In R.A. Wilson and F.C. Keil (Eds.), The MIT Encyclopedia
of the Cognitive Sciences.Cambridge, MA: MIT Press. (1999).

Spoken-Word Recognition

Listening to speech is a recognition process: SPEECH PERCEPTION identifies phonetic structure in the incoming speech signal, allowing the signal to be mapped onto representations of known words in the listener's LEXICON. Several facts about spoken-word recognition make it a challenging research area of PSYCHOLINGUISTICS. First, the process takes place in time—words are not heard all at once but from beginning to end. Second, words are rarely heard in isolation but rather within longer utterances, and there is no reliable equivalent in speech of the helpful white spaces that demarcate individual words in a printed text such as this article. Thus the process entails an operation of segmentation whereby continuous speech is effectively divided into the portions that correspond to individual words. Third, spoken words are not highly distinctive; language vocabularies of tens of thousands of words are constructed from a repertoire of on average only 30 to 40 phonemes (Maddieson 1984; see PHONOLOGY for further detail). As a consequence, words tend to resemble other words, and may have other words embedded within them (thus *steak* contains possible pronunciations of *stay* and *take* and *ache,* it resembles *state* and *snake* and *stack,* it occurs embedded within possible pronunciations of *mistake* or *first acre,* and so on). How do listeners know when to recognize *steak* and when not?

Methods for the laboratory study of spoken-word recognition are comprehensively reviewed by Grosjean and Frauenfelder (1996). This field of study is very active, but it began in earnest only in the 1970s; before then, models of word recognition such as Morton's (1969) *logogen* model were not specifically designed to deal with the characteristics of speech. Now. spoken-word recognition research is heavily model-driven, and the models differ, inter alia, as to which of the above challenges they primarily address. The first model specifically in this area was Marslen-Wilson and Welsh's (1978) *cohort* model; it focused on the temporal nature of spoken-word recognition and proposed that the initial portion of an incoming word would activate all known words beginning in that way, with this "cohort" of activated word candidates gradually being reduced as candidates incompatible with later-arriving portions of the word drop out. Thus /s/ could activate *sad, psychology, steak,* and so on: if the next phoneme were /t/, only words beginning with /st/ *{stay, steak, stupid,* etc.) would remain activated; and so on until only one word remained in the cohort. This could occur before the end of the word—thus *staple* could be identified by the /p/ because no other English words would remain in the cohort.

The *neighborhood activation model* (Luce, Pisoni, and Goldinger 1990) concentrates on similarities between words in the vocabulary and proposes that the probability of a word being recognized is a function of the word's frequency of occurrence (see VISUAL WORD RECOGNITION for more extensive discussion of this factor) and the number and frequency of similar words in the language; high-frequency words with few, low-frequency neighbors will be most easily recognized.

The currently most explicit models are TRACE (McClelland and Elman 1986) and SHORTLIST (Norris 1994), both implemented as connectionist networks (see COMPUTATIONAL PSYCHOLINGUISTICS; also Frauenfelder 1996). They both propose that the incoming signal activates potential candidate words that actively compete with one another by a process of interactive activation in which the more active a candidate word is, the more it may inhibit activation of its

competitors. Activated and competing words need not be aligned with one another, and thus the competition process offers a potential solution to the segmentation problem; so although the recognition *of first acre* may involve competition from *stay, steak,* and *take,* this will eventually be overcome by joint inhibition from *first* and *acre.*

TRACE and SHORTLIST differ primarily in one other feature that is an important characteristic of most psycholinguistic processing models—namely, whether or not they allow unidirectional or bidirectional flow of information between levels of processing. TRACE is highly interactive. That is, it allows information to pass in both directions between the lexicon and prelexical (and in principle post lexical) processing levels. SHORTLIST allows information to flow from prelexical processing of the signal to the lexicon but not vice versa. In contrast to TRACE, SHORTLIST also has a two-stage architecture, in which initial word candidates are generated on the basis of bottom-up information alone, and competition occurs only between the members of this "shortlist." TRACE allows competition in principle within the entire vocabulary, which renders it less computationally tractable, whereas SHORTLIST's structure has the practical advantage of allowing simulations with a realistic vocabulary of tens of thousands of words.

All theoretical issues separating the models are still unresolved. There is abundant experimental evidence confirming the subjective impression that spoken-word recognition is extremely rapid and highly efficient (Marslen-Wilson 1987). Concurrent activation of candidate words is supported by a wide range of experimental findings from different experimental paradigms, and active competition between such simultaneously activated words—such that concurrent activation can produce inhibition—is also supported (McQueen et al. 1995). Many findings have been interpreted in terms of interaction between levels of processing (e.g., Pitt 1995; Samuel 1997; Tabossi 1988) but noninteractive models in general can account for these findings as well (Cutler et al. 1987; Massaro and Oden 1995). In some cases, apparent demonstrations of top-down information flow have proven to be spurious, arising instead from independent bottom-up processing (for example, Elman and McClelland 1988 reported an apparent effect of lexically determined compensation for coarticulation, but Pitt and McQueen 1998 showed that the finding was actually due to transitional probability effects and hence could be accounted for without postulating top-down lexical influences on prelexical processing).

