To appear in: H. Pashler (Ed) "The Encyclopedia of Mind"
SAGE Reference, Thousand Oaks, CA, USA

The Cohort Model of Auditory Word Recognition

Matthew H. Davis

Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF, UK

matt.davis@mrc-cbu.cam.ac.uk

Cross-References:

Language Processing, Neural Basis

Neuroimaging Studies of Language

Electrophysiological Studies of Mind

Speech Perception

Speech Perception, Computational Perspectives

Statistical Models of Language Comprehension

Further reading:

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes, 12*, 613-656.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

Marslen-Wilson, W. (1984). Function and processing in spoken word recognition: a tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processing*. Hillsdale NJ: Erlbaum.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1-71.

Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, *362*(1481), 823-836.

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

Zwitserlood, P. (1989) The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32(1), 25-64.

The core idea at the heart of the Cohort model is that human speech comprehension is achieved by processing incoming speech continuously as it is heard. At all times, the system computes the best interpretation of currently available input combining information in the speech signal with prior semantic and syntactic context. Originally proposed in 1980 by William Marslen-Wilson and Lorraine Tyler the Cohort account has been subject to ongoing refinement in response to new empirical data and neural network simulations. Predictions of the model for neural responses to speech are currently being tested.

Origins of the Cohort Model (1973 to 1985):

During the 1970s, response time data collected by Marslen-Wilson and others demonstrated the speed and accuracy of speech perception and comprehension. Native speakers can shadow (i.e. repeat aloud) heard sentences with minimal delays between perception and production whilst correcting for mispronunciations of key words. Detection tasks similarly show rapid word identification in sentences with participants responding within 300ms of the start of a target word – substantially before all of the relevant speech has been heard. Early identification is achieved by comparing incoming speech with known lexical items (for example, the word *trespass* is distinct from all other words once the segment /p/ has been heard) in combination with contextual cues provided by the sentential context.

The Cohort model proposed by Marslen-Wilson and Tyler in 1980 thus suggests that word identification begins with initial activation of a set of candidates that match the start of a spoken word (the word-initial cohort which for *trespass* would include words like *tread*, *treasure*, *treble*, etc). Activated candidates are rejected as incompatible speech segments are heard such that word recognition occurs when information in the speech signal uniquely matches one single word (the uniqueness point). The importance of the sequential structure of spoken words in predicting word recognition has been confirmed in a range of response time tasks. For example, people can decide that speech does not match a real word (i.e. make lexical decisions to spoken pseudowords) as soon as they hear a segment that deviates from all spoken words. Hence, decision responses occur with a constant delay when measured from the /s/ of the *tromsone* or the /p/ of *trombope*; it is at these positions that participants can determine that these pseudowords are not the familiar word *trombone*. These findings uniquely support accounts of speech perception in which word recognition can occur when sufficient information has been perceived in the speech signal and provide strong support for the first instantiation of the Cohort model.

Computational Instantiations (1986 to 1999):

The predictions of the Cohort model for the timing of word recognition depend on assessing phonetic information that is shared by or distinguishes between different spoken words. These can be estimated from computerised pronunciation dictionaries revealing the proportion of items (like *doll* embedded in *dolphin*) that challenge early identification. A further methodological development was the use of cross-modal priming to test the time-course of word recognition. Consistent with Cohort assumptions, Pienie Zwitserlood showed that speech that matches multiple words (e.g. the start of *captain* or *captive*) activates multiple meanings to a degree that is modulated by prior sentence context. Word frequency also affects activation of word candidates, so *captain* would be more active than *capstan* due to its higher frequency of occurrence. These findings motivated a revision to the Cohort theory in which word activation and identification are graded processes that combine the speech input, lexical information and contextual cues.

This revised version of the Cohort model shares a number of important characteristics with the TRACE computational model of spoken word recognition developed by Jay McClelland and Jeff Elman in 1986. This is a computer simulation of a network of simple neuron-like processing units (neural network), that models the processes by which people recognize spoken words. These simulations motivated further tests to distinguish between the Cohort and TRACE theories. Two findings that are contra-TRACE are that: (1) ambiguous phonemes (e.g. a sound that is acoustically-intermediate between an /s/ and an /f/) do not slow down the identification of pseudowords in the same way that they slow down word identification, and (2) mispronounciations that create pseudowords impair word recognition as much as mispronounciations that make real words (e.g. hearing *bandin* for *bandage* has similar effects to *cabin* for *cabbage*). However, eye-movement effects observed for rhyme competitors (e.g. looking at a picture of a *speaker* when hearing the word *beaker*) are more consistent with the predictions of TRACE.

Much of this behavioural data is well simulated by a different neural network simulation of word recognition: the Distributed Cohort Model (DCM) developed by Gareth Gaskell and William Marslen-Wilson in the 1990s. The DCM differs from TRACE both by using distributed representations (i.e. encoding speech information over multiple rather than single artificial neurons) and by simulating phoneme perception and word identification as two parallel processes rather than two sequential stages. This allows the model to simulate the differential impact of ambiguous speech sounds on phoneme and word perception. It also provides the most accurate simulation to-date of cross-modal priming results similar to those reported by Pienie Zwitserlood. Lexical identification in DCM is achieved by activating output representations in proportion to their conditional probability given the current speech input, a proposal that to some extent anticipates recent Bayesian theories of word recognition.

Expansion and new horizons (2000 to date):

Following the success of the Distributed Cohort Model in explaining adult recognition, behavioural studies have focussed on extensions of this account to new empirical data, populations and explanatory domains. One focus of recent research is the recognition of words like *darkness* and *darkly* which compromise a stem (*dark*) and affix (-*ness*, -*ly*). These polymorphemic words are decomposed into their constituents during identification and behave differently from morphologically-unrelated pairs like *dolphin* and *doll*. A further extension of the Cohort model has been to account for behavioural evidence concerning word learning in adults and infancy. In adults, new spoken word do not show cohort-like recognition behaviour straight-away, but only if a period of sleep follows initial learning.

Infants show adult-like recognition behaviour early on in development (under 2 years old) since they also recognise words as soon as the speech signal allows similar sounding words to be ruled out.

Perhaps the most ambitious extension of the Cohort account, however, is to incorporate constraints provided by functional brain imaging. The continuous processing of speech in the Cohort account have been validated by Electro- and Magnetoencephalography (E/MEG) data showing that neural responses to spoken words are timed to the perception of critical segments in speech. Support for separate semantic and phonological pathways as in the DCM has come from brain imaging studies consistent with dual-pathway accounts of the neuro-anatomy of spoken language. A challenge for future imaging work is to establish the unique functional contribution and timing of neural responses within these distinct pathways.