Advanced Review

Models of spoken-word recognition

Andrea Weber* and Odette Scharenborg

All words of the languages we know are stored in the mental lexicon. Psycholinguistic models describe in which format lexical knowledge is stored and how it is accessed when needed for language use. The present article summarizes key findings in spoken-word recognition by humans and describes how models of spoken-word recognition account for them. Although current models of spoken-word recognition differ considerably in the details of implementation, there is general consensus among them on at least three aspects: multiple word candidates are activated in parallel as a word is being heard, activation of word candidates varies with the degree of match between the speech signal and stored lexical representations, and activated candidate words compete for recognition. No consensus has been reached on other aspects such as the flow of information between different processing levels, and the format of stored prelexical and lexical representations. © 2012 John Wiley & Sons, Ltd.

How to cite this article:
WIREs Cogn Sci 2012, 3:387–401. doi: 10.1002/wcs.1178

INTRODUCTION

In order to understand the utterance '*The sun began to rise*', listeners must recognize the individual words in that utterance. This decoding of the message must be achieved by mapping the auditory information in the speech input onto stored representations of words in the mental lexicon. Although the mapping task is usually perceived to be effortless for listeners, the underlying decoding process is in fact very complex. Particularly, three aspects of spoken language make the mapping difficult. First, words resemble each other. As languages build large vocabularies from a limited set of phonemes, words are necessarily alike (e.g., *sun, sum, suck*, and *such* only differ in their final consonant), and short words are often embedded within longer ones (e.g., *rye* and *eye* in *rise*). Second, speech is highly variable. The acoustic realization of sounds and words is different for each speaker; speaking style, speaking rate, and phonological context additionally cause variability in the signal (e.g., *sun* is usually pronounced as *sum* when followed by a bilabial stop consonant as in *began*). Third, speech is transitory and continuous. Not only is spoken language distributed in time and fades quickly from the perceptual field, the acoustic speech signal is also continuous with no clear boundaries for individual words. As shown in the spectrogram in Figure 1, breaks in the speech signal do not necessarily correspond to word boundaries (e.g., there is no break between *the sun* but there is one in *began*). This also implies that embedded words can span word boundaries (e.g., *ant* is embedded in *began to*).

Psycholinguistic research investigates how listeners master the decoding of speech. In the following, we will selectively describe those findings that highlight some of the most important findings in spoken-word recognition. Probably the most central finding is that the comprehension process is incremental. That is, listeners do not wait until the end of a word or an utterance before they interpret the input. Rather they consider multiple word candidates simultaneously that are consistent with the incoming speech. Parallel activation has been shown repeatedly in priming studies in which a word onset that is consistent with two words (e.g., /kæp/ can start *capital* and *captain)* facilitates the recognition of semantically related words for both possible continuations (*money* and *ship*[1]). Parallel activation has also been found for embedded words (e.g., *bone* in *trombone* facilitates recognition of the semantically related *rib*[2,3])

*Correspondence to: andrea.weber@mpi.nl

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

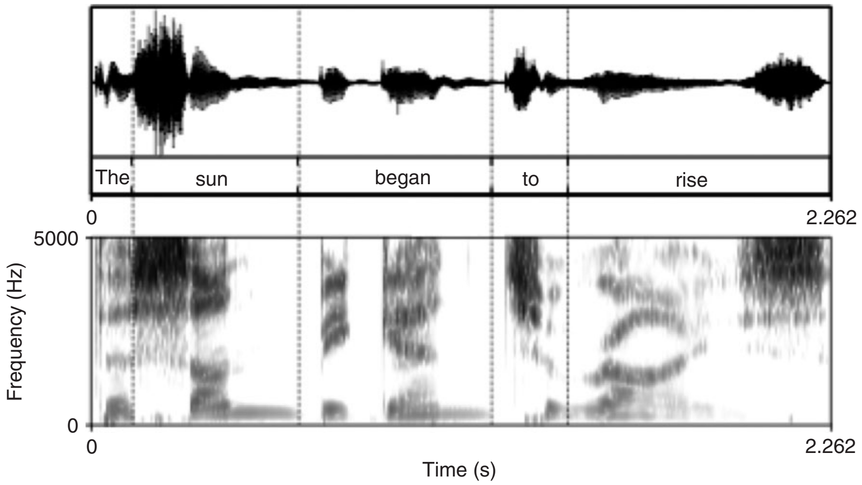

FIGURE 1 | Waveform and spectrogram of the utterance ''The sun began to rise''. The horizontal axis represents time and the vertical axis amplitude for the waveform and frequency for the spectrogram, with greater energy being represented by darker shading. White spaces in the spectrogram correspond to breaks in the speech signal. The vertical dotted lines are aligned as closely as possible with word boundaries.

and words that are spanning a word boundary.[4] More recently, the eye-tracking paradigm has been used to demonstrate this core process of spoken-word recognition: when being presented with a display of objects, listeners take longer to look at a target object mentioned in an utterance when the display includes objects with similar names (e.g., later looks to *candle* when the display also shows *candy*[5]). This suggests that listeners temporarily consider the two objects with similar names as possible targets.

Lexical activation is furthermore modulated by acoustic-phonetic detail. Specifically, the goodness of match between the speech signal and stored representations co-determines how strongly word candidates are activated. It has been shown, for instance, that *soup* is recognized more slowly when the formant transitions following the /s/ are manipulated to be typical of a different fricative than when they are typical of /s/.[6] Connine et al.[7,8] have found in priming studies that words that mismatch with the signal on multiple articulatory features are less strongly activated than words that mismatch on one feature. But also phonological context influences the effect of mismatch: [gri:m] can be recognized as *green* in the context of *green bench* where the bilabial stop onset in *bench* licenses the place assimilation of the preceding nasal /n/.[9–11] Many more studies using different paradigms have confirmed that the degree of lexical activation varies in response to fine-grained acoustic–phonetic information.[12–14]

Word candidates are not only activated in parallel, they also compete for recognition. That is, a candidate's activation is not independent of the activation of other candidates, and the more candidates are active, the more they inhibit each other. Competition has been shown with a variety of behavioral paradigms, and it is a generally assumed component of spoken-word recognition. Listeners in a word-spotting task, for instance, find it more difficult to spot short embedded words in word onsets of longer words than in onsets of nonwords: *sack* is harder to spot in /sækrəf/, the beginning of *sacrifice,* than in [sækrək] which has no possible word continuation.[15] This presumably reflects the competition between *sack* and *sacrifice*. Furthermore, it has repeatedly been found that lexical decision times are influenced by the number of similar-sounding words in the lexicon,[16,17] as well as by preceding words that are phonologically related.[18] Although competition alone can correctly parse an utterance into a sequence of individual words, listeners use a variety of cues to likely word boundaries to further help the segmentation process. These cues include phonotactic constraints and probabilities,[19–22] metrical cues,[23–25] and fine-grained acoustic information.[26–28]

MODELS OF SPOKEN-WORD RECOGNITION

Early models of word recognition were developed on the basis of data obtained in reading tasks,[29,30] but were often assumed to account for spoken-language processing as well. Morton[30] introduced in his logogen model the powerful metaphor of activation, which conveys that multiple words in our mental lexicon are responsive to the speech signal. The metaphor of activation still features in many subsequent models as it captures the notion of parallel availability that behavioral studies have shown to be at the heart of spoken-word recognition. Only later was it realized that the temporal nature of the speech signal has far reaching consequences for the comprehension process, and that models of spoken-word recognition must account for the transitory nature of speech, acoustic–-phonetic and phonologically conditioned variation, as well as the continuity of the speech signal. Since the 1980s, a number of models have then been developed specifically for spoken-word recognition.

