

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

Chapter 8

Spoken Word Recognition

Delphine Dahan and James S. Magnuson

1. INTRODUCTION

We solve an astounding array of information-processing challenges when we perceive a speaker's intended message. Apparently effortlessly, we accommodate variability in talker characteristics, dialect, speaking rate, and acoustic environment, all of which perturb the mapping between speech and linguistic categories. Without the aid of invariant cues to phonetic categories or word boundaries, we map acoustics onto phonetic categories, phonetic categories onto words in memory, words onto phrases and syntactic structures, words and syntax onto semantics, etc. Or do we?

On this view of language understanding, spoken word recognition is a distinct subsystem providing the interface between low-level perception and cognitive processes of retrieval, parsing, and interpretation. The narrowest conception of the process of recognizing a spoken word is that it starts from a string of phonemes, establishes how these phonemes should be grouped to form words, and passes these words onto the next level of processing. Some theories, though, take a broader view and blur the distinctions between speech perception, spoken word recognition, and sentence processing (e.g., Elman, 2004; Gaskell & Marslen-Wilson, 1997; Klatt, 1979; McClelland, St John, & Taraban, 1989).

What motivates the narrow and broad conceptions? There are empirical, pragmatic, and theoretical motivations for the narrow view. Empirically, psycholinguistic levels of processing map roughly onto linguistic levels of description. The fact that linguistic knowledge can be described as a hierarchically structured set of levels leads to the reasonable hypothesis that speakers (or signers) and perceivers may represent and operate on those structures. Indeed, this hypothesis is given face validity by the fact that humans can make decisions about levels like phonemes and words and that perception can be influenced by manipulations at those levels (though there is a long history of debate over their psychological reality; see Pisoni & Luce, 1987, for a review).

1 The pragmatic motivation for the narrow view stems from the fact that over a century
2 of concerted study of speech perception has led to a catalog of complex empirical phe-
3 nomena and candidate cues for speech perception, but little understanding of the specific
4 components of the speech signal that humans use to decode speech and achieve phonetic
5 constancy (Nusbaum & Magnuson, 1997; Remez, 2005). Rather than wait for a complete
6 understanding of early perceptual processes, psycholinguists have made significant
7 progress in understanding the processing of words and sentences by making the simpli-
8 fying assumption that a string of phonemes makes a reasonable proxy for the results of
9 initial perception, and that a series of sound forms associated with lexical entries makes
10 a reasonable proxy for the input to sentence processing.

11
12 Theoretically, the narrow view is motivated in part by the assumption that the division
13 of labor in staged systems affords significant processing efficiencies (Fodor, 1983;
14 Norris, 1994; Norris, McQueen, & Cutler, 2000). Breaking the problem into distinct
15 stages is argued to provide cognitive economy if the result is a series of mappings that
16 are straightforward relative to the complexity of the full mapping from lowest to highest
17 level (restrictions on the information available to each level are also key to the *interac-*
18 *tion vs. autonomy* debate discussed below).

19
20 The broader view of spoken word recognition (in the extreme, as the mapping from
21 speech to meaningful units that may be larger than words) has empirical and theoretical
22 motivations. One consideration is that by assuming that the input to spoken word recog-
23 nition is a string of abstract, phonemic category labels, one implicitly assumes that the
24 nonphonemic variability carried on the speech signal is not relevant for spoken word
25 recognition and higher levels of processing. However, if this variability and detail is not
26 random but is lawfully related (even partially) to linguistic categories, the simplifying
27 assumption that the output of speech perception is a string of phonemes may actually be
28 a complicating assumption. Indeed, there is growing evidence that spoken word recogni-
29 tion is influenced by information in the signal that cannot be captured in a string of
30 phonemes. For example, misleading coarticulatory cues caused by splicing the onset and
31 most of the vowel of one consonant–vowel–consonant (CVC) word or nonword onto the
32 last consonant of another CVC word (creating “subcategorical mismatches”; Whalen,
33 1984) changes the time course of lexical activation and competition (Dahan, Magnuson,
34 Tanenhaus, & Hogan, 2001a; Marslen-Wilson & Warren, 1994; McQueen, Norris, &
35 Cutler, 1999).

36
37 What purpose might this fine-grained sensitivity serve? One challenge posed by
38 assuming that words are identified from a string of phonemes is the *embedding problem*;
39 most long words have multiple shorter words embedded within their phonemic tran-
40 scriptions (e.g., depending on dialect, and neglecting all subphonemic cues, *unitary*
41 contains *you*, *unit*, *knit*, *it*, *tarry*, *air*, and *airy*) and conversely, many short words embed
42 in one or more other words (McQueen, Cutler, Briscoe, & Norris, 1995). Successful
43 spoken word recognition depends on distinguishing intended words from embeddings.
44 However, the embedding problem is significantly mitigated when subphonemic informa-
45 tion in the input is considered. For example, listeners are sensitive to very subtle

1 durational differences (in the range of 15–20 ms) that distinguish *phonemically identical*
2 syllables that occur in short words (*ham*) from those embedded in longer words (*hamster*)
3 (Salverda, Dahan, & McQueen, 2003; see also Davis, Marslen-Wilson, & Gaskell, 2002).
4

5 Thus, the bottom-up signal contains vital information that simplifies the mapping from
6 speech to words that would be lost were words identified from a string of phonemes.
7 Might the same be true for subsequent processes? There is increasing evidence that the
8 construction of syntactic and semantic structures relies on more than just a sequence of
9 words. Indeed, a sequence of words is almost always temporarily compatible with multiple
10 structures. For example, the structure associated with the word sequence *John knew*
11 *the answer* differs whether it is followed by *was wrong* or *to the question* (e.g., Altmann,
12 1999). A growing body of work has documented the role played by the prosodic structure
13 of an utterance (marked by prosodic breaks and intonational prominences) in favoring
14 some structures over others (e.g., Kjølgaard & Speer, 1999; for a review see Cutler,
15 Dahan, & van Donselaar, 1997; Speer & Blodgett, 2006, this volume). This literature
16 indicates that information from the speech signal is passed onto higher levels of processing.
17 This supports an integrated view of phonetic, lexical, and sentential processing.
18

19 Sentence-level top-down constraints on lexical activation have received some attention
20 in spoken word recognition, but chiefly with respect to how top-down information might
21 constrain the set of activated lexical items (e.g., Marslen-Wilson & Welsh, 1978;
22 Marslen-Wilson, 1987, 1990). Immediate access to syntactic, semantic, and nonlinguistic
23 context could provide significant constraints on spoken word recognition, by influencing
24 the activation of homophones, semantic associates, or context-appropriate lexical
25 items (Shillcock & Bard, 1993), helping resolve lexical ambiguity resulting from phonological
26 assimilations (Gaskell & Marslen-Wilson, 2001), or by restricting the set of
27 possible referents (Brown-Schmidt, Campana, & Tanenhaus, 2004).
28

29 Throughout this chapter, as we describe the central themes of current research on the
30 recognition of spoken words, we will adopt the more prevalent, narrow view (except
31 when noted) that most current work assumes. As we will discuss at the end of the chapter,
32 the growing evidence for subcategorical specificity may herald a dramatic shift in theories
33 of spoken word recognition. Taking the broad view – confronting the speech signal
34 itself and considering how higher levels of representation might constrain lexical access
35 – may be the key to significant progress in understanding spoken word recognition.
36

37 The recognition of a spoken word can be viewed as the process of classifying an
38 auditory stimulus as belonging to one “word-form” category, chosen from many alternatives.
39 As this description stresses, this process requires matching the spoken input with
40 mental representations associated with word candidates, and selecting one among several
41 candidates that are at least partially consistent with the input. Frauenfelder and Tyler
42 (1987) classified the functions required of any theory of spoken word recognition into
43 three stages. *Initial contact* is how input interfaces with and activates lexical representations.
44 *Selection* describes how the set of activated lexical alternatives is evaluated with
45 respect to the sensory input. *Integration* refers to how candidates are evaluated with

1 respect to the linguistic and nonlinguistic context, in order to identify which is the like-
2 liest candidate for recognition as well as to build larger linguistic structures.

3
4 Early models viewed these processes as discrete, only partially overlapping stages, in
5 particular predicting a temporal delay between access and selection (e.g., Marslen-
6 Wilson, 1987; Zwitserlood, 1989). More recent models allow for continuous uptake and
7 evaluation of the input, thus blurring functional and temporal distinctions between ac-
8 cess, selection, and integration (for behavioral evidence supporting the continuous view,
9 see Dahan & Tanenhaus, 2004). Nonetheless, the theoretical distinctions are useful as
10 each process poses challenges of a different nature.

11
12 In the course of this chapter, we will review the central issues pertaining to contact and
13 selection in turn, and how they have been conceptualized in different models and theo-
14 ries of spoken word recognition. We will also address whether categorizing the input as
15 a member of a word category changes listeners' percept of the input. This question hinges
16 on the architecture of the processing system, i.e., whether higher levels of representations
17 (such as words) can affect lower levels, such as speech sounds or phonemes. Finally, we
18 will briefly review integration, and close with a discussion of what we see as the most
19 crucial challenges to theories of spoken word recognition and spoken language generally,
20 and the approaches we find most promising and most likely to lead to solutions.

21 22 23 **2. INITIAL CONTACT**

24
25 When someone speaks, the linguistic content and speaker characteristics (e.g., physi-
26 ology of the vocal tract, gender, regional origin, emotions, identity) simultaneously in-
27 fluence the acoustics of the resulting spoken output. Additional sources of variability
28 include rate of elocution, prosodic prominence, and the phonetic context in which each
29 word is pronounced. Nonetheless, listeners are able to recognize acoustically different
30 stimuli as instances of the same word, thus extracting the similarity that exists between
31 these different tokens, and perceiving them as members of the same category. How are
32 words mentally represented to allow for this complex categorization?

33
34 The traditional (and dominant) view assumes that people represent the form of words
35 as categories that abstract away from variability. Drawing on linguistic theories, the
36 mental representation of a word form is usually conceived as a sequence of phonemes
37 (sometimes themselves decomposed into a bundle of contrastive features). Within this
38 framework, the ease with which a given pronunciation is categorized as a token of a given
39 word is assumed to depend upon the degree to which its components have characteristics
40 typically associated with the word's phonemes. Speaker-specific information is often
41 viewed as a source of noise which does not contribute to the process of identifying the
42 linguistic units present in the signal.

