploring whether the cortical zones that mediate language processing in monolinguals are the same as those involved when processing either one of the languages of a bilingual. These and other issues are crucial to understanding the constraints involved in language usage and language acquisition.

References

1 Takata, Y. and Nabelek, A.K. (1990) English consonant recognition in noise and in reverberation by Japanese and American listeners *J. Acoust. Soc. Am.* 88, 663–666

2 Polka, L. (1992) Characterizing the influence of native language experience on adult speech perception *Percept. Psychophys.* 52, 37–52

3 Werker, J.F. and Lalonde, C.E. (1988) Cross-language speech perception: initial capabilities and developmental change *Dev. Psychol.* 24, 672–683

4 Best, C.T. (1994) The emergence of native-language phonological influence in infants: a perceptual assimilation model, in *The Development of Speech Perception: the Transition from Speech Sounds to Spoken Words* (Goodman, J.C. and Nusbaum, H.C., eds), pp. 167–224, MIT Press

5 Werker, J.F. (1994) Cross-language speech perception: developmental change does not involve loss, in *The Development of Speech Perception: the Transition from Speech Sounds to Spoken Words* (Goodman, J.C. and Nusbaum, H.C., eds), pp. 93–120, MIT Press

6 Otake, T. *et al.* (1996) The representation of Japanese moraic nasals *J. Acoust. Soc. Am.* 100, 3831–3842

7 Mehler, J., Pallier, C. and Christophe, A. Language and cognition, in *Cognitive and Biological Aspects* (Sabourin, M., Craik, F.M.I. and Roberts, M., eds), Erlbaum (in press)

8 Kakehi, K., Kato, K. and Kashino, M. (1996) Phoneme/syllable perception and the temporal structure of speech, in *Phonological Structure and Language Processing: Cross-linguistic Studies* (Otake, T. and Cutler, A., eds), pp. 125–143, Mouton de Gruyter

9 Dupoux, E. *et al.* (1997) A destressing 'deafness' in French? *J. Mem. Lang.* 36, 406–421

10 Cutler, A. *et al.* (1992) The monolingual nature of speech segmentation by bilinguals *Cogn. Psychol.* 24, 381–410

11 Kolinsky, R., Morais, J. and Cluytens, M. (1995) Intermediate representations in spoken word recognition: evidence from word illusions *J. Mem. Lang.* 34, 19–40

12 Pallier, C. *et al.* (1993) Attentional allocation within syllabic structure of spoken words *J. Mem. Lang.* 32, 373–389

13 Sebastian-Gallés, N. *et al.* (1992) Contrasting syllabic effects in Catalan and Spanish *J. Mem. Lang.* 31, 18–32

14 Cutler, A. and Otake, T. (1994) Mora or phoneme? Further evidence for language-specific listening *J. Mem. Lang.* 33, 824–844

15 Otake, T. *et al.* (1993) Mora or syllable? Speech segmentation in Japanese *J. Mem. Lang.* 32, 258–278

16 Otake, T., Hatano, G. and Yoneyama, K. (1996) Speech segmentation by Japanese listeners, in *Phonological Structure and Language Processing: Cross-linguistic Studies* (Otake, T. and Cutler, A., eds), pp. 183–201, Mouton de Gruyter

17 McQueen, J.M. *et al.* (1995) Models of continuous speech recognition and the contents of the vocabluary *Lang. Cognit. Process.* 10, 309–331

18 Schreuder, R. and Baayen, R.H. (1994) Prefix stripping re-revisited *J. Mem. Lang.* 33, 357–375

19 Cutler, A. and Butterfield, S. (1992) Rhythmic cues to speech segmentation: evidence from juncture misperception *J. Mem. Lang.* 31, 218–236

20 Norris, D.G., McQueen, J.M. and Cutler, A. (1995) Competition and segmentation in spoken word recognition *J. Exp. Psychol. Learn. Mem. Cognit.* 10, 309–331

21 Vroomen, J., van Zon, M. and de Gelder, B. (1996) Cues to speech segmentation: evidence from juncture misperceptions and word spotting *Mem. Cognit.* 24, 744–755

22 Suomi, K., McQueen, J.M. and Cutler, A. (1997) Vowel harmony and speech segmentation in Finnish *J. Mem. Lang.* 36, 422–444

23 Pinker, S. (1994) *The Language Instinct*, Morrow

24 Mehler, J. and Dupoux, E. (1994) *What Infants Know*, Blackwell

25 Trehub, S.E. (1976) The discrimination of foreign speech contrasts by infants and adults *Child Dev.* 47, 466–472

26 Kuhl, P. *et al.* (1992) Linguistic experience alters phonetic perception in infants by 6 months of age *Science* 255, 606–608

27 Bertoncini, J. *et al.* (1995) Morae and syllables: rhythmical basis of speech representations in neonates *Lang. Speech* 38, 311–329

28 Bijeljac-Babic, R., Bertoncini, J. and Mehler, J. (1993) How do four-day-old infants categorize multisyllabic utterances? *Dev. Psychol.* 29, 711–721

29 Mehler, J. *et al.* (1996) Coping with linguistic diversity: the infant's viewpoint, in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Morgan, J.L. and Demuth, K., eds), pp. 101–116, Lawrence Erlbaum

30 Nazzi, T., Bertoncini, J. and Mehler, J. Language discrimination by newborns: towards an understanding of the role of rhythm *J. Exp. Psychol. Hum. Percept. Perform.* (in press)

31 Jusczyk, P.W., Cutler, A. and Redanz, N.J. (1993) Infants' preference for the predominant stress patterns of English words *Child Dev.* 64, 675–687

32 Friederici, A.D. and Wessels, J.M.I. (1993) Phonotactic knowledge of word boundaries and its use in infant speech-perception *Percept. Psychophys.* 54, 287–295

33 Jusczyk, P.W. *et al.* (1993) Infants' sensitivity to the sound pattern of native language words *J. Mem. Lang.* 32, 402–420

34 Jusczyk, P.W., Luce, P.A. and Charles-Luce, J. (1994) Infants' sensitivity to phonotactic patterns in the native language *J. Mem. Lang.* 33, 630–645

35 Cutler, A. *et al.* (1989) Limits on bilinguism *Nature* 340, 229–230