From a theoretical point of view, current models of spoken-word recognition differ particularly in two aspects. First, they vary in their assumptions about the abstractness of the representations that make contact with the lexicon, as well as the nature of the lexical representations themselves. Second, the models differ with respect to information flow between levels of the processing system. Different levels are responsible for different processing stages and are ordered from relatively low-level acoustic–phonetic processing to higher stages involving the lexicon. Interactive models not only allow information to flow from lower to higher levels but also top-down information flow, whereas autonomous models assume that flow of information is unidirectional from the bottom up.

Models not only vary in their theoretical assumptions but also in type, and different terms have been suggested in the literature for the varying types of models.[31] In the present review we use the term *verbal* for models that explain the stages and mechanisms of spoken-word recognition descriptively, the term *mathematical* is used for models that capture the processes of spoken-word recognition with a mathematical form, and *simulation* models are models that aim to account for the cognitive processes in speech comprehension. All mathematical and simulation models are computationally implemented as computer programs. Most current models of spoken-word recognition are computational models. An advantage of computational models is that they can be used to simulate the conditions of behavioral research and to compare a model's predictions with behavioral results obtained from human listeners. A criticism of computational models is that in order to build a functioning model, theoretical and implementational assumptions need to be made that are possibly unspecified in the behavioral research.[32,33] An example of this are the different assumptions models make about the form of prelexical representations (e.g., multidimensional features in TRACE and phonemes in Shortlist; see also below). Thus, behavioral findings are not necessarily a direct validation of how aspects of spoken-word recognition are incorporated in a computational model (see the demand for a linking hypothesis in Ref 34).

The Cohort Model

The Cohort model[35,36] was the first psycholinguistic model of word recognition specifically developed for spoken language. Central to this verbal model is the temporal aspect of spoken language, that is, the availability of acoustic–phonetic information over time. The Cohort model provided many predictions about the time-course of recognition, and it motivated substantial research that paved the way for the further development of models.

In the Cohort model, spoken-word recognition takes place in three stages: *access*, *selection*, and *integration*. During *access*, acoustic–phonetic elements in the speech signal are mapped onto words in the lexicon. Words that match with the input are activated simultaneously and make up the cohort. This simultaneous consideration of multiple candidate words is central to all subsequently developed models. In the Cohort model, however, only words that are aligned with the onset of the input are activated. For example, the Cohort model assumes that after the initial 150–200 ms (roughly consistent with the first two phonemes of a word), all words beginning with those phonemes will be activated. During *selection*, candidate words that mismatch the incoming speech signal by more than a single feature are removed from the cohort. For example, on hearing /fɛ/, all words beginning with /fɛ/ are activated; when the subsequent sound is /b/, words that do not begin with /fɛb/ drop out of the cohort. This process repeats until (ideally) the cohort is reduced to one member. The focus on onset overlap implies that words can be recognized before their offset. *February*, for instance, can be recognized by the third segment, because no other English word begins with /fɛb/. During *integration*, the syntactic and semantic properties of activated words are retrieved and checked for integrability with higher levels. A mismatch with contextual constraints, for instance, can result in the removal from the cohort. Sentential context can thus affect the selection stage in the original Cohort model. The candidate words in the cohort do not actively compete with one another; it is just the presence of other candidate words that forms the recognition process. Segmentation of the speech stream follows implicitly from the recognition of individual words: a word's offset signals the start of a new word.

A number of behavioral findings challenged the Cohort model. It had been found, for example, that listeners can recognize words that mismatch acoustically or contextually,[37] but the removal of mismatching words from the cohort entails that the model cannot recover from mismatches. Also listeners recognize frequent words more easily than infrequent ones,[38] but the Cohort model cannot capture word-frequency effects.

The successor version Cohort II[39–41] adjusted its architecture to account for these findings. In contrast to the original version, Cohort II is a fully bottom-up model. Words that (minimally) mismatch the input can now enter the cohort and can therefore be recognized; one part of the solution for handling mismatches is the introduction of word activation with selection and activation being dependent on the goodness of fit with the word input. In addition, the input to the model is now more fine-grained. To account for word-frequency effects, candidate words were assigned resting activation values in Cohort II, with higher values for frequent than for infrequent words, which makes frequent words reach the threshold for recognition faster.

The main challenge for the Cohort model, however, proved to be analyses of on-line dictionaries that showed that relatively few words can be uniquely identified before word offset,[42] and that listeners do not recognize the majority of words correctly until after word offset.[43] It was therefore a logical consequence that ensuing models should no longer consider only word candidates that match in onset with the speech input but also allow later parts of a word to be relevant. Allowing the activation of candidates that match with later parts is also a prerequisite for being able to handle the segmentation of continuous speech.

TRACE

TRACE[44,a] was the first computationally implemented model of spoken-word recognition. It is a localist (i.e., one node represents one representational unit) connectionist interactive-activation framework[45] with three layers of nodes: a feature, a phoneme, and a word layer (see Figure 2). The input to TRACE consists of multidimensional features, and words are represented as phonemic strings. TRACE was the first model that instantiated the activation of multiple word candidates that match any part of the speech input. That is, nodes are activated in proportion to their degree of fit to the input, with activation spreading through the layers (e.g., activated feature nodes spread activation to matching phoneme nodes and on to word nodes), so that on hearing the word *sun*, overlapping words like *under* and *run* are also considered in parallel. Moreover, this mechanism ensures that TRACE is able to handle ambiguous or distorted speech. Activated nodes on the phoneme and word layer receive active inhibition from other nodes compatible with the same portion of input. The word with the highest activation will inhibit candidate words with lower activation during competition, and finally the candidate word best matching the input will be recognized. Inhibitory connections on the word level help to solve activation of multiple word candidates (i.e., the fewer the word candidates that actively compete with each other, the easier recognition is). There is no inhibition between layers in TRACE, and word activation does not decrease in the presence of mismatching input. The temporal aspect of speech is handled by TRACE by duplicating all phoneme and word nodes across time (e.g., the phoneme node /s/ is duplicated for all time slices of the word *sun*, but it is activated the strongest when the feature nodes representative of /s/ are aligned in time). Feedback connections from the word layer to the phoneme layer make TRACE an interactive model. Through these connections, lexical knowledge can affect perception.

Word-frequency effects were not accounted for in the original TRACE model. However, they were later implemented by Dahan, Magnuson, and Tanenhaus,[46] who proposed three possible ways of incorporating frequency in TRACE: by adjusting resting-activation levels, by adjusting connection strengths, or as a post-activation decision bias.

TRACE successfully simulated a wide range of behavioral findings, including the Ganong effect[47] and the finding that lexical information is not used for phoneme monitoring.[48] For simulations, TRACE relies on a large number of parameters that have to be set correctly. A strength of TRACE is that the parameter settings as determined by McClelland and Elman have been used for all simulations in the original paper and were only changed slightly for later simulations. Thus, TRACE's parameters do not have to be tweaked to fit individual data.

Continuous mapping of speech input to lexical representations as in TRACE predicts activation of word candidates that overlap in onset earlier than those that overlap in rhyme with the speech input. Such a difference in the time course of activation was indeed found in a seminal eye-tracking study by Allopenna et al.[49] In this study, listeners looked earlier and more at onset competitors than rhyme competitors, with the pattern of eye fixations closely matching the pattern predicted by TRACE. The results convincingly showed that continuous mapping models can generate quantitative predictions about the word recognition process over time.

