# The MIT Encyclopedia of Communication Disorders

Edited by Raymond D. Kent

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

## Segmentation of Spoken Language by Normal Adult Listeners

Listening to spoken language usually seems effortless, but the processes involved are complex. A continuous acoustic signal must be translated into meaning so that the listener can understand the speaker's intent. The mapping of sound to meaning proceeds via the lexicon—our store of known words. Any utterance we hear may be novel to us, but the words it contains are familiar, and to understand the utterance we must therefore identify the words of which it is composed.

We know a great many words; an educated adult's vocabulary has been estimated at around 150,000 words. Entries in the mental lexicon may include, besides stand-alone words, grammatical morphemes such as prefixes and suffixes and multiword phrases such as idioms and cliches. Languages also differ widely in how they construct word forms, and this too will affect what is stored in the lexicon. But in any language, listening involves mapping the acoustic signal onto stored meanings.

The continuity of utterances means that boundaries between individual words in speech are not overtly marked. Speakers do not pause between words but run them into one another. The problem of segmenting a speech signal into words is compounded by the fact that words themselves are not highly distinctive. All the words we know are constructed of just a handful of different sounds; on average, the phonetic repertoire of a language contains 30-40 contrasting sounds (Maddieson, 1984). As a consequence, words inevitably resemble other words, and may have other words embedded within them (thus *strange* contains *stray, strain, train, rain,* and *range).* Word recognition therefore involves identifying the correct form among a large number of similar forms, in a stream in which they abut one another without a break *(strange act* contains *jack* and *jacked).*

The only segmentation that is logically required is to find the words in speech. Whether listening also involves some intermediate level of coding is an issue of contention among speech researchers. Do listeners extract whole syllables from the speech stream and use this syllabic representation to contact the lexicon? Do they extract phonemes from the input, so that listening involves an intermediate stage in which heard utterances are represented as strings of phonemes? Or does listening involve matching speech input against holistic stored forms? The available evidence does not yet allow us to distinguish among these positions (and other variants).

There is agreement, however, on other aspects of the spoken-word recognition process. First, information in the signal is evaluated continuously and the results are passed to the lexicon. Coarticulatory effects that cause cues to adjacent phonemes to overlap in time are efficiently used. Thus *robe, rope, wrote, road,* and *rogue* all begin with *ro-*. but the vowel will in each case include anticipatory information about the place of articulation of the following consonant, and listeners can exploit this (e.g.. to narrow the field of candidates to only *rope* and *robe,* eliminating *rogue, road,* and *wrote).*

Evidence for continuous evaluation comes from experiments in which listeners perform lexical decision (judging whether a spoken string is indeed a real word) on speech that has been cross-spliced so that the coarticulatory effects are no longer reliable. Thus, when listeners hear *troot* they should respond "no" *-troot* is not a word. If *troot* is cross-spliced so that a final *-t* is appended to a *troo-* from either *trook* or *troop* (which give coarticulatory cues to an upcoming velar or bilabial consonant, respectively), then responses are slower than if the cues match. This shows that listeners are sensitive to the coarticulatory mismatch and must have processed the consonant place cues in the vowel. However, the responses are still slower when the mismatching *troo-* comes from *troop* than when it comes from *trook.* This suggests that the processing of consonant cues in the vowel has caused activation of the existing compatible real-word *troop* (Marslen-Wilson and Warren, 1994; McQueen, Norris, and Cutler, 1999).

Second, multiple candidate words are simultaneously activated during the listening process, including words that are merely accidentally present in a speech signal. Thus, hearing *strange-acting* may activate *stray, train, range, jack,* and so on, as well as the intended words.

Evidence for multiple activation comes from cross-modal priming experiments in which a word-initial fragment facilitates recognition of different words that it might become. Thus, lexical decision responses for visually presented "captain" or "captive" are both facilitated when listeners have just heard the fragment *capt-* (compared with some other control fragment). Moreover, both are facilitated even if only one of them matches the context (Zwitserlood, 1989).

Third, there is active competition between alternative candidate words. The more active a candidate word is, the more it may suppress its rivals, and the more competitors a word has, the more suppression it may undergo. Evidence for competition between simultaneously activated candidate words comes from experiments in which listeners must spot any real words occurring in spoken nonsense strings. If the rest of the string partially activates a competitor word, then spotting the real embedded word is slowed. For instance, listeners spot *mess* less rapidly in *domess* (which partially activates *domestic,* a competitor for the same portion of the signal that supports *mess)* than in *nemess* (which supports no other word; McQueen, Norris, and Cutler, 1994; see also Norris, McQueen, and Cutler, 1995; Vroomen and de Gelder, 1995; Soto-Faraco, Sebastian-Galles, and Cutler, 2001).

Because activated and competing words need not be aligned with one another, the competition process offers a potential means of segmenting the utterance. Thus, although recognition of *strange-acting* may involve competition from *stray, range, jack,* and so on, this will eventually yield to joint inhibition from the two intended words, which receive greater support from the signal.

Adult listeners can also use information which their linguistic experience suggests to be correlated with the presence of a word boundary. For instance, in English the phoneme sequence [mg] never occurs word-internally, so the occurrence of this sequence must imply a word boundary *(some go, tame goose);* sequences such as [pf] or [ml] or [zw] never occur syllable-internally, so this sequence implies at least a syllable boundary *(cupful, seemly, beeswax).* Listeners more rapidly spot embedded words whose edges are aligned with such a boundary-correlated sequence (e.g., *rock* is spotted more easily in *foomrock* than in *foogrock;* McQueen, 1998). Also, words that begin with a common phoneme sequence are easier to extract from a preceding context than words that begin with an infrequent sequence (e.g., in *golnook* versus *golnag,* it will be easier to spot *nag,* which shares its beginning with *natural, navigate, narrow, nap,* and many other words; van der Lugt, 2001; see also Cairns et al., 1997).

