HOW HUMAN SPEECH RECOGNITION IS AFFECTED BY PHONOLOGICAL DIVERSITY AMONG LANGUAGES

Anne Cutler

MPI for Psycholinguistics, Wundtlaan 1, NL-6525 XD Nijmegen.

ABSTRACT - Listeners process spoken language in ways which are adapted to the phonological structure of their native language. As a consequence, non-native speakers do not listen to a language in the same way as native speakers; moreover, listeners may use their native language listening procedures inappropriately with foreign input. With sufficient experience, however, it may be possible to inhibit this latter (counter-productive) behaviour.

INTRODUCTION

When we listen to speech, we recognise the words it contains. We cannot recognise spoken utterances as undivided wholes, because it would be impossible to store every complete utterance we might ever hear. Therefore listening to speech involves finding the individual words in the speech signal input, and matching them to stored lexical entries in order to determine their meaning. But speech is continuous: in most spoken language, few cues are available to signal reliably where one word ends and the next begins, so that segmentation of utterances into words cannot be achieved by relying on simple cues in the signal.

The subjective experience of listening to speech (in our native language), however, it that it is very easy - we just seem to hear one word after another. Of course, what *does* seem hard is listening to speech in a foreign language - even when we know the words of a language well (for instance, when reading it seems easy), it is often difficult to comprehend spoken utterances. Here subjective experience suggests that foreign utterances are often spoken very fast; certainly the clear impression of one word after another, offered by our native language, seems not to be present.

This paper summarises a series of studies which have addressed the questions of how listeners achieve segmentation so easily, and why listening to a foreign language seems not just difficult but different. These studies suggest that listeners effectively exploit the phonological structure of language to achieve speech segmentation; because languages are phonologically diverse, speakers of different languages come to rely upon different listening procedures. This in turn makes listening to foreign languages, to which one's native listening procedures may be ill-suited, in certain circumstances difficult.

SEGMENTATION OF ENGLISH

Experiments in English have suggested that listeners segment speech at the onset of each strong syllable (i.e. each syllable containing a full vowel). For example, finding a real word in a spoken nonsense sequence is hard if the word is spread over two strong syllables (e.g. *mint* in [mɪntef]) but easier if the word is spread over a strong and a following weak syllable (e.g. *mint* in [mɪntəf]; Cutler & Norris, 1988; McQueen, Norris & Cutler, 1994). This finding suggests that listeners divide [mɪntef] at the onset of the second strong syllable, so that detecting *mint* requires recombination of speech material across a segmentation point; [mɪntəf] on the other hand offers no such obstacles to detection of *mint*, as the second syllable is weak and so the sequence is simply not divided. Similarly, when English-speakers make slips of the ear which involve mistakes in word boundary placement, they tend most often to insert boundaries before strong syllables (e.g. hearing *by loose analogy* as *by Luce and Allergy*) or delete boundaries before weak syllables (e.g. hearing *how big is it*? as *how bigoted*?; Cutler & Butterfield, 1992). This so-called "Metrical Segmentation Strategy" for English (Cutler, 1990) is efficient in that most English lexical words do indeed begin with strong syllables, and most strong syllables in typical utterances are indeed word-initial (Cutler & Carter, 1987).

LANGUAGE-SPECIFIC SEGMENTATION

The Metrical Segmentation Strategy for English is founded on the opposition between strong and weak syllables which is an important feature of English phonology. Other languages, however, have quite different phonologies; thus the English procedure may not work at all for other languages. Many other languages, for example, do not have English-like stress, so that no opposition between strong and weak syllables exists. French, for instance, is one such language, and Japanese another. Experimental evidence concerning segmentation procedures exists for both these languages.

In French, evidence from a wide variety of experimental tasks suggests that listeners segment speech into syllable-sized units (Mehler, Dommergues, Frauenfelder & Seguí, 1981; Seguí, Frauenfelder & Mehler, 1981; Cutler, Mehler, Norris & Seguí, 1986; Pallier, Sebastián-Gallés, Felguera, Christophe & Mehler, 1993). Confirming evidence suggests that syllabic segmentation can be observed under certain conditions in other languages also - for instance, in Spanish (Sebastián-Gallés, Dupoux, Seguí & Mehler, 1992; Bradley, Sánchez-Casas & García-Albea, 1993) and in Catalan (Sebastián-Gallés et al., 1992).

Syllabic segmentation is by no means the same process as the stress-based segmentation proposed, in the form of the Metrical Segmentation Strategy, for English. Yet there is a sense in which the procedures which have been experimentally demonstrated for English and for French are closely parallel. Both stress in English and the syllable in French are the basis of rhythmic structure in their respective languages. This led Cutler, Mehler, Norris and Seguí (1992) to propose that listeners might in fact adopt a universally applicable solution to the word boundary problem, in that to solve it they exploit whatever rhythmic structure happens to characterise their language.