The two most controversial components of TRACE are the implausible duplication of the network,[44,50] and the existence of the lexical feedback loop.[51] In order to recognize words over time, the entire lexical network in TRACE needs to be duplicated many times. Consequently, TRACE can

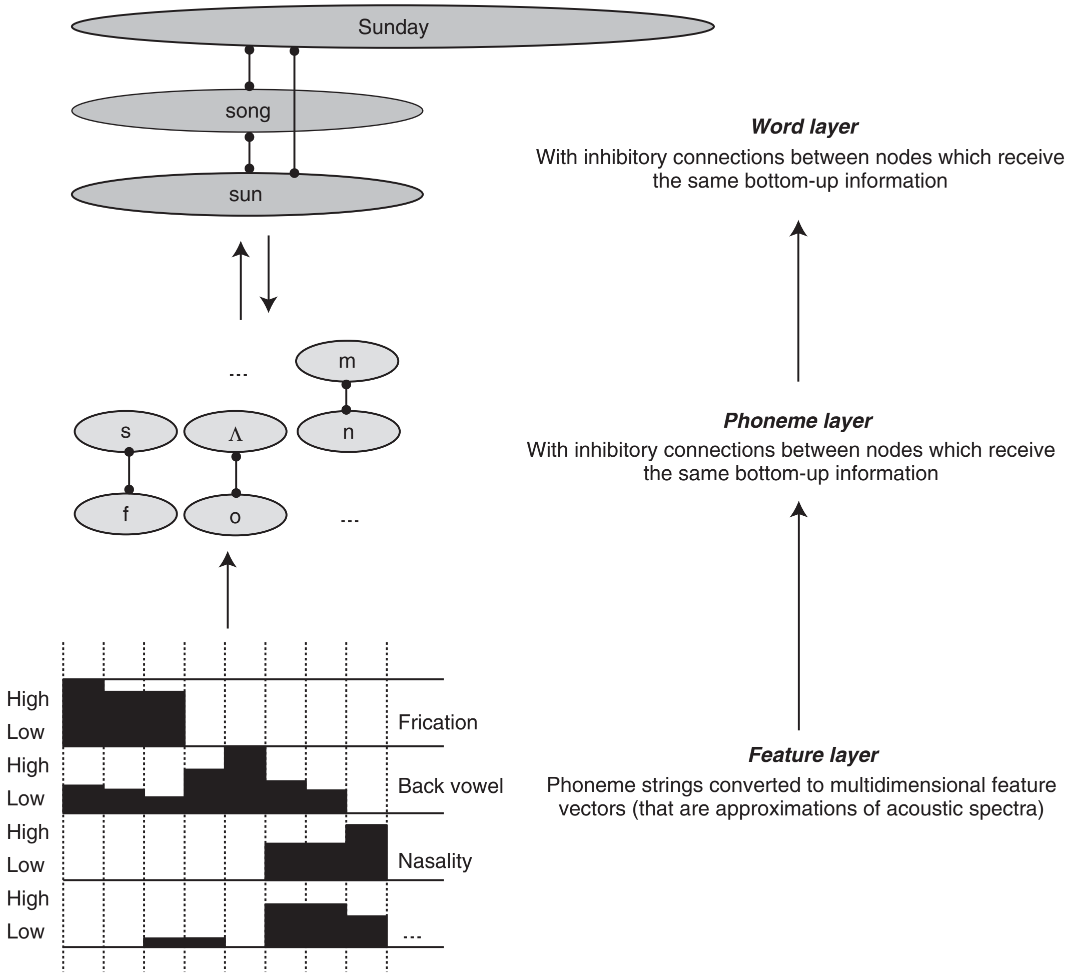
Sunday
song
sun
Word layer
With inhibitory connections between nodes which receive the same bottom-up information
m
...
s
Λ
n
f
o
...
Phoneme layer
With inhibitory connections between nodes which receive the same bottom-up information
High
Low
Frication
High
Low
Back vowel
High
Low
Nasality
High
Low
...
Feature layer
Phoneme strings converted to multidimensional feature vectors (that are approximations of acoustic spectra)

FIGURE 2 | Recognition process of the word *sun* by TRACE. For every time slice the entire network is copied, for better visualisation this duplication is not depicted in the figure. Activation in the lower layers flows upwards to the higher levels to all nodes that incorporate the lower layer node. Activation from the word layer also flows back to the phoneme layer.

only handle unrealistically small lexicons. Simulations typically involve lexicons of just a few hundred words, and use only a limited subset of English phonemes. Lexical feedback on the other hand, has been argued to be unnecessary since it cannot speed processing or improve accuracy,[52] and it can furthermore prevent recovery from mispronunciations.[53] Proponents of interactive models have pointed out that lexical feedback is in line with research showing that lexical knowledge allows listeners to quickly adapt to speakers with unfamiliar pronunciation,[54] but proponents of feed-forward models have countered that feedback for perceptual learning is different from online feedback as is implemented in TRACE.[55,56]

Shortlist

Shortlist[50] was developed in response to the criticism of duplication and lexical feedback in TRACE, and combines aspects of feed-forward models, such as the phoneme decision model Race[57] and Cohort II, with the competition mechanism of TRACE. The duplication of the entire network for each input feature in TRACE is avoided by implementing Shortlist as a two-stage model in which the generation of lexical candidates and the competition process are separated (see Figure 3). The first stage consists of an exhaustive serial lexical search (although it is assumed that the search in humans occurs in parallel), which results in a shortlist of maximally 30 candidate words that match the input processed so far. Subsequently, these word candidates are wired into a small interactive-activation network (the second, competition stage) in which the words that receive support from the same section of the input are connected via inhibitory links and compete with one another. Activation of candidate words is

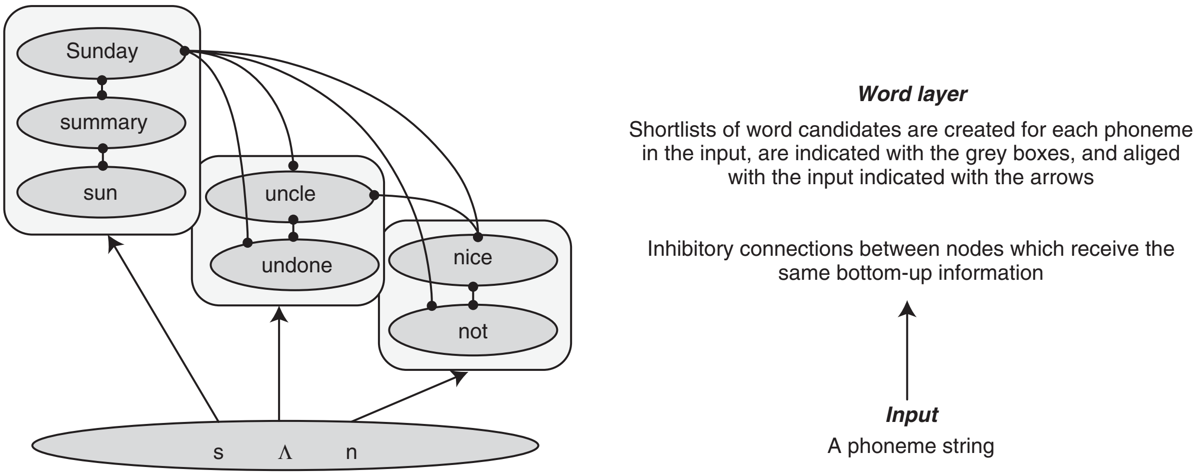
Sunday
summary
sun
uncle
undone
nice
not
s ʌ n
Word layer
Shortlists of word candidates are created for each phoneme in the input, are indicated with the grey boxes, and aliged with the input indicated with the arrows
Inhibitory connections between nodes which receive the same bottom-up information
Input
A phoneme string

FIGURE 3 | Recognition process of the word *sun* by Shortlist. For every time slice, a new shortlist (indicated with the gray box) is created, which is subsequently wired into the competition stage. For better visualisation this repetition is not depicted in the figure. Candidate words that overlap with each other at any position compete with one another. In this example, all candidate words would inhibit one another; however, for better visualisation not all inhibitory connections are shown.

determined by their degree of fit with the input, where word activations decrease with mismatching information. The word with the highest activation will inhibit candidate words with lower activation during competition, and finally the candidate word best matching the input will be recognized. The interactive-activation network is equivalent to the word layer of TRACE. The entire process is repeated with each new phoneme that becomes available, so that there is a separate shortlist and word layer for each input phoneme. Shortlist is a feed-forward only model.