43
44 This view has not gone uncontested. An episodic view, most forcefully argued for by
45 Goldinger (1996, 1998), conceptualizes lexical representations as ensembles of detailed

1 memory traces (or episodes) of word instances. Several recognition memory studies have
 2 shown that people implicitly retain in memory nonlinguistic aspects of spoken words
 3 (e.g., Hintzman, Block, & Inskip, 1972; Goldinger, 1996, 1998; Palmeri, Goldinger, &
 4 Pisoni, 1993). The question at stake is whether these memory traces of words constitute
 5 the knowledge that people access and use when identifying spoken words. Goldinger
 6 (1998) applied (Hintzman's) (1986) MINERVA2 model of episodic memory to spoken
 7 word recognition. In this model, a speech episode (a word) is simultaneously compared
 8 to all memory traces. Activation of a trace is proportional to its acoustic similarity with
 9 the stimulus. The aggregate of all activated traces (the so-called *echo*) is sent to working
 10 memory and corresponds to the listener's percept. Because the echo consists of a blend
 11 of the memory traces that resemble the stimulus, it tends to capture the aspects that are
 12 common among the traces but not the aspects that differ. This principle enables the model
 13 to make generalizations and categorize new tokens without assuming the existence of
 14 abstract mental categories. A critical challenge to the episodic view is how the similarity
 15 between an actual speech stimulus and memory traces would be computed, if no normal-
 16 ization or other data-reducing process abstracting from surface variability is assumed.
 17 Goldinger's model has thus far assumed word-length episodes and remains agnostic
 18 about how words would be isolated from the utterances they are embedded in, which is
 19 problematic given the challenges posed by word segmentation (see below). Given its rad-
 20 ical departure from classical approaches, this theory may well have the potential to bring
 21 new leverage to problems of speech perception and spoken word recognition. However,
 22 until similarity mapping and segmentation are spelled out, the episodic view faces the
 23 same challenges as the traditional, abstract view.¹

AQ1

AQ2

24
 25 The traditional view has influenced much of the research on spoken word recognition.
 26 Thus, the recognition of a spoken word is generally viewed as the mapping of the speech
 27 input onto abstract lexical representations, with abstract units standing for the word's
 28 subcomponents, the phonemes, mediating this mapping. An extended line of research has
 29 documented how listeners accommodate the variability inherent to speech rate and pho-
 30 netic context in the perception and recognition of individual phonemes (Miller &
 31 Liberman, 1979). We will not review this literature here, but rather will focus on how the-
 32 ories of spoken word recognition have embodied, or sometimes departed from, the clas-
 33 sical approach to spoken word recognition.

34
 35
 36
 37
 38 ¹ Goldinger's (1998) simulations have two critical problems. The model assumes the input is pre-segmented into
 39 word-length episodes (the primitive unit), which are represented as vectors of units (with values of -1, 0, or
 40 +1), with some units representing the word type, and others representing talker and context information. While
 41 Goldinger claimed such a model can achieve phonetic constancy without normalization of talker differences,
 42 the solution depends on this unrealistic assumption about the input. In real speech, talker variability conditions
 43 phonetic realization. In Goldinger's simulations, the input is in effect pre-normalized. The episodic model's
 44 promise to solve phonetic constancy without normalization may be possible, but tests with more realistic input
 45 are needed to evaluate it (see Goldinger & Azuma, 2003, for a discussion of how adaptive resonance theory may
 provide the means to test the theory with more realistic inputs and mechanisms).

2.1. Initial Contact and Similarity Metrics

The first question for any model is the nature of the input representation: How do the products of sensory information interface with the lexicon? As mentioned earlier, the input to word recognition has traditionally been assumed to be a string of phonemes, output by a speech perception system (as in the original COHORT model; Marslen-Wilson & Welsh, 1978). This representation was also adopted by the model SHORTLIST (Norris, 1994), although mainly for practical reasons. However, the string-of-phonemes encoding of the speech input assumes that subphonemic variation in the signal is lost, while such variation has been shown to affect listeners' word recognition. For example, Andruski, Blumstein, and Burton (1994) demonstrated that, as the realization of the initial segment of a word like *king* is modified as to differ from a prototypical /k/ (by shortening the duration of the stop voice onset time) but not enough to change the ultimate categorization of this segment, people are nonetheless less likely to categorize the word as an instance of *king* (see McMurray, Tanenhaus, & Aslin, 2002, for converging evidence; for demonstrations of listeners' sensitivity to subcategorical cues in vowels, see (Dahan et al., 2001a,b); Marslen-Wilson & Warren, 1994; Whalen, 1984). Evidence for graded activation of words based on subphonemic similarity requires a finer representational grain than phonemes.

AQ3

Another issue related to the string-of-phonemes assumption is that it imposes a dissociation between the process of recognizing words from that of recognizing its components. A recent attempt to add an automatic phone recognizer to the SHORTLIST model exposed the limitations of this assumption (Scharenborg, ten Bosch, Boves, & Norris, 2003). In these simulations, the automatic phone recognizer took real speech (naturalistic speech samples from telephone conversations) as input and generated a sequence of phonemes. From this string of phonemes, the activation and competition mechanisms implemented in the SHORTLIST model yielded the best matching word candidate. Word recognition accuracy performance was poor (around 25%), but improved considerably when one of the model's parameters, the penalty assigned to candidates that mismatch the phonemic input, was set to zero. This result may be interpreted as evidence that SHORTLIST, originally tested on an unrealistically accurate phonemic input, must be revised to accommodate likely erroneous input from a phone recognizer. On the other hand, this result can be taken as reflecting the shortcoming of a phonemic string as input. If "hard" phonemic decisions are made by the phone recognizer, the fact that other phonemic interpretations were substantially supported by the signal is lost. Most of all, these simulations illustrate how much the modeling of spoken word recognition hinges on assumptions about the input representation.

The next simplest solution is to assume that the input takes the form of localist phoneme activation units (as in the REVISED COHORT model (Marslen-Wilson, 1987, 1989), and MERGE (Norris et al., 2000)). Subphonemic detail can be approximated in the distributed representation afforded by the entire set of phonemes. Thus, a segment ambiguous between /k/ and /g/ can be represented by partial activation of both units. A slightly more fine-grained representation can be achieved with units representing a set of (usually binary) acoustic-phonetic features (as in the DISTRIBUTED COHORT MODEL

1 (Gaskell & Marslen-Wilson, 1997)). However, the realism of these schemes is limited, as
2 they fail to incorporate a critical aspect of speech, coarticulation.

3
4 As demonstrated by the seminal work by Liberman, Cooper, Shankweiler, and
5 Studdert-Kennedy (1967) and contrary to a listener's subjective impression, a spoken
6 utterance is not a concatenated sequence of discrete speech sounds. The gestures involved
7 in the articulation of each sound overlap temporally with the gestures that generate adja-
8 cent sounds. One of the consequences of this temporal overlap has been coined the "seg-
9 mentation" problem. A spoken utterance cannot be divided into smaller portions, each
10 one representing a single segment.² If the recognition of a spoken word involves the map-
11 ping of the input onto word representations where segments are in temporal order, the lis-
12 tener must assign the presence of a given acoustic feature in the input to a given segment.

13
14 The TRACE model (McClelland & Elman, 1986) uses the most complex input repre-
15 sentations of any current model of speech perception and spoken word recognition. The
16 input is a "pseudo-spectral" representation based on seven acoustic-phonetic features,
17 each represented with a nine-unit continuous vector, which encode the degree to which
18 the feature is represented in the input. Features spread over time by ramping up to a
19 phoneme center and then ramping off. Phoneme centers are close enough together and
20 features spread far enough that there is substantial overlap between phonemes, creating
21 a rough analog to coarticulation.

22
23 TRACE's architecture is also a critical aspect in the way it accounts for the processing of
24 coarticulated speech. In TRACE, units that stand for hypotheses at the featural, phonemic, or
25 word level, are replicated every three time slices. Each unit stands for a linguistic unit
26 potentially present in the input at a different point in time. The extensive unit reduplication
27 has often been criticized as an implausible feature of the model (beginning with McClelland
28 & Elman, 1986). However, this is central to solving the segmentation issue, as it accom-
29 modates the fact that features that result from the overlap of articulatory gestures coincide in
30 time. A given time slice may provide evidence supporting different phonemes, thus activat-
31 ing several incompatible phoneme units. However, within each level, units that span the
32 same portion of the input inhibit each other. Consequently, the phoneme hypothesis for
33 which the evidence is the strongest can win the competition. Thus, TRACE's architecture al-
34 lows the segmentation of coarticulated speech into a sequence of discrete segments.

35
36 However, TRACE's input scheme provides a very rough approximation of coarticulation
37 in real speech. While it accommodates the temporal overlap of gestures, it fails to
38 accommodate the fact that this temporal overlap affects the articulatory (and therefore
39 acoustic) realization of segments (i.e., the "lack of invariance" issue, cf. Liberman et al.,
40 1967). There have been very few even moderately successful attempts to devise

41
42
43
44
45

² This is true despite the fact that the duration of single segments, such as consonants or vowels, are often reported. Such segmentation is based on conventions on how to define boundaries between segments based on their relative prominence (see Fowler, 1984).

1 psychologically tractable models that work directly on the actual speech signal or a min-
 2 imally transformed speech signal. (The hidden-Markov models and similar mechanisms
 3 used in automatic speech recognition systems arguably substitute the black box of the
 4 brain with largely opaque statistical approximations; see Nusbaum & Magnuson, 1997,
 5 for discussion.) Klatt's (1979) LEXICAL ACCESS FROM SPECTRA (LAFS) model is perhaps
 6 the best known, but the mapping from spectra to lexical items is at least as variable as the
 7 mapping from speech to phonemes. Work in the adaptive resonance framework has grap-
 8 pled with real speech signals (the ARTSTREAM model; Grossberg, Govindarajan, Wyse, &
 9 Cohen, 2004) but has yet to be extended to the recognition of phonemic or lexical forms.
 10 The strategy of Plaut and Kello (1999) may well be the best hope for progress toward
 11 more realistic input. They use a collection of articulatory and acoustic cues that *might*
 12 turn out to be tractable to extract from speech (auditory and visual cues to jaw move-
 13 ments, changes in formants, etc.), and in combination, *might* prove a sufficient basis for
 14 speech perception and spoken word recognition.

15 16 2.2. Initial Constraints on Activation

17
18 Theories differ on the patterns of activation that follow initial contact. More specifi-
 19 cally, they differ in the theories of similarity they assume. The ORIGINAL (Marslen-Wilson
 20 & Welsh, 1978), REVISED (Marslen-Wilson, 1987, 1989) and DISTRIBUTED COHORT mod-
 21 els (Gaskell & Marslen-Wilson, 1997, 1999, 2002) place great emphasis on word onsets.
 22 The real-time constraints of the speech signal motivate an emphasis on optimal use of
 23 bottom-up information as it becomes available. Since a word's onset is heard first, it
 24 should determine which lexical items are first activated. Thus, in the original COHORT
 25 model, the set of activated lexical alternatives was constrained to a *word-initial cohort* of
 26 items that matched perfectly the phonemic representation of the first approximately 150
 27 ms of a word's onset. In light of evidence that a word might be recognized even when its
 28 first sounds are altered (for example, due to mispronunciation, cf. Cole, 1973), the
 29 revised and DISTRIBUTED COHORT models abandon the strict, all-or-none match constraint.
 30 Instead, lexical representations are activated as a function of their similarity to a spoken
 31 word, with this similarity being continuously evaluated rather than limited to the initial
 32 portion of the spoken word. Nonetheless, the models' emphasis on real-time processing
 33 maintains a special status to the spoken word's initial sounds, as they contribute to the ac-
 34 tivation of some words, and thereby the interpretation of subsequent spoken material will
 35 be biased in favor of these words (see the discussion of *Selection* below for a full de-
 36 scription of how these biases might be implemented).