Orthogonal to these principal questions of model architecture are further issues such as the nature of the primary prelexical unit of representation (Mehler, Dupoux, and Segui 1990; Pisoni and Luce 1987); the relative contribution to word activation of matching versus mismatching phonetic information (Connine et al. 1997); the phonological explicitness of lexical representations (Frauenfelder and Lahiri 1989); the processing of contextually induced phonological transformations such as *sweek girl* for *sweet girl* (Gaskell and Marslen-Wilson 19%); the role of prosodic structure in recognition (Cutler et al. 1997); and the role of word-internal morphological structure in recognition (Marslen-Wilson et al. 1994).

See also CONNECTIONIST APPROACHES TO LANGUAGE; LANGUAGE PROCESSING; PROSODY AND INTONATION, PROCESSING ISSUES

—***Anne Cutler***

References

Comrine, C. M., D. Titone, T. Deelman, and D. Blasko. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language* 13: 291-299.

Cuder, A., D. Dahan, and W. van Donselaar. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40: 141-201.

Cutler, A., J. Mehler, D. G. Norris, and J. Segui. (1987). Phoneme identification and the lexicon. *Cognitive Psychology* 19: 141—177.

Elman, J. L. and J. L. McClelland. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27: 143-165.

Frauenfelder, U. H. (1996). Computational models of spoken word recognition. In T. Dijksta and K. de Smedt, Eds., *Computational Psycholinguistics*. London: Taylor and Francis, pp. 114—138.

Frauenfelder, U. H., and A. Lahiri. (1989). Understanding words and word recognition: Can phonology help? In W. D. Marslen-Wilson, Ed., *Lexical Representation and Process*. Cambridge, MA: MIT Press., pp. 319-341.

GaskelLG. M., and W. M. Marslen-Wilson. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Peformance* 22: 144—156.

Grosjean, F., and U. H. Frauenfelder, Eds. (1996). Spoken word recognition paradigms. Special issue of *Language and Cognitive Processes* 11: 553-699.

Luce, P. A., D. B. Pisoni, and S. D. Goldinger. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann, Ed., *Cognitive Models of Speech Processing*. Cambridge, MA: MIT Press, pp. 122-147.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Marslen-Wilson, W. D. (1987). Parallel processing in spoken word recognition. *Cognition* 25: 71-102.

Marslen-Wilson, W., L. K. Tyler, R. Waksier, and L. Older. (1994). Morphology and meaning in me English mental lexicon. *Psychological Review* 101:3-33.

Marslen-Wilson, W. D., and A. Welsh. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10: 29-63.

Massaro, D. W., and G. C. Oden. (1995). Independence of lexical context and phonological information in speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2: 1053-1064.

McClelland, J. L., and J. L. Elman. (1986). The TRACE model of speech perception. *Cognitive Psychology* 18: 1-86.

McQueen, J. M., A. Cuder, T. Briscoe, and D. G. Norris. (1995). Models of continuous speech recognition and the contents of die vocabulary. *Language and Cognitive Processes* 10: 309-331.

Mehler, J., E. Dupoux, and J. Segui. (1990). Constraining models of lexical access: The onset of word recognition. In G. T. M. Altmann, Ed., *Cognitive Models of Speech Processing*. Cambridge, MA: MIT Press, pp. 236-262.

Morton, J. (1969). Interaction of information in word perception. *Psychological Review* 76: 165—178.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52: 189-234.

Pisoni, D. B., and P. A. Luce. (1987). Acoustic-phonetic representations in word recognition.'Cognmon 25: 21-52.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1037-1052.

Pitt, M. A., and J. M. McQueen. (1998). Is compensation for coarticulation mediated by me lexicon? *Journal of Memory and Language* 39: 347-370.

Samuel, A. G. (1997). Lexical activation produces potent phonetic percepts. *Cognitive Psychology* 32: 97-127.

Tabossi., P. (1988). Effects of context on me immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14: 153-162.

Further Readings

Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller and P. D. Eimas, Eds., *Speech, Language, and Communication,* of E. C. Carterette and M. P. Friedman. Eds., *Handbook of Perception and Cognition,* vol. 11. New York: Academic Press, pp. 97-136.

Friederici, A., Ed. (1998). *Language Comprehension: A Biological Perspective*. Heidelberg: Springer.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson, Ed.. *Lexical Representation and Process*. Cambridge, MA: MIT Press, pp. 169-226.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology* 21: 398-421.