This hypothesis could be tested against a language with a rhythm characterised neither by stress nor by syllabic regularity. Japanese is such a language; its rhythm is described in terms of a sub-syllabic unit, the mora. A mora can be a CV structure, or a single vowel, or a syllabic coda; thus *Honda*, for example, has three morae: *ho-n-da*. Experimental studies using perception tasks analogous to those used in experiments with English and French listeners indeed produced evidence of mora-based segmentation by Japanese listeners (Otake, Hatano, Cutler & Mehler, 1993; Cutler & Otake, 1994).

MORE LANGUAGE-SPECIFIC LISTENING

Thus the evidence from English, French and Japanese suggests that speakers of different languages have different procedures available to them for solving the speech segmentation problem. This means that not only do languages differ in the obvious ways - different repertoires of sounds, different words, different grammars - but they also differ in the ways that their native speakers listen to them. The segmentation of continuous speech input into words is, furthermore, not the only aspect of human speech recognition which shows language-specific features. Even something as (apparently) simple as detecting a phoneme can differ across languages.

Thus English listeners detect vowel targets more slowly and less accurately than stop consonant targets, even when the former are highly distinct (/a/, /i/) while the latter are confusable (/p/, /t/; van Ooyen, Cutler & Norris, 1991); vowels are also detected more slowly than fricatives (Norris, van Ooyen & Cutler, 1992). In contrast, Spanish subjects listening to Spanish words show no significant difference between detection of vowels and of stop consonants (van Ooyen & Sánchez-Casas, 1993). The vowel repertoires of English and Spanish are, of course, very different: English has many, highly confusable vowels, while Spanish has only five vowels, which occupy distinct positions in vowel space. Also, dialect distinctions are signalled primarily by vowel quality in English but not in Spanish. Japanese resembles Spanish more closely than it resembles English: Japanese, too, has five highly distinct vowels, and dialect distinctions are signalled less by vowel quality than they are in English. As in Spanish, no vowel/consonant differences in phoneme detection appear in Japanese. Recall, however, that the mora in Japanese can be a single vowel or a single syllable-final consonant; and indeed, moraic effects appear in phoneme detection in Japanese. Both vowel targets and consonant targets are detected more rapidly and more accurately if they correspond to a mora (e.g. *O* in *aoki, taoru; N* in *enka, kinri*) than if they do not (e.g. *O* in *kokage, itoku; N* in *enoki, kanojo*; Cutler & Otake, 1994).

THE LOCUS OF LANGUAGE-SPECIFICITY

There is thus considerable evidence that native speakers listening to their own language use language-specific listening procedures. Note, however, that the locus of language-specificity is not in (the phonological structure of) the input. For instance, the presence of a particular rhythmic structure in the input does not of itself produce segmentation based on that structure: English listeners show no evidence of syllabic segmentation with French input, for example (Cutler et al., 1986), and neither do Japanese listeners (Otake, 1992); English listeners likewise show no evidence of mora-based segmentation of Japanese input (Otake et al., 1993; Cutler & Otake, 1994), and nor do French listeners (Otake et al., 1993). The segmentation procedures are, instead, part of the processing repertoire of the listener rather than an input-driven phenomenon. Non-native speakers do not seem to be able to use the same procedures as native speakers.

Moreover, non-native speakers do not seem to be able in some cases to switch off the listening procedures they use for their native language. Indeed, they apply their native language-specific procedures even to foreign language input which is phonologically quite diufferent from the native language (so that the native procedures may work very inefficiently). Thus French listeners apply syllabic segmentation to English input (Cutler et al., 1986) and to Japanese input (Otake et al., 1993), and Japanese listeners apply moraic segmentation where possible to English input (Cutler & Otake, 1994).

Using a native procedure when listening to a foreign language, even though the native procedure is inappropriate for the foreign language, is unlikely to be the best means of processing spoken language; thus it is reasonable to suppose that it may in part be responsible for the inordinate difficulty of listening to foreign languages - for instance, difficulty with listening to a language which can be *read* quite fluently.

LIMITING LANGUAGE-SPECIFIC LISTENING

To determine whether it is possible for one listener to use more than one procedure, Cutler, Mehler, Norris and Seguí (1992) studied a group of balanced English-French bilinguals - i.e. speakers who were equally in command of both languages to indistinguishable native levels. These speakers, they found, commanded only one of the two segmentation procedures specific respectively to English and to French - either syllabic segmentation or stress-based segmentation, but not both. A measure of language preference determined which procedure was available - if on this measure a subject was classed as "English-dominant", he or she used stress-based segmentation with English but did not use syllabic segmentation with French. If a subject was "French-dominant", then syllabic segmentation was used with French but stress-based segmentation was not used with English. This suggested that it is possible to command only one such procedure (e.g., only one way to exploit speech rhythm in segmentation).