37
38 The NEIGHBORHOOD ACTIVATION MODEL (NAM; Luce, 1986; Luce, Pisoni, & Goldinger,
 39 1990; Luce & Pisoni, 1998) differs from any instantiation of the COHORT model by pre-
 40 dicting activation of words that reflects their global similarity with the spoken word.³

41
42
43 ³ NAM is not a processing model *per se* – it is more properly considered a formal similarity model. However,
 44 its similarity metric imposes significant constraints on an underlying processing mechanism, and as such, it is
 45 appropriate to consider what NAM predicts in terms of lexical activation.

1 Two similarity metrics were developed within the model. The more complex one is
2 derived from observed similarity measures, such as position-specific diphone confusion
3 probabilities. Similarity between the spoken word and other words is computed by
4 comparing the confusability of each of its segments with other words' segments in the
5 same position within the word. Similarity is gradient, although limited to words that have
6 the same number of segments. The simpler metric, sometimes called the *one-phoneme*
7 *shortcut metric*, distinguishes words that are predicted to become activated during the
8 perception of a spoken word (i.e., its neighbors) from those that are not, with no
9 gradiency in the degree of activation of the former. Activated words (i.e., neighbors of the
10 spoken words) are defined as words that differ from the spoken word by no more than
11 one phoneme, whether by substitution, deletion, or addition, in any position. Thus,
12 neighbors of *cat* include *bat*, *kit* and *cap* (substitutions), *at* (deletion), and *scat* and *cast*
13 (additions).
14

15 The COHORT and NEIGHBORHOOD models make different predictions about what items
16 may be activated by a spoken word. The COHORT model predicts that hearing *cat* also
17 activates *castle* but should activate *bat* to a negligible degree. NAM predicts that *cat* will
18 activate *bat* but not *castle*, as it differs by too many phonemes. There is empirical sup-
19 port for each prediction. Marslen-Wilson (1993) reported a series of studies in which the
20 auditory presentation of a word primes visual lexical decisions to semantic associates of
21 words overlapping in onset, but not in rhyme (e.g., *beaker* would prime *insect*, an
22 associate of *beetle*, but not *stereo*, an associate of *speaker*). But Luce and Pisoni (1998)
23 reported that neighborhoods based on global similarity provide the best prediction of pro-
24 cessing time for large sets of words in tasks like lexical decision and naming, although
25 they did not separate out the contribution of the cohort-type neighbors from that of non-
26 cohort ones (we discuss this result further in the *selection* section).
27

28 TRACE makes an intermediate prediction: It activates both onset- and rhyme-overlap-
29 ping words, because, as in the NEIGHBORHOOD model, words can be activated even if
30 they mismatch at onset. However, unlike the NEIGHBORHOOD model, TRACE represents
31 time: Words that become activated early in the spoken input have an advantage over
32 words that become activated later, because more of the spoken word has been heard and
33 selection mechanisms are then more effective at favoring the best matching candidate.
34 Thus, TRACE predicts activation of both onset- and rhyme-overlapping candidates,
35 although at different times and of different amplitude. Allopenna, Magnuson, and
36 Tanenhaus (1998) provided behavioral data supporting this prediction. They estimated
37 lexical activation to word candidates by monitoring eye movements to pictures as par-
38 ticipants followed verbal instructions to move an item on a computer screen. Fixations
39 were closely time-locked to the speech (with a lag only slightly larger than that
40 required to plan and launch an eye movement), and mapped closely onto phonetic sim-
41 ilarity over time (with higher and earlier fixation proportions to onset-overlapping com-
42 petitor than rhyme-overlapping competitor) as well as response probabilities generated
43 by TRACE. This study highlights the importance of a measure of lexical activation over
44 time, given the rapid evolution of lexical activation over time, as the spoken input is
45 heard.

1 The Allopenna et al. (1998) study highlights one shortcoming of the similarity model
 2 embodied in NAM to the study of spoken word recognition. The temporal distribution of
 3 similarity is not considered; *dab* and *bad* are assumed to be equally active upon hearing
 4 *dad* (ignoring frequency for the sake of the example). NAM fails to capture the temporal
 5 dimension of speech and the special status that the initial sounds have due to their tem-
 6 poral precedence (Marslen-Wilson & Zwitserlood, 1989). It also gives too much weight
 7 to the match in the number of segments or syllabic structure by entirely excluding the
 8 contribution of words that are more than one phoneme longer than the word to be recog-
 9 nized, despite evidence suggesting that words of different lengths affect the processing of
 10 a given word (Marslen-Wilson, 1984, 1987). The algorithm cannot be easily extended to
 11 the computation of competition environment for polysyllabic words, as most of these
 12 words have very few, if any, competitors under the one-phoneme difference definition.⁴
 13 Finally, the one-phoneme shortcut metric, which has been most widely used by
 14 researchers and has proven useful in stimulus selection and experimental control, treats
 15 any phoneme deviation equally, regardless of its phonetic nature. Confusion between two
 16 words differing by one-phoneme addition or substitution, or confusion between two
 17 words differing by a vowel or a consonant, are all assumed to be equivalent, despite em-
 18 pirical evidence that the nature of the phonetic feature(s) that differ between two words
 19 is an important factor in accounting for word confusions (e.g., Bailey & Hahn, 2005;
 20 Hahn & Bailey, 2004; see also van Ooijen, 1996).⁵

21 2.3. Plasticity in Mapping the Speech Signal onto the Lexicon_

22
 23 As pointed out in the introduction to this section, the acoustic form that a given word
 24 takes can vary greatly. Nonetheless, listeners have little difficulty accommodating this
 25 variability, which has sometimes been interpreted as reflecting plasticity in the mapping
 26 of the speech signal onto the lexicon. Here we review some of this work.
 27
 28

29 A substantial number of studies have examined the processing of spoken words that
 30 have undergone phonological assimilation. In connected speech, the value of a segment's
 31 feature (i.e., place of articulation or voicing) may assimilate to that of the same feature
 32 from its surrounding segments. For instance, the place of articulation of the final alveo-
 33 lar sound of the word *green*, may be altered to become (or approach) the bilabial place
 34 of articulation of the initial sound of the subsequent word *boat*, so that the sequence may
 35 sound a little like *gream boat*. The conditions under which assimilation may occur are dic-
 36 tated by the phonology of the language. Research on the perception of assimilated words
 37
 38
 39

40 ⁴ Cluff and Luce (1990) used the one-phoneme difference algorithm to compute the competition environment
 41 of bisyllabic words composed of two monosyllabic words (e.g., jigsaw) by establishing competitors for each
 42 syllable independently, thereby considering only monosyllabic competitors.

43 ⁵ Luce, Goldinger, Auer, and Vitevitch (2000) report examples of cases where a more complex metric, based on
 44 positional similarity ratings, makes distinctly different predictions than the one-phoneme shortcut metric, e.g.,
 45 predicting competition between *veer* and *bull* due to high similarity at every segment despite no complete
 phoneme matches.

1 has shown that this deviation does not preclude the identification of the assimilated token
2 as an instance of the intended word. Gaskell and colleagues (Gaskell, 2003; Gaskell,
3 Hare, & Marslen-Wilson, 1995; Gaskell & Marslen-Wilson, 1996, 1998) have suggested
4 that listeners have learned to accept the assimilated form as a token of the intended word
5 in the appropriate context, especially if the assimilation was only partial (thus, maintain-
6 ing some of the acoustic characteristics of the original segment). This proposal was sup-
7 ported by simulations from a connectionist model that was trained to learn to map
8 acoustically variable (but arguably simplified) input onto canonical, fixed representation
9 of words (see Gow, 2001, 2002, 2003a, 2003b, for a critique of Gaskell and colleagues'
10 proposal and for a competing account of the perception of assimilated words).
11

12 Other rule-based variations do not involve a phonemic or subphonemic alteration.
13 These pronunciations are characteristic of casual (as opposed to careful) speech, and
14 often described as including an atypical or reduced realization of some of the segments
15 of words. For example, the final consonant of the word *flute* can be realized with an alve-
16 lar closure and an audible release (the typical realization of the phoneme /t/), or realized
17 as a glottal stop, with no release. Similarly, the vowel of the unstressed syllable of a poly-
18 syllabic word can be so drastically reduced that it is not acoustically present in the sig-
19 nal (e.g., *police*, pronounced roughly as [plis]). How can two fairly different realizations
20 be interpreted as instances of the same word? Do people represent the multiple forms that
21 a word can take in order to accommodate such variation? And if so, at what level of
22 abstraction are they represented, and does frequency of occurrence of variants determine
23 whether a variant is represented or not?
24

25 Recent work has addressed these questions by examining whether variations are
26 equally effective at making contact with the lexical representation of the intended word
27 (or, put slightly differently, whether variations are equally categorized as members of the
28 intended word category). Some studies have probed the degree to which the meaning of
29 the word becomes available (e.g., Deelman & Connine, 2001; Sumner & Samuel, 2005).
30 Other studies have examined the degree to which variants map onto the same or different
31 representations by assessing whether having heard one variant facilitates the subsequent
32 processing of the alternative (e.g., LoCasto & Connine, 2002; McLennan, Luce, &
33 Charles-Luce, 2003; Sumner & Samuel, 2005; Utman, Blumstein, & Burton, 2000). The
34 findings that emerge from these studies are complex and often conflicting. Some results
35 suggest that any variant facilitates the processing of any alternative, which is sometimes
36 interpreted as evidence for a single, abstract and general representation; other results
37 argue for specificity. Some researchers found evidence for a special status of the most
38 frequent variant (Connine, 2004), while others did not (Sumner & Samuel, 2005). This
39 line of research has only begun, and it is too early to draw definite conclusions.
40 Nonetheless, one aspect that has been little considered is the relevance of the context in
41 which these variants occur. Listeners may be sensitive to how likely and expected a given
42 variation is, given what is known of the talker's speaking style, speed of elocution, and
43 perhaps geographic or dialectal origin. Such expectations (or use of context) may deter-
44 mine the degree to which a token will be mapped onto or activate representation(s)
45 associated with the intended word.