The two-stage set-up makes it possible for Shortlist to use a more realistically sized lexicon of over 26,000 words. As in TRACE, words in the lexicon are represented as phonemic strings, and word candidates can be activated at any moment in time; word beginnings and endings are not explicitly marked. Word-frequency effects are not accounted for in Shortlist.

Shortlist has two unique features: lexical stress can constrain word activation (as has been found for speakers of stress-timed languages such as English and Dutch, who use the rhythmic distinction between strong and weak syllables for segmentation[15,23,25,58–60]), and activation of candidate words is decreased when they leave adjacent input that cannot constitute a viable word (e.g., since a single consonant cannot be a word in English, activation of *apple* in *fapple* is reduced[61]) through the implementation of the possible-word constraint. Much like TRACE, Shortlist can make detailed predictions of word activation over time. Shortlist successfully simulated various behavioral findings, such as the right-context problem[43,58,62] as well as results from cross-model priming studies regarding the time course of multiple word activation, competition, and selection.[15,59]

Shortlist B is a newer version of the original Shortlist model (ever since called Shortlist A, A for activation[53]), which argues that human listeners are ‘optimal Bayesian recognizers’ (p. 357). The theoretical assumptions underlying Shortlist B are identical to Shortlist A, but the implementation of the model is fundamentally different. First, Shortlist B is based on Bayesian principles; word candidates no longer have word activations, but word probabilities that were developed using techniques from the field of automatic speech recognition.[33,63] Second, the input no longer consists of handcrafted phoneme strings, instead it is a sequence of phoneme probabilities over three time slices per segment, derived from a large-scale gating study.[64] Shortlist B incorporates word frequencies as prior probabilities, and is able to handle mismatches in the input through the computation of likelihoods. Shortlist B successfully simulates various behavioral findings, including data on the segmentation of continuous speech[43] and word frequency effects.[46,65] Shortlist B can be used to make detailed predictions on the ‘optimality of the word recognition process’[53] (p. 391).

Fine-Tracker

Fine-Tracker[33] was specifically developed to account for the accumulating evidence that fine-phonetic detail, as provided in durational and prosodic information, is important in word recognition.[12,66,67] The

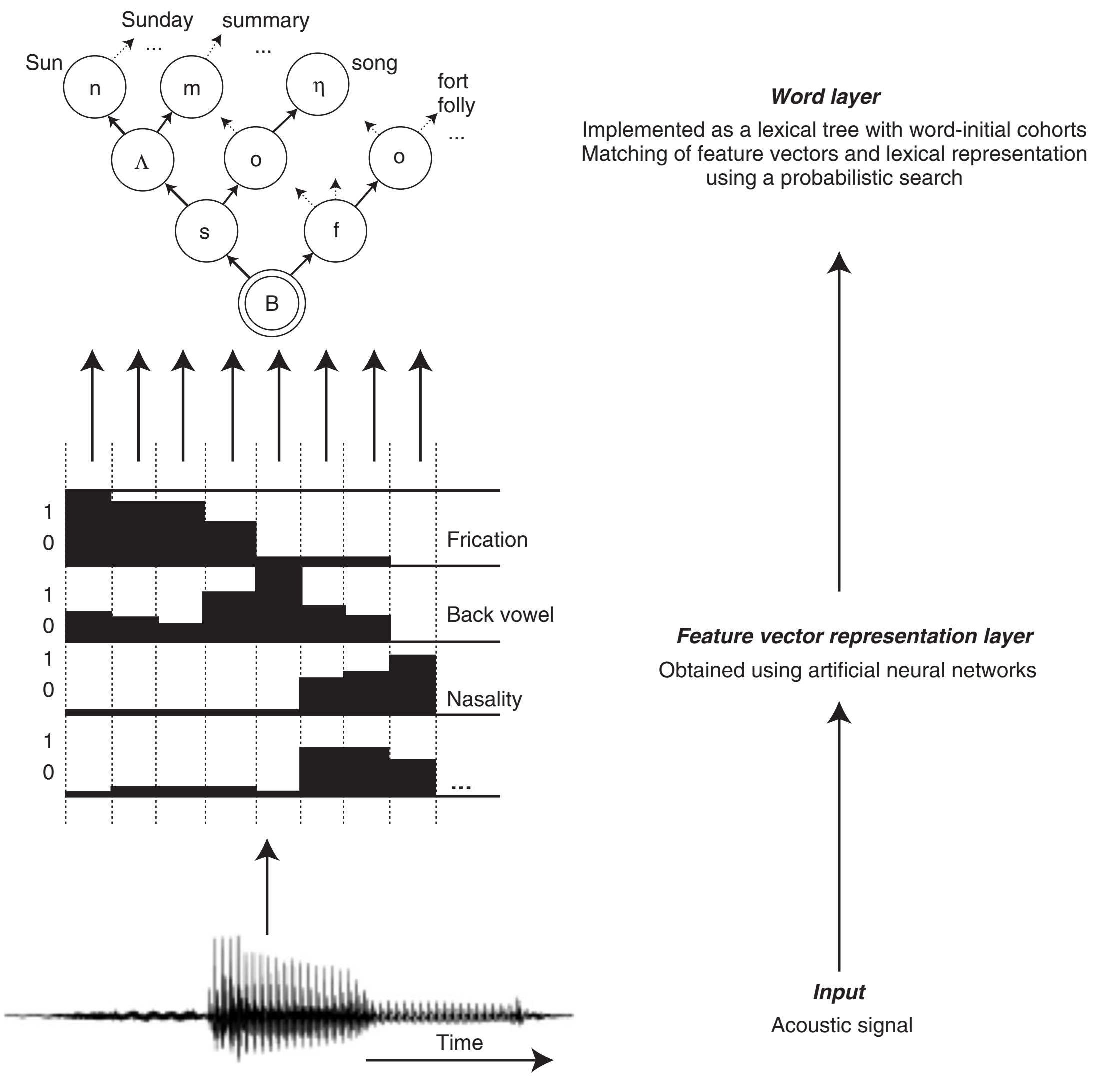
Sunday
summary
Sun
song
fort
folly
...
n
m
ŋ
ʌ
o
o
s
f
B
Word layer
Implemented as a lexical tree with word-initial cohorts
Matching of feature vectors and lexical representation
using a probabilistic search
1
0
Frication
Back vowel
Nasality
...
Feature vector representation layer
Obtained using artificial neural networks
Input
Acoustic signal
Time

FIGURE 4 | Recognition process of the word *sun* by Fine-Tracker. The acoustic signal is transformed into a sequence of feature vectors over time by a set of artificial neural networks. At the word layer, words are represented as feature vectors, for better visualisation they are depicted as phonemes in the figure. Fine-Tracker's lexicon is implemented as a lexical tree, with 'B' as the beginning of the tree. Not all possible paths in the lexical tree are shown. Each node can be followed by multiple other nodes, indicated with the dotted arrows as examples. The input feature vectors and the lexical feature vectors are mapped onto one another using a probabilistic word search.

role of subtle phonetic information is problematic for computational models that assume a discrete, abstract level between input and lexicon, because the abstract representations are too coarse to capture phonetic details. Unlike humans, these models cannot use durational information to avoid activation of (the slightly longer) *ham* in *hamster*.[67]