These latter sources of information are, of course, necessarily language-specific. It is a characteristic of a particular vocabulary that more words begin with the *na-* of *nag* than with the *noo-* of *nook;* likewise, it is vocabulary-specific that sequences such as [pf] or [zw] or [ml] cannot occur within a syllable. Each of these three sequences is in fact legitimately syllable-internal in some language ([pf], for instance, in German: *Pferd, Kopf).*

Other language-specific information is also used in segmentation, notably rhythmic structure. In languages such as English and Dutch, most words begin with stressed syllables, and listeners find it easier to segment speech at the onset of stressed syllables (Cutler and Norris, 1988; Vroomen, van Zon, and de Gelder, 1996). This can be clearly seen in segmentation errors, as when a pop song line *She's a must to avoid* is widely misperceived as *She's a muscular boy*—the strong syllable *void* is taken to be the onset of a new word, while the weak syllables *to* and *a-* are taken to be noninitial (Cutler and Butterfield, 1992).

The stress rhythm of English and Dutch is not universal; many other languages have different rhythmic structures. Indeed, syllabically based rhythm in French is accompanied by syllabic segmentation in French listening experiments (Mehler et al., 1981; Cutler et al., 1986; Kolinsky, Morais, and Cluytens, 1995), while moraic rhythm in Japanese likewise accompanies moraic segmentation by Japanese listeners (Otake et al., 1993; Cutler and Otake, 1994).

Thus, although the type of rhythm is language-specific, its use in speech segmentation seems universal. Other universal constraints on segmentation exist, for example, to limit activation of spurious embedded competitors. It is harder to spot a word if the residual context contains only consonants (thus, *apple* is harder to find in *fapple* than in *vuffapple;* Norris et al., 1997), an effect explained as a primitive filter selecting for possible words—*vuff* is not a word, but it might have been one, while *f* could never be a word. This constraint would operate to rule out many spuriously present words in speech (such as *tray* and *ray* in *stray).* It is not affected by what may be a word in a particular language (Norris et al., 2001; Cutler, Demuth, and McQueen, 2002) and thus appears to be universal.

The ability to extract words from continuous speech starts early in life, as shown by experiments in which infants listen longer to passages containing words that they had previously heard in isolation than to wholly new passages (Jusczyk and Aslin, 1995); none of the passages can be comprehended by these young listeners, but they can recognize familiar strings embedded in the fluent speech. One-year-olds also detect familiar strings less easily if they are embedded in a context without a vowel (e.g., *rest* is found less easily in *crest* than in *caressed;* Johnson et al., 2003); that is, they are already sensitive to the apparently universal constraint on possible words.

Finally, segmentation of second languages in later life is not aided by the efficiency with which listeners exploit language-specific structure in recognizing speech. Segmentation procedures suitable for the native language can be inappropriately applied to non-native input (Cutler et al., 1986; Otake et al., 1993; Cutler and Otake, 1994; Weber, 2001). This is one effect making listening to a second language paradoxically harder than, for instance, reading the same language.

*See also* PHONOLOGY AND ADULT APHASIA.

—*Anne Cutler*

## References

Cairns, P., Shillcock, R., Chater, N., and Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33,* 111-153.

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

Cutler, A., Demuth, K., and McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science, 13,* 258-262.

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

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

Cutler, A., and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33,* 824-844.

Johnson, E. K., Jusczyk, P. W., Cutler, A., and Norris, D. (2003). Lexical viability constraints on speech segmentation by infants without a lexicon. *Cognitive Psychology, 46.* 65 97.

Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1-23.

Kolinksy, R., Morais, J., and Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language. 34.* 19-40.

Maddieson, I. (1984). *Patterns of sounds.* Cambridge, U.K.: Cambridge University Press.

Marslen-Wilson, W. D., and Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101,* 653 675.

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 19,* 21 46.

McQueen, J. M., Norris, D., and Cutler, A. (19941. Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621-638.

McQueen, J. M., Norris, D., and Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1363 1389

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

Norris, D., McQueen, J. M., and Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1209-1228.

Norris, D., McQueen, J. M., Cutler, A., and Butterneld, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34,* 191-243.

Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., and Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes, 16,* 637-660.

Otake, T., Hatano, G., Cutler, A., and MehJer, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 258-278.

Soto-Faraco, S., Sebastian-Galles, N., and Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412—432.

van der Lugt, A. (2001). The use of sequential probabilities in the segmentation of speech..*Perception and Psychophysics, 63,* 811-823.

Vroomen, J., and de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 98-108.

Vroomen, J., van Zon, M., and de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition, 24,* 744-755.

Weber, A. (2001). *Language-specific listening: The case of phonetic sequences.* Doctoral dissertation, University of Nijmegen, The Netherlands.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25-64.

## *Further Readings*

Cutler, A., and Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. Brown and P. Hagoort (Eds), *Neurocognition of language* (pp. 123-166). Oxford: Oxford University Press.

Frauenfelder, U. H., and Floccia, C. (1998). The recognition of spoken words. In A. Friederici (Ed.), *Language comprehension, a biological perspective* (pp. 1-49). Heidelberg: Springer-Verlag.

Grosjean, F., and Frauenfelder, U. H. (Eds.). (1997). *Spoken word recognition paradigms.* Hove, U.K.: Psychology Press.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

McQueen, J. M., and Cutler, A. (Eds.). (2001). *Spoken word access processes.* Hove, U.K.: Psychology Press.