1 Indeed, we know that listeners do adapt to the characteristics of the talker or speech
 2 that they hear. Evidence of such adaptation comes from studies showing that word iden-
 3 tification is impaired by trial-to-trial changes in the voice of the talker (Mullennix, Pisoni,
 4 & Martin, 1989; Nusbaum & Morin, 1992) and/or in his/her speaking rate (Sommers,
 5 Nygaard, & Pisoni, 1994), and from studies showing advantage for the identification of
 6 words spoken in a familiar vs. unfamiliar voice ((Nygaard), Sommers, & Pisoni, 1994; **AQ4**
 7 although see Luce & Lyons, 1998). This suggests plasticity in the process of perceiving
 8 and interpreting speech. Listeners' ability to adapt to the characteristics of the speech or
 9 the talker they are exposed to has long been acknowledged (e.g., Joos, 1948; Ladefoged
 10 & Broadbent, 1957; Peterson & Barney, 1952). More recently, a number of studies have
 11 documented how adaptation to distorted or foreign-accented speech proceeds. The
 12 process appears to operate quite rapidly, with measurable improvement in comprehension
 13 observed after as little as two to four sentences (Clarke & Garrett, 2004). Importantly,
 14 from relatively short exposure to distorted speech, people acquire knowledge that can
 15 generalize to sentences containing unheard words (Davis, Johnsruide, Hervais-Adelman,
 16 Taylor, & McGettigan, 2005; Greenspan, Nusbaum, & Pisoni, 1988), or to similarly
 17 distorted speech from a different talker (e.g., Dupoux & Green, 1997). Furthermore,
 18 listeners' perceptual adaptation to unusual speech or talker characteristics seems to be (at
 19 least largely) mediated by lexical knowledge. Listeners who were exposed to 20 distorted
 20 nonsense sentences prior to testing on sensible sentences fared no better than people with
 21 no prior exposure to distorted speech (Davis et al., 2005; for similar conclusions, see
 22 Eisner & McQueen, 2005; Norris, McQueen, & Cutler, 2003).

23
 24 Evidence for plasticity in the mapping of spoken input onto lexical representations
 25 may help explain how listeners cope with the extreme variability found in speech. So
 26 long as this variability is context-dependent, and thus lawful, prior (even brief) exposure
 27 to speech from a new talker may trigger the learning of a new mapping between speech
 28 input and linguistic units.

31 **2.4. The Interaction Debate: Is the Interface Bidirectional?**

32
 33 Is the architecture underlying spoken word recognition autonomous (feedforward
 34 only) or interactive (lexical representations feed information back over the interface to
 35 the sublexical representations)? Bottom-up and top-down information *is* integrated: The
 36 literature is full of examples of lexical effects on tasks that tap sublexical representations.
 37 Phonemes are detected more quickly in words than nonwords (the word superiority ef-
 38 fect; Rubin, Turvey, & Van Gelder, 1976). Listeners report hearing phonemes consistent
 39 with lexical or sentential context in locations completely replaced with noise (the
 40 phoneme restoration effect; e.g., Warren, 1970; Samuel, 1981, 1997). If a phoneme con-
 41 tinuum is attached to a context that makes one endpoint a word and the other a nonword
 42 (e.g., /t/-/d/ attached to *-ash* or *-ask*), categorical perception boundaries shift such that
 43 more steps are identified as consistent with the lexical endpoint (Ganong, 1980; a bias is
 44 also found in word-word contexts with a frequency differential; Fox, 1984). Helpful
 45 visual contexts are integrated quickly to resolve ambiguities in sentence processing

1 (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The crux of the interaction
2 debate is *when* integration occurs.

3
4 This debate has recently taken center stage in spoken word recognition research, hav-
5 ing been energized by forceful empirical and theoretical arguments for autonomous mod-
6 els of spoken word recognition by Norris et al. (2000, 2003). In brief, the *autonomous*
7 *view* is that processing stages can be optimized by allowing them access only to bottom-
8 up information (disallowing interaction of top-down information). This view of stages
9 within a processing system is related to arguments for modularity between processing
10 systems (Fodor, 1983). In both cases, the idea is that veridical perception depends upon
11 transparent processing of the incoming signal. On this view, if top-down information is
12 integrated directly with sensory information, an organism *ipso facto* loses the possibility
13 of veridical perception, as there is no distinction between information in the environment
14 and information in the organism. Autonomous models account for lexical effects on sub-
15 lexical tasks by proposing parallel, competing lexical and sublexical routes (as in the
16 Race model; Cutler & Norris, 1979), or that the locus of sublexical decisions is, counter-
17 intuitively, post-lexical. In the Merge model (Norris et al., 2000), for example, there are
18 two banks of phoneme units. One is the source of bottom-up input to the lexical layer.
19 The second receives input from the bottom-up phoneme nodes and the lexical nodes. This
20 decision layer can thus integrate lexical and phonological knowledge without changing
21 the prelexical interpretation of the sensory input. The separate bank of decision nodes is
22 justified on the grounds that phonemic awareness is a late-developing artifact of learning
23 to read, based on evidence that phonemic awareness does not develop if one does not
24 learn to read (see Norris et al., 2000; Marslen-Wilson & Warren, 1994, for discussion; but
25 there is evidence that sublexical awareness (if not precisely phonemic) does emerge in
26 preliterate children (see Liberman, Shankweiler, Fischer, & Carter, 1974) and illiterate
27 adults (Bertelson & de Gelder, 1989); see Shankweiler & Fowler, 2004, for a review).
28 Falsifying this position would require showing that top-down lexical effects have a
29 perceptual, rather than decisional, locus.

30
31 On the *interactive view*, if top-down information *can* usefully constrain interpretation
32 of bottom-up information, it should be used, and veridical perception can be maintained
33 by properly weighting bottom-up and top-down information. Falsifying this position is
34 more difficult. Alternative explanations for lexical effects must be proposed, and evi-
35 dence must show that when those explanations make predictions that are different from
36 lexical feedback predictions, the lexical feedback predictions are incorrect. Over the past
37 two decades, the debate appeared to be settled at least two or three times, with alterna-
38 tive apparent falsifications of autonomous and interactive positions.

39
40 Elman and McClelland (1988) seemingly falsified the autonomous position, by
41 showing lexical effects on sublexical processing rather than sublexical decisions. They
42 conducted a study designed to demonstrate *lexically mediated compensation for coartic-*
43 *ulation*. Compensation for coarticulation (Mann & Repp, 1981) refers to the fact that in
44 normal production, if a segment with a front place of articulation follows one further
45 back (or vice versa), physical and temporal constraints may prevent the articulation from

1 reaching its ideal location, with the result that in this context, the front segment will have
2 a place of articulation further back than normal. When a front-back continuum (e.g., /t/
3 /k/) is presented following a back segment (e.g., /Σ/) the category boundary shifts toward
4 the back (i.e., more steps on the continuum are identified as the +front segment /t/), and
5 the opposite happens after a front segment (e.g., /s/). In Elman and McClelland's (1988)
6 study, this low-level perceptual phenomenon was coupled with the Ganong (1980) effect.
7 The Ganong effect shows that the interpretation of an ambiguous sound (symbolized by
8 ?, intermediate between, e.g., *p* and *b*) embedded in a larger spoken stimulus (e.g., ?*eace*)
9 is biased toward the interpretation that turns the spoken stimulus into a real word (e.g.,
10 *peace*). Elman and McClelland (1988) reasoned that if the basis for the Ganong effect is
11 feedback to the perceptual level, a restored phoneme in that paradigm should have similar
12 *consequences* as an intact phoneme, and in particular, it should drive compensation
13 for coarticulation. They found exactly that result: the boundary of a *tapes-capes* contin-
14 uum shifted following a segment ambiguous between /s/ and /Σ/ as a function of the lex-
15 ical bias preceding the ambiguous segment (e.g., *Christma-* or *fooli-*). For the next
16 decade, many regarded this as strong evidence in support of interaction.

17
18 However, Pitt and McQueen (1998) explored the hypothesis that the basis for the ef-
19 fect was diphone transitional probabilities (TPs), based on an analysis by Cairns,
20 Shillcock, Chater, and Levy (1995), purportedly showing that Elman and McClelland's
21 lexical contexts were confounded with TP. Under the TP hypothesis, compensation for
22 coarticulation after a segment ambiguous between /s/ and /Σ/ is driven by the higher
23 probability of /s/ after the final vowel of *Christma*, /↔/, than after the final vowel of *fooli*,
24 /l/, and, conversely, the higher probability of /Σ/ after /l/ than after /↔/. Because these
25 transitional probabilities can be viewed as involving sublexical knowledge only, Elman
26 and McClelland's (1986) results would not be proof of lexical influence on sublexical
27 processing. Pitt and McQueen directly tested this hypothesis and found compensation for
28 coarticulation with nonword contexts as a function of TP, but failed to find it in lexical
29 contexts where TP was controlled. For the next several years, this was regarded by many
30 as strong evidence that TP was the basis for "lexically" mediated compensation for coar-
31 ticulation.

32
33 Samuel and Pitt (2003) provided a thorough empirical and acoustic analysis of the
34 paradigm. They reported new studies in which they found lexically mediated compensa-
35 tion for coarticulation with several contexts with opposite lexical and diphone TP biases.
36 They also provided plausible perceptual explanations for the minority of cases where
37 lexically mediated compensation for coarticulation has not been found (e.g., Pitt &
38 McQueen, 1998; and some contexts tested by Samuel and Pitt themselves). Magnuson,
39 McMurray, Tanenhaus, and Aslin (2003a) reported converging evidence as well as a new
40 corpus analysis of transitional probabilities in American English that revealed that not all
41 of Elman and McClelland's lexical contexts were confounded with diphone TP. They also
42 used corpus analyses to show that no particular *n*-phone TP could predict observed lex-
43 ical effects. Instead, the appropriate TP context seems to be an *n*-phone of dynamic length,
44 where *n* resolves to word length, and thus the knowledge driving mediated compensation
45 for coarticulation seems to be lexical

1 Further evidence for feedback comes from *selective adaptation to restored phonemes*.
 2 Samuel (1997, 2001a, b) has shown that “restored” phonemes (phonemes replaced with
 3 noise, but which subjects report hearing in a manner consistent with lexical or larger con-
 4 texts) can drive the selective adaptation found with fully articulated phonemes. If a seg-
 5 ment at one end of a categorical perception continuum is repeated many times, the
 6 boundary shifts toward that stimulus, such that a smaller step toward the opposite end of
 7 the continuum leads to a change in perception. Restored phonemes have similar (though
 8 weaker) effects, suggesting the locus is prelexical.
 9

10 Norris et al. (2003) added a new wrinkle to the debate. Based on evidence for short-
 11 term changes in phonemic categories based on implicit perceptual learning, they
 12 acknowledged the need for feedback, but argued that it need not occur on-line. Instead,
 13 they make a distinction between on-line feedback (as in interactive models) and *feedback*
 14 *for learning*, although without specifying how feedback for learning is triggered or
 15 timed; if it is not to happen during processing, the learning signal must be stored until
 16 some opportune “down-time” during which the learning signal may be transmitted. The
 17 idea is that since (according to their arguments) on-line feedback can serve no useful role,
 18 and since a principled division can be made between on-line and “for learning” feedback
 19 in computational models, the most parsimonious account remains an autonomous model
 20 with feedback for learning. Norris et al. acknowledge the possibility that feedback might
 21 be implemented in such a way that it simultaneously provides on-line and for-learning
 22 feedback (see (Mirman), McClelland, & Holt, in press, for just such an implementation, **AQ5**
 23 which incorporates Hebbian learning into TRACE), but again, that such an architecture is
 24 not necessary; on this view, on-line feedback might exist, but only because it either
 25 allows a convenient medium for or is an epiphenomenon of feedback-for-learning.
 26