Fine-Tracker is based on the theory underlying Shortlist, and, like its predecessor SpeM,[63] takes the actual acoustic signal as input. It consists of two modules (see Figure 4). The first module is an artificial neural network (ANN) consisting of an input, hidden, and output layer, which converts the acoustic signal into 'articulatory feature' vectors, created over small time steps. The value for each of the articulatory features can be regarded as the likelihood of that articulatory feature. The feature vectors are then the input to the word recognition module. In the Fine-Tracker lexicon, words are represented in terms of articulatory feature vectors. Because these vectors can take any value between 0 and 1 (which are the canonical values for lexical vectors), contextual phenomena like assimilation and nasalization of vowels can be encoded through feature spreading. Fine-Tracker's word recognition module uses a probabilistic word search (dynamic time warping, a standard technique in automatic speech recognition) to match the prelexical feature vectors onto the candidate words in the lexicon in order to find the most likely sequence of words; multiple prelexical vectors (one for every 5 ms of speech) are sequentially mapped onto a single lexical feature vector. For each of the prelexical vectors the degree of fit with the lexical vector is calculated and affects the likelihood of a word. The number of feature vectors can be set for each phoneme or word separately in the lexical vector thus ensuring that the model can deal with durational information. In this way, lexical representations can vary in duration (e.g., the duration of 'ham' in *ham* and *hamster*). Fine-Tracker incorporates word and word co-occurrence frequencies. Similar to Shortlist B, multiple activations, competition, and selection are thus implemented as a probabilistic word search. Words can start and end at any time, and there is no explicit segmentation process. Unlike in TRACE or Shortlist, candidate words do not actively suppress or inhibit each other. The output of the word recognition module consists of an ordered list of the best matching hypothesized parses.

A strength of Fine-Tracker is that it can be tested with real speech rather than an abstract form of input representation as is used by other models of word recognition. Moreover, the activation flow of candidate words over time in Fine-Tracker has been successfully linked to word activation in eye-tracking studies[33] that examined the use of durational cues in word recognition.[67,68] A shortcoming of real-speech models is that due to limitations of the speech conversion module, i.e., the imperfect conversion of the speech signal to prelexical representations, such models are currently only able to use a small subset of a language's vocabulary.[63,69] Obviously, if the speech conversion module fails, everything downstream will as well. Better speech conversion modules are therefore of paramount importance in the development of better real-speech models.

NAM/PARSYN

The neighborhood activation model (NAM[65]) is a mathematical model of spoken-word recognition. It was developed to examine effects of number of similar words and their word frequencies on spoken-word recognition. In NAM, the input is assumed to activate a set of words (stored as acoustic–phonetic patterns) that differ maximally by one phoneme from the input. The difference can be by deletion, addition, or substitution. Activation is determined by degree of fit with the input; that is, NAM computes a frequency-weighted neighborhood probability for each word. The acoustic–phonetic patterns then activate word decision units. Activation of word decision units is determined by activation of the acoustic–phonetic patterns, by higher-level lexical information (i.e., word frequency, which is calculated by weighting each neighbor in the metric by its log frequency), and by overall level of activity in the entire system of word decision units. Decision values are computed on the basis of a frequency-biased, activation-based version of R.D. Luce's choice rule.[70] The choice rule in NAM approximates the competition process. A word is recognized if its decision value is above a certain threshold. NAM makes several predictions about the effects of the number of similar words and their word frequency on spoken-word recognition, for which there now is considerable evidence from behavioral studies.[42,65,71,72] As such, NAM had a large impact on theories of spoken-word recognition and research on spoken-word recognition in general, as studies on spoken-word recognition now often control for neighborhood density.[73]

PARSYN[18] is the connectionist instantiation of NAM. It consists of three levels: an input level of position-specific allophones, a level of allophones, and a word level. Like in the previously discussed connectionist models, activation spreads bottom-up through the levels. Competition is implemented as inhibitory connections between the words on the word level. Word boundaries are explicitly marked in the input. Unlike TRACE, Shortlist, and Fine-Tracker, NAM and PARSYN are only able to recognize words in isolation, but not in continuous speech. PARSYN successfully replicated the findings NAM was able to simulate and extended on that, e.g. with the simulation of findings from priming studies which showed that phonetic priming does not depend on target degradation, but that it affects processing times.[18]

Minerva2

Minerva2 is an episodic (or exemplar) model of memory.[74] Whereas all earlier described models assume abstract prelexical and lexical representations, an episodic theory of spoken-word recognition considers acoustic variability due to speaking rate or voice characteristics, for instance, an integral part of the theory and keeps this information in memory. Goldinger[75] used Minerva2 to investigate an episodic view of spoken-word recognition, motivated by the fact that the speech signal is highly variable (i.e., the lack of invariance), and that listeners' good memory for surface forms of words is well attested.[76] Minerva2 simulates episodic memory by storing numerous, independent memory traces for every word. When a new word is presented at the model's input (the 'probe' in the form of a vector of numeric elements), it is compared to all traces in memory. Activation of the traces is then dependent on the degree of fit with the probe. Subsequently, an 'echo' is retrieved, which constitutes essentially a weighted composite of all activated traces, and which may contain information not present in the probe, such as its word class. The intensity of an echo corresponds to word activation in abstract models. In Minerva2, words are represented by vectors of numeric elements. Note that although Minerva2 is a pure episodic model, it can mimic abstract behavior due to the blending of probes and stored traces, forming experience. Repeated presentation of multiple tokens of a word will thus result in an echo that mainly captures common aspects of traces (thereby eliminating the idiosyncratic characteristics stored in individual traces).

The issue of feedback from the lexical to the prelexical level does not arise, because episodic models like Minerva2 do not have an intermediate level between input and lexicon. Abstract intermediate representations have been argued to render word recognition more efficient by avoiding redundancy at the lexical level: when acoustic knowledge about a sound is stored prelexically, it need not be stored separately for every word containing that sound on the lexical level.[56,77,78] However, recoding the speech signal into abstract representations is very difficult due to the high variability and complexity of the speech signal.

Because of its nature, Minerva2 incorporates fine-grained speaker-specific information and uses it for word recognition. Minerva2 correctly predicts for instance the tendency of participants in a shadowing task to imitate the acoustic pattern of the word they have to repeat,[75] and the sensitivity of listeners to words spoken in the same voice and different voices.[79]

The model offers currently no solution for recognizing continuous speech; episodes are always single words, and it is not clear how multiple words in an utterance could be identified. Furthermore, no mechanism has been suggested for how the similarity mapping between speech signal and stored memory traces could be achieved (without reducing the surface variability in some form).

Distributed Cohort Model

The Distributed Cohort Model (DCM[80]) works from the key assumption of connectionist theory that information is represented in a distributed manner[81,82] and as such deviates from all previously discussed models in that it combines recognition of form and meaning. DCM is a connectionist model, but unlike TRACE, Shortlist, and PARSYN, information is represented in a distributed manner, that is, there is no one-on-one mapping of word and node in the model. Importantly, nodes in DCM stand for phonological and semantic features of words. The model has an input layer which takes binary phonetic features as input and a hidden layer, which is connected to two sets of output units, one for the phonological features of a word and one for semantic features.

Because DCM is a distributed model, explicit intermediate levels of representation are not needed; instead DCM regards the speech recognition process as a direct mapping from phonetic features onto distributed abstract representations of both form and meaning simultaneously. As in the previous models, the mapping process is based on similarity. The goal of the model is not to explicitly recognize the phonological form of words, but rather to retrieve phonological and semantic information from speech input. Immediate access to semantic information in continuous speech can help, for example, to reduce the activation of semantically implausible candidate words.[83]

Since all words are represented with the same set of nodes in DCM, there is no explicit activation of a candidate word and no direct competition between them. Instead, activation and competition are implicit in the blend formed by the patterns of the candidate words. Word activation is inversely related to the distance of the model's output and the target word's representation. Competition in DCM is mediated by the number of candidate words in the set; the higher the number of candidate words, the lower their activation. Semantic information starts out as a blend of the semantic vectors of all candidate words. As the number of candidate words is reduced with more input being available, the blend consists of the semantic vectors of fewer candidates and eventually results in the vector activation of the remaining single word candidate.