However, one further aspect of Cutler et al.'s results has interesting implications. The bilingual French-English speakers who were "French-dominant", and showed monolingual-French response patterns for syllable detection in French, did not show monolingual-French response patterns for syllable detection in English. Instead, they showed no effect at all of the syllable structure of the words. Thus although listeners normally use their native language procedures when listening to a foreign language, even when these procedures are inappropriate for the foreign input, these bilinguals did not use the inappropriate procedure. Cutler et al. suggested that they had learned to inhibit the inappropriate application of the syllabic segmentation procedure, as a result of accrued experience of its inefficiency with English input. This suggests that the counter-productive use of native listening procedures with ill-suited non-native input may be something which a listener can, with sufficient experience, unlearn. It remains to be determined how much experience - along the continuum from the fleeting acquaintance of the new second-language learner to the essentially native experience of the balanced bilingual - counts as sufficient for this purpose.

ACKNOWLEDGEMENTS

The work reported here was supported by the Human Frontier Scientific Program, and was carried out in collaboration with Dennis Norris and Brit van Ooyen (MRC-APU, Cambridge), Jacques Mehler and Juan Seguí (CNRS, Paris), Takashi Otake and Giyoo Hatano (Dokkyo University) and Rosa Sánchez-Casas (St Louis University, Madrid), to all of whom sincere thanks are due.

REFERENCES

Bradley, D.C., Sánchez-Casas, R.M. & García-Albea, J.E. (1993) *The status of the syllable in the perception of Spanish and English.* Language & Cognitive Processes, 8, 197-233.

Cutler, A. (1990) *Exploiting prosodic probabilities in speech segmentation* In G. Altmann (Ed.) Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives. Cambridge, MA: MIT Press; 105-121.

Cutler, A. & Butterfield, S. (1992) *Rhythmic cues to speech segmentation: Evidence from juncture misperception* Journal of Memory & Language, 31, 218-236.

Cutler, A. & Carter, D.M. (1987) *The predominance of strong initial syllables in the English vocabulary.* Computer Speech & Language, 2, 133-142.

Cutler, A., Mehler, J., Norris, D.G. & Seguí, J. (1986) *The syllable's differing role in the segmentation of French and English.* Journal of Memory & Language, 25, 385-400.

Cutler, A., Mehler, J., Norris, D. and Seguí, J. (1992) *The monolingual nature of speech segmentation by bilinguals.* Cognitive Psychology, 24, 381-410.

Cutler, A. & Norris, D.G. (1988) *The role of strong syllables in segmentation for lexical access.* Journal of Experimental Psychology: Human Perception & Performance, 14, 113-121.

Cutler, A. & Otake, T. (1994) *Mora or phoneme? Further evidence for language-specific listening.* Journal of Memory & Language, 33.

McQueen, J.M., Norris, D.G. & Cutler, A. (1994) *Competition in spoken word recognition: Spotting words in other words.* Journal of Experimental Psychology: Learning, Memory & Cognition, 20, 621-638.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Seguí, J. (1981) *The syllable's role in speech segmentation.* Journal of Verbal Learning & Verbal Behavior, 20, 298-305.

Norris, D.G., van Ooyen, B. & Cutler, A. (1992) *Speeded detection of vowels and steady-state consonants.* Proceedings of the Second International Conference on Spoken Language Processing, Banff, Canada; Vol. 2, 1055-1058.

van Ooyen, B., Cutler, A. & Norris, D. (1991) *Detection times for vowels versus consonants.* Proceedings of EUROSPEECH 91, Genoa; Vol. 3, 1451-1454.

van Ooyen, B. & Sánchez-Casas, R.M. (1993) *A cross-linguistic difference in phoneme detection.* Paper presented to the Experimental Psychology Society, Cambridge.

Otake, T. (1992) *Morae and syllables in the segmentation of Japanese.* Paper presented to the XXV International Congress of Psychology, Brussels.

Otake, T., Hatano, G., Cutler, A. & Mehler, J. (1993) *Mora or syllable? Speech segmentation in Japanese.* Journal of Memory & Language, 32, 358-378.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A. & Mehler, J. (1993) *Attentional allocation within the syllabic structure of spoken words.* Journal of Memory & Language, 32, 373-389.

Sebastián-Gallés, N., Dupoux, E., Seguí, J. & Mehler, J. (1992) *Contrasting syllabic effects in Catalan and Spanish.* Journal of Memory & Language, 31, 18-32.

Seguí, J., Frauenfelder, U.H. & Mehler, J. (1981) *Phoneme monitoring, syllable monitoring and lexical access.* British Journal of Psychology 72, 471-477.