27 One might argue that in light of the added complexity of post-perceptual decision units
 28 in Merge (cf. Samuel, 2001a), the need for feedback to account for perceptual learning,
 29 and the ability of a single feedback system to incorporate on-line feedback (accounting
 30 for lexical effects on phonemes) and feedback for learning, interaction provides the more
 31 parsimonious account. However, given the alternative explanations for the empirical
 32 record provided by Norris et al. (2000, 2003), along with their evolving theoretical per-
 33 spective, there remains room for reasonable disagreement on this debate. Stronger theo-
 34 retical and empirical cases are required to settle it.
 35

37 **3. SELECTION: HOW IS ACTIVATION REGULATED AND RECOGNITION** 38 **ACHIEVED?** 39

40 Once the activation set is specified, a mechanism is needed to evaluate the items in the
 41 set and eventually *select* an item for lexical access (and a comprehensive theory must also
 42 specify under what conditions selection will fail to occur, e.g., in the case of a nonword
 43 input). All current theories assume that a form of competition is required for selection.
 44 As a spoken word is heard, multiple lexical items are considered as a function of their
 45 phonological similarity to the input and of their frequency of occurrence, or prior

1 probability, and activated lexical items compete for selection. The two key factors we will
2 discuss here are the role of frequency and a sampling of the competition mechanisms pro-
3 posed under different theories. We will also include a discussion on the issue related to
4 recognizing words in utterances (i.e., the word segmentation issue), as it requires com-
5 petition among incompatible hypotheses (those that claim the same portion of the input).
6

7 8 **3.1. Frequency**

9 It has long been established that words that occur frequently in the language (as
10 reflected by counts of large text corpora) are recognized faster, and more accurately under
11 noisy conditions, than words that occur rarely (e.g., Howes & Solomon, 1951; Savin,
12 1963). This frequency effect can be couched in Bayesian terms as the impact on percep-
13 tual decisions of the prior probability of encountering a given word. The influence of fre-
14 quency has been instantiated in various ways within theories and models of spoken-word
15 recognition. In search models (e.g., the AUTONOMOUS SEARCH model (Forster, 1989)), word
16 forms are mentally organized into bins, arranged by frequency of occurrence within each
17 bin, with the result that *initial contact* with the lexicon is ordered by frequency. The recog-
18 nition of a spoken word is viewed as a self-terminating search. The search terminates
19 sooner for high-frequency words, for which a match between the input and a word form
20 can be established early in the search, than for low-frequency words. In localist activation
21 models, which characterize the dominant view in the field, word forms are conceived as
22 independent processing units that accumulate activation proportionally to their match with
23 the incoming signal. In such models, word frequency can directly influence the activation
24 of word units by modulating the units' threshold for response (e.g., the LOGOGEN model
25 (Morton, 1969)), the units' resting (i.e., default) activation (e.g., the COHORT model
26 (Marslen-Wilson, 1987)), the strength of connections between sublexical and lexical rep-
27 resentations (MacKay, 1982, 1987), or can act as a post-activation, decision bias, thus act-
28 ing on *selection* (as in the NAM (Luce, 1986; Luce & Pisoni, 1998; Luce et al., 1990)).
29

30 In an attempt to contrast the *initial contact* and *selection* instantiations of frequency,
31 some researchers hypothesized that frequency operating as a decision bias should be ob-
32 served late, with respect to the onset of spoken input (e.g., Connine, Titone, & Wang,
33 1993; Goldinger, Luce, & Pisoni, 1989). Although such delay was reported in some stud-
34 ies (Connine et al., 1993), Dahan, Magnuson, and Tanenhaus (2001b) showed that fre-
35 quency effects could be observed in the earliest moments of lexical processing. They
36 monitored participants' eye movements as they followed spoken instructions to interact
37 with items in a visual display. When fixation proportions over time to low-frequency
38 targets, low-frequency cohorts, high-frequency cohorts and unrelated distractors were
39 compared, Dahan et al. found frequency effects in the earliest signal-driven changes in
40 fixation proportions (within about 200 ms of word onset) – although the magnitude of
41 frequency effects grew as more of a word was heard. Dahan et al. added three frequency
42 mechanisms to TRACE to compare predictions of different proposals for how frequency
43 might be instantiated. Resting level and post-activation bias mechanisms yielded virtu-
44 ally identical predictions (when the post-activation bias was applied continuously, though
45

1 for it to have a “late” locus it would have to be applied suddenly after a certain amount
2 of bottom-up evidence accrued). A bottom-up connection strength instantiation (in which
3 connections between phonemes and high-frequency words were stronger than those be-
4 tween phonemes and low-frequency words) provided the best fit to the data. This account
5 predicts a continuous effect of frequency, but a gradual one, since the frequency effect
6 depends on the strength of the bottom-up input. The bottom-up connection strength
7 account would also be consistent with learning models in which connection strengths are
8 tuned to prior probabilities through experience.
9

10 3.2. Competition

11
12 There is now considerable evidence that the recognition of a spoken word is affected
13 by the set of lexical alternatives that are partially compatible with the input. A word
14 that is phonetically similar to few and/or rare other words is recognized more easily
15 than a word similar to many and/or frequent other words, above and beyond effects of
16 the frequency of the word itself (Luce, 1986; Luce & Pisoni, 1998). This indicates that
17 the recognition process does not solely depend on the degree to which the spoken input
18 matches the representation of a given word, but also on the degree to which the input
19 matches the representations of alternative words. All current theories of spoken word
20 recognition acknowledge the need for competition, but differ in the mechanisms they
21 assume accomplishes it. The primary mechanisms are *decision rules* and *direct*
22 *competition*. We will focus on these, and then turn to a third alternative, *emergent*
23 *competition*.
24

25 *Decision rule competition*. The original COHORT model (Marslen-Wilson & Welsh,
26 1978) predicted that the recognition of a spoken word depends on the activation of mul-
27 tiple candidates (the *word-initial cohort*) but only indirectly; the cohort determines the
28 *uniqueness point* of the target word – the point at which the target is the last lexical
29 candidate compatible with the input. The model assumed that the onset of a spoken word
30 activates all word candidates sharing that onset. As more input becomes available,
31 candidates are pruned from the competitor set as soon as they mismatch (e.g., *cat* is re-
32 moved from *castle*’s cohort when /s/ is heard), until only one candidate remains.
33 Inclusion or exclusion of a candidate from the competitor set was viewed as an all-or-
34 none and frequency-insensitive process. Revisions to the model, prompted by theoretical
35 and empirical arguments (Marslen-Wilson, 1987), changed the mechanism for cohort in-
36 clusion and exclusion into a gradient activation process reflecting the degree of evidence
37 for a candidate in the input and its frequency. In this revised model, candidates cannot be
38 described as simply in or out of the cohort. Instead, they are more or less activated, and
39 the criterion for recognition was changed into a decision rule that evaluates a unit’s ac-
40 tivation level with respect to the activation level of all other units (Marslen-Wilson, 1987,
41 1993). This, in effect, allows the recognition of a given word to be affected by other can-
42 didates’ match to the input, but without direct competition between units; any lexical
43 item’s activation reflects its goodness of fit to the input. Competition only exists at the
44 level of the decision rule.
45

1 A similar mechanism was proposed earlier as part of the NAM developed by Luce and
2 colleagues (Luce, 1986; Luce et al., 1990; Luce & Pisoni, 1998). The model states that
3 the probability of recognizing a given word can be approximated by the ratio of the tar-
4 get word's log frequency to the summed log frequencies of all items in its neighborhood,
5 including the target word; in other words, ease of recognition is predicted to be propor-
6 tional to the amount of frequency the target contributes to the total frequency of its neigh-
7 borhood. Computed over large sets of words, this probability rule was shown to account
8 for more unique variance in tasks like lexical decision or naming (about 15%) than any
9 other factor (the next best was target frequency alone, which only accounted for 5%). The
10 NEIGHBORHOOD model stands out among current theories in that it is a formal mathemat-
11 ical model of activation and competition, but not a processing model. It also stands out
12 for its power and simplicity. The frequency-weighted probability rule compactly embod-
13 ies general principles shared by current theories, as well as the specifics of the neigh-
14 borhood conception of competitors, and generates precise, testable predictions. Nonetheless,
15 as noted above, the NAM fails to incorporate the dynamics of a spoken word's competi-
16 tion environment.

17
18 *Direct competition.* Connectionist models like TRACE (McClelland & Elman, 1986),
19 SHORTLIST (Norris, 1994), and more recently PARSYN (Luce et al., 2000) assume compe-
20 tition among lexical units via lateral inhibition. Units within the lexical layer (and the
21 phoneme layer, in the case of TRACE and PARSYN) send each other inhibition as a function
22 of their respective activation, which depends on their similarity to the input. For exam-
23 ple, upon hearing the input /kɑt/ (*cot*), the units *cat* and *cap* would also both be acti-
24 vated; *cat* is more similar to the input than *cap*, and so would be activated more strongly,
25 and send more inhibition to *cap* than viceversa (assuming equal word frequency). The
26 end result is that a lexical item with an activation advantage will eventually suppress its
27 competitors. The recurrent loops created by lateral inhibition in these sorts of models
28 give them temporal dynamics, which allow fine-grained predictions of the activations of
29 targets and competitors over time.

30
31 Distinguishing between an implementation of lexical competition in terms of decision
32 rule or lateral inhibition has proven difficult, as they make very similar predictions
33 (Marslen-Wilson, Moss, & van Halen, 1996; see also Bard, 1990). Similar debates are
34 taking place among models of perceptual choice (Usher & McClelland, 2001). Decision-
35 rule competition is arguably a simpler computational mechanism than lateral inhibition.
36 In the decision-rule implementation, the temporal dynamics of candidates' activation can
37 only reflect changes in the evidence supporting each candidate, as the spoken input
38 unfolds over time. By contrast, competition via lateral inhibition predicts temporal
39 dynamics that reflect both the impact of evidence from the input and recurrent loops on
40 candidates' activation. Distinguishing between these two implementations is thus likely
41 to require consideration of lexical activation over time.

42
43 *Emergent competition.* Gaskell and Marslen-Wilson (1997, 1999, 2002) have proposed
44 a distributed architecture, where words are represented by overlapping, distributed pat-
45 terns of node activation. One portion of these nodes stands for phonological features,

1 while another stands for semantic features. A given word is represented as a pattern of
 2 activation among phonological and semantic feature nodes, thus capturing the form and
 3 the meaning of that word. When the initial portion of a word is presented to the model,
 4 patterns learned by the network that are consistent with the input are simultaneously
 5 activated. However, because there is only one substrate for activation—the same set of dis-
 6 tributed nodes—the outcome is an activation pattern that *blends* the consistent patterns.
 7 Thus, competition takes the form of interference between the patterns associated with
 8 candidates consistent with partial input. The activation pattern resulting from processing
 9 partial input may be more or less coherent depending on the nature of the information
 10 that the nodes encode (phonological vs. semantic) and the number of compatible
 11 hypotheses simultaneously considered.