Word beginnings and endings are not explicitly marked in the input. The binary input features are chosen such that fine-grained information regarding the representation of vowel transitions can be captured, which makes DCM able to simulate the effect of mismatching vowel transitions.[14] Word frequency can be taken into account through repeated presentation of the word during the training phase of the model.

One prediction of DCM is that word beginnings with few completion possibilities (e.g., /ga:m/ can only be completed as *garment*) should exhibit stronger semantic activation than words with many possibilities (e.g., /kæpt/ can start *captive* and *captain)* since for the later semantic information is still a blend of words. This is exactly what Gaskell and Marslen-Wilson found in a priming study.[11] It has been argued, however, that breaking the comprehension process into separate stages is cognitively more economical than a combined mapping of form and meaning as put forward in the DCM.[52,84] Additionally, evidence from priming studies supports the assumption that phonological and conceptual representations are possibly separate and to a certain extent independent components of word recognition.[85]

SUMMARY AND CONCLUSION

In the previous section, we described the basic architecture of a number of influential models of spoken-word recognition. We furthermore tried to point out for each model where its strengths and weaknesses lie. Table 1 summarizes the main aspects of the models providing a quick overview of the commonalities and differences between the models.

The list of models is however not complete. The focus of our model overview is on models of lexical processing, we therefore omitted models with an emphasis on speech sound perception such as the LAFS model,[87] the Laff model,[88] ARTWORD,[89] and FLMP.[90] Although lexical aspects can act a part in these models, the accounts usually give no explicit description of word recognition.

Furthermore, there are two relevant issues that we have not explicitly discussed in the model overview: semantic and morphological processing. With the exception of DCM,[80] the models in Table 1 are concerned with the recognition of word form and not of meaning. On the other side of the spectrum, numerous models exist that are mainly concerned with meaning and not with phonological form. In general, these models explain how meaning is organized in the mental lexicon and less which mechanisms are used to access meaning. Classical examples of semantic models are the hierarchical network model,[91] the semantic feature model,[92] the spreading activation model,[93] and the ACT model.[94] The question of whether phonological representations of words are tantamount to semantic representations is also a matter of debate in the field of spoken-word recognition. A typical empirical approach to this question is to compare form priming with semantic priming.[85] Based on these studies, it has been argued that phonological forms are separate from conceptual representations, and that during word recognition phonological representations are activated first, but that activation cascades through to conceptual representations as soon as possible (but see e.g., Ref 95).

The main question with respect to morphological structure in lexical activation is whether morphologically complex words are stored as whole forms that do not reflect their morphological complexity (full listing[96]), as multiple morphemes with separate access representations (full parsing[97]), or that storage depends on the regularity of the morphological forms (dual-route[98,99]). Although most of the research on morphology has been done with reading, a considerable amount of research by now has been conducted in the auditory domain (see Refs 100 and 101 for reviews); form priming[102] and word reconstruction[103] are typical tasks to investigate morphological processing in the auditory domain. Baayen, McQueen, Dijkstra, & Schreuder[104] proposed a model in which phonological representations of full forms, as well as of stems and affixes are all activated in parallel; such an account is in line with the competition-based models of spoken-word recognition described above.

Having summarized how standard models of spoken-word recognition relate to models of semantic and morphological processing, we want to turn now to the question of where the field goes from here. Obviously the remaining disagreement on flow of information (feed-forward versus top-down) and form of stored representations (abstract versus episodic) must be settled. With respect to flow of information, empirical evidence is needed that shows whether lexical knowledge can directly influence pre-decisional prelexical processing or not; researchers on both sides have acknowledged that it is difficult to develop studies that can convincingly make this point (for both sides[105,106]). With respect to form of representations, it has become obvious that both purely abstract models and purely episodic models are incomplete, and the challenge for the future is to develop a hybrid approach that combines both abstract and episodic representations[107,108]; an example of such a complementary system account can be found in Norman and O'Reilly,[109] and see also an account of Connine and colleagues in which abstract lexical representations encode phonological variants based on variant frequency.[110]

Models of spoken-word recognition have often been developed with a focus on particular aspects of lexical processing: the size of the phonological neighborhood in NAM,[65] for example, or lexical segmentation in Shortlist.[50] Other parts of the models are frequently underspecified. This makes it difficult to assess them. Not only is it hard to determine how well the models can simulate specific empirical findings, judging whether the theoretical assumptions in the model are consistent with an effective complete recognition system is nearly impossible.[32,33] For example, many models make the simplifying assumption that the word recognition process receives

TABLE 1 | Models of Spoken-Word Recognition

Model	Primary References	Input Representation	Prelexical Representations	Word-form Representation	Online Lexical–Prelexical Feedback	Competition Process	Handling Fine-Grained Information	Type of Model
Cohort model	Marslen-Wilson & Welsh,[36] Marslen-Wilson & Tyler[35]	Not specified	Features	Underspecified phonological structures	No	Decision-level process, no inter-word competition	No	Verbal
TRACE	McClelland & Elman[44]	Multidimensional features, which are converted into phonemes	Features and phonemes	Logogens	Yes	Interactive-activation network, with active, direct inhibition of phoneme and lexical nodes	Partly	Simulation
Shortlist	Norris[50]	Phoneme strings	Phonemes	Phoneme strings	No	Interactive-activation network, with mismatch parameter and direct competition between words	No	Simulation
Shortlist B	Norris & McQueen[53]	Sequence of phoneme probabilities over Three time slices per segment	Phoneme probabilities	Phoneme strings	No	Beam search	Partly	Simulation
Fine-Tracker	Scharenborg[33]	Acoustic signal	Articulatory-acoustic feature vectors	Feature vector strings	No	Beam search	Yes	Simulation
Neighborhood Activation Model (NAM)	Luce[42]	Acoustic–phonetic patterns	Acoustic–phonetic patterns	Logogens	No	Decision-level process, no inter-word competition	Partly	Mathematical
PARSYN	Luce et al.[18]	Context-sensitive allophones	Allophones	Logogens	Yes	Interactive-activation network, direct competition between words	Partly (due to allophones)	Simulation
Minerva2	Hintzman,[74] Goldinger[75]	Numeric vectors of −1, 0, +1	N/A	Episodic traces	N/A	Decision-level process, no inter-word competition	Yes	Simulation
Distributed Cohort Model	Gaskell & Marslen-Wilson[80]	Multidimensional features	Phonetic features	Distributed vectors	No	No direct competition between words; competition inversely related to the size of the cohort	Partly	Simulation

a sequence of abstract units (typically phonemes or features) as input rather than actual spontaneous speech. If this simplifying assumption is abandoned, it could have serious consequences for the way other components of the model work. What is therefore needed is a unifying theory that accounts for all aspects of spoken-word recognition by human listeners.

NOTE

[a]The actual name of the model is TRACE II. TRACE I[111] focused on the conversion of digitized speech into a set of phonetic features, and was never connected to TRACE II. However, TRACE is commonly used to refer to the model of spoken-word recognition.