12
 13 We refer to this as *emergent competition* because the competition dynamics arise from
 14 a complex combination of interacting causes. These include intricate patterns of excita-
 15 tory and inhibitory weights that emerge as a function of the corpus on which a recurrent
 16 network is trained, the attractors that form for phonological, semantic, and possibly com-
 17 binations of inputs and outputs.

18
 19 The model's distributed architecture makes an intriguing prediction. Although the
 20 model assumes the simultaneous activation of all the word candidates that match the
 21 input, it also predicts that the resulting pattern of activation does not represent the form
 22 or the meaning of any of these candidates individually. Rather, because this activation
 23 pattern is a blend, their common features (most often, their shared sounds) are faithfully
 24 represented, whereas their divergent features (such as their semantic features, as words
 25 that are phonologically similar are not typically semantically related) have been blended;
 26 reconstructing the divergent features of word candidates would depend, among other
 27 things, on the number of word candidates involved.

28
 29 Gaskell and Marslen-Wilson (2002) reported data supporting the model's prediction.
 30 In particular, they showed that the presentation of a spoken prime that is compatible with
 31 several possible candidates (e.g., /kθπtI/, compatible with a number of candidates,
 32 including *captain* and *captive*) does not boost participants' speed at making a lexical deci-
 33 sion on a word semantically related to one of the candidates (e.g., *commander*), suggest-
 34 ing that the semantic representations of the activated phonological forms were blended
 35 and not sufficiently distinctive to allow detectable priming. By contrast, the presentation
 36 of a spoken prime that is compatible with only one possible candidate (e.g., /γμ↔/, only
 37 compatible with *garment* [British English pronunciation]) did facilitate processing of a
 38 word semantically related to this candidate (e.g., *attire*). This result can be accounted for
 39 by the distributed architecture assumed by Gaskell and Marslen-Wilson's model because
 40 the pattern of activation in the semantic feature nodes becomes less coherent as more
 41 candidates are considered and more heterogeneous patterns (associated with form-over-
 42 lapping candidates with unrelated meanings) participate in the blend.

AQ6

43
 44 Models with localist representations could also account for this result. We are unaware
 45 of any current, implemented model that could do so without modification, but the general

1 principles of, e.g., interactive activation are consistent with the result. An explanation
 2 parallel to that of Gaskell's and Marslen-Wilson's is that the larger the phonological
 3 competitor set is, the weaker the activation that each of their semantic representation
 4 receives. The phonological competitors initially receive equivalent support from the
 5 phonological input (*mutatis mutandis* for differences in frequency, etc.). As long as no
 6 phonological representations are strongly favored by the bottom-up input, however, their
 7 corresponding semantic representations receive too little activation to be detected via
 8 priming. An analogous mechanism exists in ARTWORD (Grossberg & Myers, 2000), where
 9 a perceptual resonance (assumed to lead to conscious perception) is established only once
 10 the level of activation of one candidate (or "chunk") has sufficiently overcome that of its
 11 competitors.
 12
 13

14 3.3. Word Segmentation in Continuous Speech: Competition Across Word 15 Boundaries 16

17 A spoken utterance cannot easily be segmented into the words that compose it because
 18 boundaries between words are not reliably marked in the acoustic signal, and have often
 19 been blurred through phonological phenomena such as coarticulation and resyllabifica-
 20 tion. This is not to say that word boundaries are never acoustically marked. For instance,
 21 silent pauses between phrases mark the boundaries of the words that appear at the edges
 22 of these phrases. In fact, an extensive literature has demonstrated that listeners make use
 23 of word-boundary cues when present (phonotactic cues: McQueen, 1998; prosodic cues:
 24 Salverda et al., 2003; phonetic cues: Quené, 1992, 1993; Gow & Gordon, 1995). What
 25 this literature has shown is that word-boundary cues are used as a source of evidence sup-
 26 porting word candidates that are consistent with the hypothesized word boundary, and not
 27 used prelexically, to chunk the signal into words *before* initiating contact with the lexi-
 28 con, as had been previously proposed (e.g., Cutler, 1990).
 29

30 Because word boundary cues are probabilistic at best, and because words tend to share
 31 many of their components with other words, multiple words are consistent with virtually
 32 any portion of an utterance. For example, McQueen et al. (1995) established that 84% of
 33 English polysyllabic words contain at least one shorter embedded word (e.g., *ham* in
 34 *hamster*, or *bone* in *trombone*). This lexical ambiguity sometimes applies across word
 35 boundaries, as in *ship inquiry*, where (in British English) *shipping* matches *ship* and the
 36 initial portion of *inquiry*. Thus, competition among word candidates that start at different
 37 points in time is required. As mentioned earlier, TRACE models inter-word competition by
 38 assuming that all word units that overlap in time, i.e., competing for the same portion of
 39 the input, inhibit one another. Because a unit representing the same word is replicated
 40 many times over time/space, a given word unit can become activated as soon as the input
 41 provides some evidence supporting it, regardless of where in time the information
 42 appears. For instance, after the sequence /ΣIπIN/ (the initial portion of the phrase *ship*
 43 *inquiry*), *inquiry* can start receiving activation from the input, and eventually be recog-
 44 nized, even though *shipping* is already strongly activated. Note that some words can com-
 45 pete even when they do not share any segments. In the example above, the candidate

1 *shipment* competes with *inquiry* because both are competing for the same portion of the
2 input. Thus, TRACE solves the problem of segmenting words out of a continuous spoken
3 input by using the same mechanism it uses to segment a coarticulated signal into a se-
4 quence of phonemic units.

5
6 Alternatives to TRACE's solution to word segmentation and recognition have been pro-
7 posed. Norris (1994) criticized the multiple replications of the lexical network in TRACE.
8 He developed SHORTLIST, a model in which a limited set of candidates that are most acti-
9 vated by (i.e., consistent with) the input is compiled. The model consists of two compo-
10 nents. A lexical search network, implemented as a simple dictionary lookup, provides a
11 list of the best matches to the input at each phoneme position. The second component is
12 a competition network including as many as the top 30 candidates aligned with each
13 input position (SHORTLIST is often described as allowing a maximum of 30 words to enter
14 the competition network, but this is inaccurate; D. Norris, personal communication).
15 Items selected for each shortlist compete with one another proportionally to the number
16 of sounds they share in an interactive activation network. Items in different shortlists also
17 compete *if* they overlap. For example, given the input *ship inquiry*, *ship* and *shipping* will
18 enter the shortlist aligned with the first phoneme. *Inquiry* will eventually dominate the
19 shortlist aligned with the fourth phoneme, i.e., after *ship*, and will inhibit *shipping*, be-
20 cause the two overlap in input positions 4 and 5, but it will not inhibit *ship*, since it does
21 not overlap with *ship*. Thus, *ship* and *inquiry* create pressure for a parse into nonover-
22 lapping words, and eventually inhibit *shipping* sufficiently to allow *ship* to be recognized.
23 The selection-competition cycle repeats itself as input is presented to the model. At each
24 time step, a new lexical search is done for *every position encountered so far*. The com-
25 position of the shortlist changes dynamically as spoken input becomes available, with
26 some candidates dropping and being replaced by new candidates, depending on bottom-
27 up match/mismatch scores from the lexical search network and inhibition within the
28 competition network.

29
30 Despite the important computational economy offered by establishing the competitor
31 set in a dynamical fashion, compared to a hard-wired manner as in TRACE, SHORTLIST also
32 has several limitations. First, the lexical search mechanism is called recursively—a new
33 search is done at each position as each new phoneme is heard. If the lexical search were
34 implemented as a recurrent network, this would require one copy of the lexical network
35 for each phoneme position, and so the model would require the same number of nodes as
36 TRACE, plus those used in the shortlists (but would use many fewer *connections*). Second,
37 the biological plausibility of the dynamic programming required by SHORTLIST must be
38 addressed (cf. Protopappas, 1999). Finally, it has yet to be shown that SHORTLIST can
39 account for the broad range of data TRACE can.

40
41 ARTWORD (Grossberg & Myers, 2000) is a model specifically designed to account for
42 the dynamics of inter-word competition and how later-arriving information can modulate
43 the perception of earlier occurring speech. In this model, the spoken input activates sen-
44 sory features. Activation of these features is transformed into a sequence of items in
45 working memory. The sequential order of these items is encoded by a gradient of activity

1 within the representation (with the most active item representations corresponding to the
2 most recent event). The activity pattern in working memory in turn activates “list chunks”
3 that match the active items *and* their order. List chunks consist of unitized linguistic units
4 (e.g., phonemes, syllables, words). Activated chunks compete with one another, propor-
5 tionally to their level of activation and to the number of items they compete for. Once an
6 activated list chunk reaches an activation threshold, it sends back activation to the con-
7 sistent items in working memory, and inhibition to inconsistent items.

8
9 The excitatory loop between list chunks and items in working memory corresponds to
10 a process known as *resonance*. In Grossberg and Myers’s (2000) own words, “when lis-
11 teners perceive fluent speech, a wave of resonant activity plays across the working
12 memory, binding the phonemic items into larger language units and raising them into the
13 listener’s conscious perception” (p. 738). Thus, in this model, recognizing a spoken word
14 can be described as having associated a given linguistic interpretation to a portion of
15 speech represented in working memory, where time is encoded.

16
17 The dynamics of the resonance wave is the major factor that determines how continuous
18 speech is perceived as a succession of segmented and unitized word units. First, the model
19 includes two reset mechanisms that can terminate one resonance to allow for the next one
20 to be initiated (see Grossberg, Boardman, & Cohen, 1997, for more details). Thus, the per-
21 ception of a multi-word utterance can be described as a sequence of resonance waves.
22 Second, because of competition among activated chunks, ARTWORD accounts for recogni-
23 tion despite the activation of multiple candidates at various points in the signal. Third, the
24 model allows for later-arriving information to modify the resonance wave by *resonant*
25 *transfer*: The resonance associated with a short word (e.g., *ham*) can be transferred to a
26 longer one (e.g., *hamster*) as the second syllable of the word *hamster* is processed. Finally
27 and critically, ARTWORD can account for the impact of some word-boundary cues (such as
28 segmental lengthening, e.g., Salverda et al., 2003) without invoking additional mechanisms.
29 Indeed, a resonance transfer can only occur within a very limited, speech-rate-dependent
30 time window. Thus, if the first sounds of the second syllable of *hamster* are delayed (be-
31 cause of lengthening of the last sounds of *ham*, a silent pause, or lengthening of the sound
32 following *ham*), the resonance established between the word chunk *ham* and items in work-
33 ing memory may have been reset, and the items’ activation fallen to low activation levels.
34 No resonance transfer is then possible, and listeners will perceive the word *ham* followed
35 by another word starting with the sounds /st/. This is consistent with Salverda et al.’s re-
36 sults, showing that long /ham/ syllables tend to be interpreted as monosyllabic words.

37 38 39 **4. INTEGRATION: WHEN AND HOW IS CONTEXT INTEGRATED?**

40
41 Words occur embedded in a larger context, most often in a sentence. There exists a
42 tight interdependency between a given word and its sentential context. A word con-
43 tributes to the meaning of the sentence, but the contribution of a word to the meaning of
44 the sentence also rests on the sentence itself.

45

1 Most of the empirical work examining the interaction between a word and its senten-
2 tial context has focused on the possible constraint that the context may impose on the set
3 of word candidates compatible with the spoken input. Initial studies suggested a late im-
4 pact of context. For example, Tanenhaus, Seidenberg, and Leiman (1979; see also
5 Swinney, 1979) presented listeners with auditory sentences that were biased toward one
6 sense of a homophone (e.g., *she held the rose* vs. *they all rose*), and then used visual lex-
7 ical decision to probe semantic activation. They found statistically equivalent priming for
8 associates of both senses (e.g., *flower* and *stand*) immediately after homophone offset,
9 but only found reliable priming for the context-appropriate sense 250 ms later. This was
10 interpreted as evidence for context-free initial lexical activation, quickly followed by an
11 integration stage where word interpretations incompatible with the context are rejected.
12 Similar conclusions were reached by Zwitserlood (1989), who reported evidence for the
13 early activation of the meaning of all words compatible with the initial sounds of a spo-
14 ken word, regardless of the context.
15

16 However, Shillcock and Bard (1993) tested the hypothesis that the Tanenhaus et al.
17 contexts contained very weak biases (other form classes besides nouns or verbs could
18 have been heard at the homophone position, and the contexts at best biased listeners to-
19 ward thousands of nouns vs. thousands of verbs). They used contexts that had been
20 experimentally established as biased towards a single item – the closed class word, *would*
21 (*John said he didn't want to do the job but his brother would, as I later found out*) – or
22 towards a large number of items: (*John said he didn't want to do the job with his*
23 *brother's wood, as I later found out*). In the closed-class case, they found no evidence of
24 priming of *wood*; its associate, *timber*, was not primed even if they probed prior to the
25 offset of *would*. This suggests that top-down context *can* affect early stages of word
26 recognition, but that top-down information is generally given much less weight than bot-
27 tom-up, and is proportional to prior probability: the more narrowly constraining the top-
28 down information is, the greater the impact it may have on early moments of processing
29 (see Dahan, Swingley, Tanenhaus, & Magnuson, 2000, who report evidence for the early
30 impact of determiners marked for grammatical gender on the recognition of subsequent
31 spoken nouns in French).
32

33 Generally speaking, theories of spoken word recognition have remained agnostic about
34 the integration of sensory information with higher level context. Notable exceptions to
35 this are the three versions of the COHORT model. In the original COHORT model, top-down
36 knowledge (e.g., semantic context) played an active role throughout selection, allowing
37 recognition prior to the uniqueness point for words strongly supported by context. It also
38 had the potential to guide initial contact, by preventing a highly inconsistent item from
39 entering the recognition cohort. In the revised COHORT model, in light of intuitive and
40 empirical evidence that clearly articulated words that have low probability in a particular
41 context are still clearly perceived, context no longer affected initial contact (i.e., could no
42 longer exclude an item from entering the cohort despite strong bottom-up support).
43 Instead, context was viewed as acting on a set of candidates first established on the basis
44 of sensory information only. The model argued in favor of a context-free, initial activation
45

1 stage. The most recent version of the model, the DISTRIBUTED COHORT model, departs
 2 from this stance by assuming no division between initial contact and selection. Semantic
 3 features are an integral part of lexical representations, and thus semantic and phonologi-
 4 cal knowledge are simultaneously activated by bottom-up input. This last instantiation,
 5 by renouncing the theoretical processing division between form and meaning, is com-
 6 patible with findings of a continuous integration of different sources of evidence in order
 7 to ultimately derive an interpretation of the spoken input.
 8
 9

10 5. AVENUES FOR PROGRESS

11 The three most crucial developments for theories of spoken-word recognition, as ar-
 12 gued throughout this chapter, are (1) increasing evidence that the input to spoken word
 13 recognition retains much if not all of the surface detail of utterances; (2) evidence that
 14 language representations are not static but instead are subject to constant change; and (3)
 15 the emergence of theoretical frameworks that deny the existence of distinct stages corre-
 16 sponding to speech perception, spoken word recognition, sentence processing, and
 17 beyond – and empirical support for these theories. These developments may herald a rad-
 18 ical reconceptualization of spoken word recognition and language processing in general,
 19 if not an all-out paradigm shift.
 20

21 There are two sets of findings that compellingly demonstrate that the input to lexical
 22 access is not limited to an abstract phonemic code. The first (reviewed briefly in our
 23 introduction) is evidence that fine-grained phonetic detail affects the time course of lex-
 24 ical activation and competition (Andruski et al., 1994; Davis et al., 2002; Salverda et al.,
 25 2003). The second (reviewed in Section 2) is evidence that even (putatively) non-lin-
 26 guistic surface detail, such as talker sex or even more fine-grained talker characteristics,
 27 is preserved in memory for spoken language (Goldinger, 1996). The fact that such detail
 28 not only affects memory but also word recognition motivates exemplar theories like
 29 Goldinger's (1998) episodic lexicon theory, in which the basis for lexical (and potentially
 30 lower and higher levels of representation) categories are clusters of memory traces of,
 31 essentially, raw speech "episodes" that preserve all surface detail. On such a view, each
 32 new memory trace has the potential to change the "category" with which it is clustered,
 33 making exemplar theories compatible with recent evidence that short-term changes in
 34 phonotactic probabilities quickly influence production (Dell, Reed, Adams, & Meyer,
 35 2000) and comprehension (Onishi, Chambers, & Fisher, 2002). These rapid changes in
 36 lexical production and processing challenge the frequent, if implicit, assumption that the
 37 adult phonological and lexical knowledge is more or less fixed.
 38

39 These developments pose significant challenges to theories of spoken word recogni-
 40 tion and spoken language processing in general. They point to a system in which there
 41 may be distinct levels of representation (given the cognitive economies of composition-
 42 ality and generativity afforded by, e.g., phonemes and words), but also parallel episodic
 43 representations that are less abstract, and without discrete stages corresponding to the de-
 44 scriptive levels of speech perception, word recognition, sentence processing, and so on.
 45

1 As mentioned earlier, Goldinger and Azuma's (2003) appeal to adaptive resonance (e.g.,
2 Grossberg, 2003) as a potentially unifying framework capable of incorporating learning,
3 sublexical and lexical effects as well as the principles of episodic lexicon theory, appears
4 to hold substantial promise.

5
6 However, integrating this view with the processing of actual speech or a close analog,
7 remains a significant challenge. While the ARTSTREAM model (Grossberg et al., 2004)
8 has demonstrated the potential of the ART framework to process the speech signal itself,
9 it has not yet been extended to contact with phonemic or lexical forms. Plaut and Kello
10 (1999) provided another framework with significant promise, in which close analogs of
11 the speech signal are used, and phonological and semantic representations are treated
12 within perception and production, as well as development.

13
14 Integrating (descriptive) levels of speech perception and word recognition upwards
15 also remains as a significant challenge. Theories of sentence processing in the constraint-
16 based framework have long blurred the boundary between lexical access and sentence
17 processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Kim, 1998),
18 assuming that lexical representations include not just phonological and semantic knowl-
19 edge, but also specify the syntactic relations in which a lexical item can participate.
20 Evidence that lexical access and sentence processing are constrained in an immediate and
21 continuous fashion by nonlinguistic context – such as the actions afforded to the listener
22 by combinations of objects and instruments (Chambers, Magnuson, & Tanenhaus, 2004),
23 or even affordances available to interlocutors (Hanna & Tanenhaus, 2003) – demands that
24 we scale our theories up and integrate them with sentence, discourse, and general cogni-
25 tive processes.

26
27 We began this chapter by contrasting the modal, narrow view of spoken word recog-
28 nition (as the mapping from phonemes to sound forms that provide access to the lexicon)
29 with a broad view, encompassing the speech signal, the word level, and higher levels of
30 structure and representation. The broad view is supported by growing evidence for con-
31 tinuous effects of subphonemic information at the lexical level and beyond on the one
32 hand, and immediate integration and interaction between descriptively low and high lev-
33 els of linguistic representation and even non-linguistic affordances of physical objects
34 (Chambers et al., 2004) on the other. Our view is that significant progress in understand-
35 ing spoken word recognition, and language processing more generally, will require
36 stretching (or possibly abandoning) current theories and models to accommodate the
37 broad view of language processing.

38

39

40

ACKNOWLEDGEMENTS

41

42

43

44

45

We thank Dan Swingley and Dan Mirman for comments and discussions that greatly improved this chapter. Preparation of this chapter was supported in part by the National Science Foundation under Grant No. 0433567 to DD, and by National Institute on Deafness and Other Communication Disorders Grant DC-005765 to JSM.

1 **REFERENCES**

2
3 Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, 38, 419–439.

6 Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory & Language*, 41, 124–145.

9 Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–187.

12 Bagley, W. C. (1900). The apperception of the spoken sentence: A study in the psychology of language. *American Journal of Psychology*, 12, 80–130. **AQ7**

15 Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, 52, 339–362.

18 Bard, E. G. (1990). Competition, lateral inhibition, and frequency: Comments on the papers of Frauenfelder & Peeters, Marslen-Wilson, and others. In: G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 185–210). Cambridge, MA: MIT Press.

22 Bertelson, P., & de Gelder, B. (1989). Learning about reading from illiterates. In: M. Galaburda, & M. Albert (Eds), *From reading to neurons* (pp. 1–23). Cambridge, MA: MIT Press.

25 Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2004). Reference resolution in the wild: Circumscription of referential domains by naive participants during an interactive problem solving task. In: J. C. Trueswell & M. K. Tanenhaus (Eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.

31 Cairns, P., Shillock, R., Chater, N. & Levy, J. (1995) Bottom-up connectionist modelling of speech. In: J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds). *Connectionist models of memory and language*, University College London Press. **AQ8**

34 Chambers, C. G., Magnuson, J. S., & Tanenhaus, M. K. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 687–696.

38 Clarke, C. M., and Garrett, M. F. (2004) Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, 116, 3647–3658.

41 Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 551–563.