REFERENCES

1. Zwitserlood P. The locus of the effects of sentential-semantic context on spoken-word processing. *Cognition* 1989, 32:25–64.
2. Shillcock RC. Lexical hypotheses in continuous speech. In: Altmann G, ed. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge: MIT Press; 1990, 24–29.
3. Luce PA, Cluff MS. Delayed commitment in spoken word recognition: evidence from cross-modal priming. *Percept Psychophys* 1998, 60:484–490.
4. Tabossi P, Collina S, Mazzetti M, Zoppello M. Syllables in the processing of spoken Italian. *J Exp Psychol: Hum Percept Perform* 2000, 26:758–775.
5. Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. *Science* 1995, 268:1632–1634.
6. Whalen D. Subcategorical phonetic mismatches and lexical access. *Percept Psychophys* 1991, 50:351–360.
7. Connine CM, Blasko DG, Titone DG. Do the beginnings of spoken words have a special status in auditory word recognition? *J Mem Lang* 1993, 32:193–210.
8. Connine CM, Titone D, Deelman T, Blasko DG. Similarity mapping in spoken word recognition. *J Mem Lang* 1997, 37:463–480.
9. Gaskell MG, Marslen-Wilson WD. Phonological variation and inference in lexical access. *J Exp Psychol: Hum Percept Perform* 1996, 22:144–158.
10. Gaskell MG, Marslen-Wilson WD. Mechanisms of phonological inference. *J Exp Psychol: Hum Percept Perform* 1998, 24:380–396.
11. Gaskell MG, Marslen-Wilson WD. Representation and competition in the perception of spoken words. *Cogn Psychol* 2002, 45:220–266.
12. Andruski JE, Blumstein SE, Burton M. The effect of subphonemic differences on lexical access. *Cognition* 1994, 52:163–187.
13. Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Tracking the time course of subcategorical mismatches: evidence for lexical competition. *Lang Cogn Process* 2001, 16:507–534.
14. Marslen-Wilson WD, Warren P. Levels of perceptual representation and process in lexical access: words, phonemes and features. *Psychol Rev* 1994, 101:653–675.
15. McQueen JM, Norris D, Cutler A. Competition in spoken word recognition: spotting words in other words. *J Exp Psychol Learn Mem Cogn* 1994, 20:621–638.
16. Vitevitch MS, Luce PA. When words compete: levels of processing in spoken-word recognition. *Psychol Sci* 1998, 9:325–329.
17. Vitevitch MS, Luce PA. Probabilistic phonotactics and neighborhood activation in spoken-word recognition. *J Mem Lang* 1999, 40:374–408.
18. Luce PA, Goldinger SD, Auer ET, Jr. Vitevitch MS. Phonetic priming, neighborhood activation, and parsyn. *Percept Psychophys* 2000, 62:615–625.
19. Dumay N, Frauenfelder UH, Content A. The role of the syllable in lexical segmentation in French: word-spotting data. *Brain Lang* 2002, 81:144–161.
20. McQueen JM. Segmentation of continuous speech using phonotactics. *J Mem Lang* 1998, 39:21–46.
21. van der Lugt A. The use of sequential probabilites in the segmentation of speech. *Percept Psychophys* 2001, 63:811–823.
22. Weber A, Cutler A. First-language phonotactics in second-language listening. *J Acoust Soc Am* 2006, 119:597–607.
23. Cutler A, Butterfield S. Rhythmic cues to speech segmentation: evidence from juncture misperception. *J Mem Lang* 1992, 31:218–236.
24. Pallier C, Sebastián-Gallés N, Felguera T, Christophe A, Mehler J. Attentional allocation within the syllable structure of spoken words. *J Mem Lang* 1993, 32:373–389.
25. Vroomen J, van Zon M, de Gelder B. Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Mem Cogn* 1996, 24:744–755.
26. Quené H. Durational cues for word segmentation in Dutch. *J Phonet* 1992, 20:331–350.
27. Quené H. Segment durations and accent as cues to word segmentation in Dutch. *J Acoust Soc Am* 1993, 94:2027–2035.
28. Turk AE, Shattuck-Hufnagel S. Word-boundary related duration patterns in English. *J Phonet* 2000, 28:397–440.

29. Forster KI. Accessing the mental lexicon. In: Wales RJ, Walker EW, eds. *New Approaches to Language Mechanisms*. Amsterdam: North-Holland; 1976.

30. Morton J. The integration of information in word recognition. *Psychol Rev* 1969, 76:165–178.

31. Marr D. *Vision* San Francisco: W. H.: Freeman; 1982.

32. Norris D. How do computational models help us build better theories? In: Cutler AM, ed. *Twenty-First Century Psycholinguistics: Four Cornerstones*. NJ: Lawrence, Erlbaum; 2005.

33. Scharenborg O, Boves L. Computational modelling of spoken-word recognition processes: design choices and evaluation. *Pragmat Cogn* 2010, 18:136–164.

34. Tanenhaus MK, Magnuson J, Dahan D, Chambers C. Eye movements and lexical access in spoken-language comprehension: evaluating the linking hypothesis between fixations and linguistic processing. *J Psycholing Res* 2000, 29:557–580.

35. Marslen-Wilson WD, Tyler LK. The temporal structure of spoken language understanding. *Cognition* 1980, 8:1–71.

36. Marslen-Wilson WD, Welsh A. Processing interactions and lexical access during word recognition in continuous speech. *Cogn Psychol* 1978, 10:29–63.

37. Cole RA. Listening for mispronunciations: a measure of what we hear during speech. *Percept Psychophys* 1973, 1:153–156.

38. Taft M, Hambly G. Exploring the cohort model of spoken word recognition. *Cognition* 1986, 22:259–282.

39. Marslen-Wilson WD. Functional parallelism in spoken word-recognition. *Cognition* 1987, 25:71–102.

40. Marslen-Wilson WD. Activation, competition and frequency in lexical access. In: Altman GTM, ed. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT Press; 1990, 148–172.

41. Marslen-Wilson WD, Brown CM, Tyler LK. Lexical representations in spoken language comprehension. *Lang Cogn Process* 1988, 3:1–16.

42. Luce PA. A computational analysis of uniqueness points in auditory word recognition. *Percept Psychophys* 1986, 39:155–158.

43. Bard EG, Shillcock RC, Altmann GE. The recognition of words after their acoustic offsets in spontaneous speech: evidence of subsequent context. *Percept Psychophys* 1988, 44:395–408.

44. McClelland JL, Elman JL. The TRACE model of speech perception. *Cogn Psychol* 1986, 18:1–86.

45. McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception, Part 1: an account of basic findings. *Psychol Rev* 1981, 88:375–405.

46. Dahan D, Magnuson J, Tanenhaus M. Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cogn Psychol* 2001, 42:361–367.

47. Ganong WF. Phonetic categorization in auditory word perception. *J Exp Psychol: Hum Percept Perform* 1980, 6:110–125.

48. Foss DJ, Blank MA. Identifying the speech codes. *Cogn Psychol* 1980, 12:1–31.

49. Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J Mem Lang* 1998, 38:419–439.

50. Norris D. Shortlist: a connectionist model of continuous speech recognition. *Cognition* 1994, 52:189–234.

51. McQueen JM, Jesse A, Norris D. No lexical-prelexical feedback during speech perception or: is it time to stop playing those Christmas tapes? *J Mem Lang* 2009, 61:1–18.

52. Norris D, McQueen JM, Cutler A. Merging information in speech recognition: feedback is never necessary. *Behav Brain Sci* 2000, 23:299–370.

53. Norris D, McQueen JM. Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev* 2008, 115:357–395.

54. Magnuson JS, McMurray B, Tanenhaus MK, Aslin RN. Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cogn Sci* 2003, 27:285–298.

55. McQueen JM. The ghost of Christmas future: didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cogn Sci* 2003, 27:795–799.

56. Norris D, McQueen JM, Cutler A. Perceptual learning in speech. *Cogn Psychol* 2003, 47:204–238.

57. Cutler A, Norris D. Monitoring sentence comprehension. In: Cooper WE, Walker ECT, eds. *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale: Erlbaum; 1979.

58. Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. *J Exp Psychol Hum Percept Perform* 1988, 14:113–121.

59. Norris D, McQueen JM, Cutler A. Competition and segmentation in spoken-word recognition. *J Exp Psychol Learn Mem Cogn* 1995, 21:1209–1228.

60. Vroomen J, de Gelder B. Metrical segmentation and lexical inhibition in spoken word recognition. *J Exp Psychol Hum Percept Perform* 1995, 21:98–108.

61. Norris D, McQueen JM, Cutler A, Butterfield S. The possible-word constraint in the segmentation of continuous speech. *Cogn Psychol* 1997, 34:191–243.

62. Grosjean F. The recognition of words after their acoustic offsets: evidence and implications. *Percept Psychophys* 1985, 38:299–310.

63. Scharenborg O, Norris D, ten Bosch L, McQueen J. How should a speech recognizer work? *Cogn Sci* 2005, 29:867–918.

64. Smits R, Warner N, McQueen JM, Cutler A. Unfolding of phonetic information over time: a database of Dutch diphone perception. *J Acoust Soc Am* 2003, 113:563–574.

65. Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear Hear* 1998, 19:1–36.

66. Marslen-Wilson WD, Gaskell MG. Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *J Exp Psychol Hum Percept Perform* 2002, 28:218–244.

67. Salverda AP, Dahan D, McQueen J. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 2003, 90:51–89.

68. Shatzman KB, McQueen JM. Segment duration as a cue to word boundaries in spoken-word recognition. *Percept Psychophys* 2006, 68:1–16.

69. Cutler A. *Native Listening: Language Experience and the Recognition of Spoken Words*. Cambridge, MA: MIT Press; 2012, in press.

70. Luce RD. *Individual Choice Behavior*. Oxford: John Wiley; 1959.

71. Cluff MS, Luce PA. Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation. *J Exp Psychol Hum Percept Perform* 1990, 16:551–563.

72. Goldinger SD, Luce PA, Pisoni DB. Priming lexical neighbors of spoken words: effects of competition and inhibition. *J Mem Lang* 1989, 28:501–518.

73. Magnuson JS, Mirman D, Harris HD. Computational models of spoken word recognition. In: Spivey M, McRae K, Joanisse M, eds. *The Cambridge Handbook of Psycholinguistics*. Cambridge University Press; in press.

74. Hintzman DL. Schema abstraction in a multiple-trace memory model. *Psychol Rev* 1986, 93:411–428.

75. Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychol Rev* 1998, 105:251–279.

76. Hintzman DL, Block R, Inskeep N. Memory for mode of input. *J Verb Learn Verb Behav* 1972, 11:741–749.

77. McQueen JM, Mitterer H. Lexically-driven perceptual adjustments of vowel categories. *Poster Presented at the ISCA Workshop on Plasticity in Speech Perception*. London: 2005.

78. Pitt MA, McQueen JM. Is compensation for coarticulation mediated by the lexicon? *J Mem Lang* 1998, 39:347–370.

79. Goldinger SD. Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp Psychol Lang Mem Cogn* 1996, 22:1166–1183.

80. Gaskell MG, Marslen-Wilson WD. Integrating form and meaning: a distributed model of speech perception. *Lang Cogn Process* 1997, 12:613–656 (Special cognitive models of speech processing: psycholinguistic and computational perspectives on the lexicon).

81. Hinton GE, McClelland JL, Rumelhart DE. Distributed representations. In: Rumelhart DE, McClelland JL, eds., vol. 1. *Parallel Distributed Processing: Explorations in the Microsstructure of Cognition*. Cambridge, MA: MIT Press; 1986.

82. Smolensky P. On the proper treatment of connectionism. *Behav Brain Sci* 1988, 11:1–74.

83. Weber A, Crocker MW. On the nature of semantic constraints on lexical access. *J Psycholing Res* 2011, Advance online publication. doi:10.1007/s10936-011-9184-0.

84. Fodor JA. *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press; 1983.

85. Norris D, Cutler A, McQueen JM, Butterfield S. Phonological and conceptual activation in speech comprehension. *Cogn Psychol* 2006, 53:146–193.

86. McQueen JM. Speech perception. In: Lamberts K, Goldstone R, eds. *The Handbook of Cognition*. London: Sage Publications; 2005, 255–275.

87. Klatt DH. Speech perception: a model of acoustic-phonetic analysis and lexical access. *J Phonet* 1979, 7:279–312.

88. Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J Acoust Soc Am* 2002, 111:1872–1891.

89. Grossberg S, Myers CW. The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychol Rev* 2000, 107:735–767.

90. Massaro DW. *Perceiving Talking Faces: From Speech Perception to A Behavioral Principle*. Cambridge, MA: MIT Press; 1997.

91. Collins AM, Quillian MR. Retrieval time from semantic memory. *J Verb Learn Verb Behav* 1969, 8:240–247.

92. Smith EE, Shoben EJ, Rips LJ. Structure and process in semantic memory: a featural model for semantic decisions. *Psychol Rev* 1974, 81:214–241.

93. Collins AM, Loftus EF. A spreading-activation theory of semantic processing. *Psychol Rev* 1975, 82:407–428.

94. Anderson JR. ACT: a simple theory of complex cognition. *Am Psychol* 1996, 51:355–365.

95. Bölte J, Coenen E. Is phonological information mapped onto semantic information in a one-to-one manner? *Brain Lang* 2002, 81:384–397.

96. Butterworth B. Lexical representation. In: Butterworth B, ed., vol. 2. *Language Production*. London: Academic Press; 1983, 257–332.

97. Taft M, Forster KI. Lexical storage and retrieval of prefixed words. *J Verb Learn Verb Behav* 1975, 14:630–647.

98. Clahsen H. Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behav Brain Sci* 1999, 22:991–1060.

99. Schreuder R, Baayen H., Modeling morphological processing. In: Feldman LB, ed. *Morphological Aspects of Language Processing*. Hillsdale, NJ: Erlbaum; 1995, 131–154.

100. Marslen-Wilson WD. Access to lexical representations: cross-linguistic issues. *Lang Cogn Process* 2001, 16:699–708.

101. Marslen-Wilson WD. Morphology and language. In: Brown K, ed. *Encyclopedia of Language and Linguistics*. Oxford: Elsevier; 2006.

102. Marslen-Wilson WD, Tyler LK, Waksler R, Older L. Morphology and meaning in the English mental lexicon. *Psychol Rev* 1994, 101:3–33.

103. Ernestus M, Baayen H. Paradigmatic effects in auditory word recognition: the case of alternating voice in Dutch. *Lang Cogn Process* 2007, 22:1–24.

104. Baayen H, McQueen JM, Dijkstra T, Schreuder R. Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In: Baayen H, Schreuder R, eds. *Morphological Structure in Language Processing*. Berlin: Mouton de Gruyter; 2003, 355–390.

105. McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? *Trends Cogn Sci* 2006, 10:363–369.

106. McQueen JM, Norris D, Cutler A. Are there really interactive processes in speech perception? *Trends Cogn Sci* 2006, 10:533.

107. Goldinger SD. A complementary-systems approach to abstract and episodic speech perception. *16th International Congress of Phonetic Sciences*., Dudweiler: Pirrot; 2007, 49–54.

108. Cutler A, Weber A. Listening experience and phonetic-to-lexical mapping in L2. *16th International Congress of Phonetic Sciences*. Dudweiler: Pirrot; 2007, 43–48.

109. Norman K, O'Reilly R. Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. *Psychol Rev* 2003, 110:611–646.

110. Connine CM, Ranbom L, Patterson DJ. On the representation of phonological variant frequency in spoken word recognition. *Percept Psychophys* 2008, 70:403–411.

111. Elman JL, McClelland JL. Exploiting lawful variability in the speech wave. In: Perkell JS, Klatt DH, eds. *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.

FURTHER READING

McQueen JM. Eight questions about spoken-word recognition. In Gaskell MG, Ed. *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press; 2007, 37–53.

McQueen JM, Cutler A. Spoken word access processes: an introduction. *Lang Cogn Process* 2001, 16:469–490.