- 1 Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech.
2 *Perception & Psychophysics*, 13, 153–156.
- 3
4 Cole, R. A., & Rudnick, A. I. (1983). What's new in speech perception? The research and ideas
5 of William Chandler Bagley, 1874–1946. *Psychological Review*, 90, 94–101.
- 6
7 Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of
8 phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, 11,
9 1084–1089.
- 10
11 Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic
12 effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*,
13 19, 81–94.
- 14
15 Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In: G. T. M. Altmann
16 (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*
17 (pp. 105–121). Cambridge: MIT Press.
- 18
19 Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken lan-
20 guage: A literature review. *Language & Speech*, 40, 141–201.
- 21
22 Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In: W. E. Cooper, & E. C. T.
23 Walker (Eds), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*.
24 Hillsdale: Erlbaum.
- 25
26 Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spo-
27 ken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317–367.
- 28
29 Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mis-
30 matches and the time course of lexical access: Evidence for lexical competition. *Language and*
31 *Cognitive Processes*, 16, 507–534.
- 32
33 Dahan, D., Swingle, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and
34 spoken-word recognition in french. *Journal of Memory and Language*, 42, 465–480.
- 35
36 Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-
37 language comprehension: Immediate effects of verb-based thematic constraints. *Journal of*
38 *Experimental Psychology: Learning, Memory, & Cognition*, 30, 498–513.
- 39
40 Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical
41 information drives perceptual learning of distorted speech: Evidence from the comprehension of
42 noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134, 222–241.
- 43
44 Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path:
45 Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology:*
Human Perception & Performance, 28, 218–244.

- 1 Deelman, T., & Connine, C. M. (2001). Missing information in spoken word recognition:
2 Nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception and*
3 *Performance*, 27, 656–663.
- 4
5 Dell, G.S., Reed, K.D., Adams, D.R. & Meyer, A.S. (2000). Speech errors, phonotactic constraints,
6 and implicit learning: A study of the role of experience in language production. *Journal of*
7 *Experimental Psychology: Learning, Memory and Cognition*, 26, 1355–1367.
- 8
9 Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of
10 talker and rate changes. *Journal of Experimental Psychology: Human Perception & Performance*,
11 23, 914–927.
- 12
13 Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing.
14 *Perception & Psychophysics*, 67, 224–238.
- 15
16 Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 8,
17 301–306.
- 18
19 Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech wave. In: J. S.
20 Perkell, & D. H. Klatt (Eds), *Invariance and variability in speech processes*. Hillsdale, NJ:
21 Lawrence Erlbaum Associates.
- 22
23 Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception:
24 Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*,
25 27, 143–165.
- 26
27 Fodor J (1983) *Modularity of mind*. Cambridge, MA: MIT Press.
- 28
29 Forster, K. I. (1989). Basic issues in lexical processing. In: W. Marslen-Wilson (Ed.), *Lexical rep-*
30 *resentation and process*. Cambridge, MA: MIT Press.
- 31
32 Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception and*
33 *Psychophysics*, 36, 359–368.
- 34
35 Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental*
36 *Psychology: Human Perception & Performance*, 10, 526–540.
- 37
38 Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduc-
39 tion. *Cognition*, 25, 1–20.
- 40
41 Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of*
42 *Experimental Psychology: Human Perception & Performance*, 6, 110–125.
- 43
44 Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech per-
45 ception. *Journal of Phonetics*, 31, 447–463.
- 46
47 Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonologi-
48 cal representation in speech perception. *Cognitive Science*, 19, 407–439.

- 1 Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical
2 access. *Journal of Experimental Psychology: Human Perception & Performance*, 22, 144–158.
- 3
4 Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed
5 model of speech perception. *Language & Cognitive Processes. Special Cognitive Models of Speech*
6 *Processing: Psycholinguistic and Computational Perspectives on the Lexicon*, 12, 613–656.
- 7
8 Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in
9 speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, 24,
10 380–396.
- 11
12 Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken
13 word recognition. *Cognitive Science*, 23, 439–462.
- 14
15 Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word
16 recognition: Bridging the gap. *Journal of Memory & Language*, 44, 325–349.
- 17
18 Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the percep-
19 tion of spoken words. *Cognitive Psychology*, 45, 220–266.
- 20
21 Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and
22 recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22,
23 1166–1183.
- 24
25 Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological*
26 *Review*, 105, 251–279.
- 27
28 Goldinger, S.D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in
29 speech perception. *Journal of Phonetics*, 31, 305–320.
- 30
31 Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words:
32 Effects of competition and inhibition. *Journal of Memory & Language*, 28, 501–518.
- 33
34 Gow, D. W., Jr. (2001). Assimilation and anticipation in continuous spoken word recognition.
35 *Journal of Memory & Language*, 45, 133–159.
- 36
37 Gow, D. W., Jr. (2002). Does english coronal place assimilation create lexical ambiguity? *Journal*
38 *of Experimental Psychology: Human Perception & Performance*, 28, 163–179.
- 39
40 Gow, D. W., Jr. (2003a). Feature parsing: Feature cue mapping in spoken word recognition.
41 *Perception & Psychophysics*, 65, 575–590.
- 42
43 Gow, D. W., Jr. (2003b). How representations help define computational problems: Commentary
44 on Grossberg, Gaskell and Greenberg. *Journal of Phonetics*, 31, 487–493.
- 45
46 Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation:
47 Evidence from priming. *Journal of Experimental Psychology: Human Perception & Performance*,
48 21, 344–359.

- 1 Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech pro-
 2 duced by rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 421–433.
- 3
 4 Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*,
 5 423–445.
- 6
 7 Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech cate-
 8 gorization. *Journal of Experimental Psychology: Human Perception & Performance*, *23*, 481–503.
- 9
 10 Grossberg, S., Govindarajan, K.K., Wyse, L.L., & Cohen, M.A. (2004) ARTSTREAM: A neural
 11 network model of auditory scene analysis and source segregation”. *Neural Networks*, *17*, 511–536.
- 12
 13 Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword
 14 integration and duration-dependent backward effects. *Psychological Review*, *107*, 735–767.
- 15
 16 Hahn, U., & Bailey, T. M. (2004). What makes words sound similar? *Cognition* (in press). **AQ9**
- 17
 18 Hanna, J.E. & Tanenhaus, M.K. (2003). Pragmatic effects on reference resolution in a collabora-
 19 tive task: evidence from eye movements. *Cognitive Science*, *28*, 105–115.
- 20
 21 Hintzman, D. L. (1986). “Schema Abstraction” in a multiple-trace memory model. *Psychological*
 22 *Review*, *93*, 411–428.
- 23
 24 Hintzman, D. L., Block, R. A., & Inskoop, N. R. (1972). Memory for mode of input. *Journal of*
 25 *Verbal Learning and Verbal Behavior*, *11*, 741–749.
- 26
 27 Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probabil-
 28 ity. *Journal of Experimental Psychology*, *41*, 401–410.
- 29
 30 Joos, M. (1948). *Acoustic phonetics*. Baltimore, MD: Linguistic Society of America.
- 31
 32 Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of
 33 temporary syntactic closure ambiguity. *Journal of Memory & Language*, *40*, 153–194.
- 34
 35 Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access.
 36 *Journal of Phonetics*, *7*, 279–312.
- 37
 38 Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- 39
 40 Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the*
 41 *Acoustical Society of America*, *29*, 98–104.
- 42
 43 Liberman, A., Cooper, F. S., Shankweiler, D., P. & Studdert-Kennedy, M. G. (1967). Perception of
 44 the speech code. *Psychological Review*, *74*, 431–461.
- 45
 46 Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme
 47 segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201–212.
- 48
 49 LoCasto, P., & Connine, C. M. (2002). Rule-governed missing information in spoken word recog-
 50 nition: Schwa vowel deletion. *Perception and Psychophysics*, *64*, 208–219.

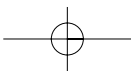
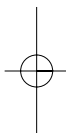
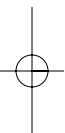
- 1 Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition.
2 *Perception & Psychophysics*, 39, 155–158.
- 3
4 Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neigh-
5 borhood activation, and parsyn. *Perception and Psychophysics*, 62, 615–625.
- 6
7 Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words.
8 *Memory & Cognition*, 26, 708–715.
- 9
10 Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation
11 model. *Ear and Hearing*, 19, 1–36.
- 12
13 Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words.
14 In: G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and compu-
15 tational perspectives* (pp. 122–147). Cambridge, MA: MIT Press.
- 16
17 MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic
18 ambiguity resolution. *Psychological Review*, 101, 676–703.
- 19
20 MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled
21 behavior. *Psychological Review*, 89, 483–506.
- 22
23 MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other
24 cognitive skills*. New York: Springer.
- 25
26 Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003a). Lexical effects on com-
27 pensation for coarticulation: The ghost of Christmas past. *Cognitive Science*, 27, 285–298.
- 28
29 Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003b). Lexical effects on com-
30 pensation for coarticulation: A tale of two systems? *Cognitive Science*, 27, 801–805.
- 31
32 Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [Σ]-[s] dis-
33 tinction. *Perception & Psychophysics*, 28, 213–228.
- 34
35 Marslen-Wilson, W. (1984). Function and process in spoken word recognition. In: H. Bouma &
36 D. G. Bouwhis (Eds), *Attention and performance x: Control of language processes* (pp. 125–150).
37 Hillsdale, NJ: Lawrence Erlbaum Associates.
- 38
39 Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- 40
41 Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In W.
42 Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.
- 43
44 Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In: G.T.M.
45 Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational
Perspectives* (pp.148–172). Cambridge, MA: MIT Press.
- Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In: G. T. M.
Altmann, & R. Shillcock (Eds), *Cognitive Models of Speech Processing: The Second Sperlonga
Meeting*, pp. 187–210. Mahwah: Erlbaum.

- 1 Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in
2 lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, 22,
3 1376–1392.
- 4 Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representations and process in lex-
5 ical access: Words, phonemes, features. *Psychological Review*, 101, 653–675.
- 6
7 Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word
8 recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.
- 9
10 Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of
11 word onsets. *Journal of Experimental Psychology: Human Perception & Performance*, 15,
12 576–585.
- 13
14 McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive*
15 *Psychology*, 18, 1–86.
- 16
17 McClelland, J. L., St John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distrib-
18 uted processing approach. *Language & Cognitive Processes. Special Parsing and Interpretation*, 4,
19 SI287–SI335.
- 20
21 McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal*
22 *of Experimental Psychology: Learning, Memory, & Cognition*, 29, 539–553.
- 23
24 McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category pho-
25 netic variation on lexical access. *Cognition*, 86, B33–B42.
- 26
27 McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of*
28 *Memory & Language*, 39, 21–46.
- 29
30 McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recog-
31 nition and the contents of the vocabulary. *Language and Cognitive Processes*, 10, 309–331.
- 32
33 McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making:
34 Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human*
35 *Perception & Performance*, 25, 1363–1389.
- 36
37 Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the per-
38 ception of stop consonant and semivowel. *Perception & Psychophysics*, 25, 457–465.
- 39
40 Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations
41 of lexically induced delays in phoneme recognition. *Journal of Memory & Language*, 52, 416–435.
- 42
43 Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76,
44 165–178.
- 45
46 Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spo-
47 ken word recognition. *Journal of the Acoustical Society of America*, 85, 365–378.

- 1 Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*,
2 52, 189–234.
- 3
4 Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition:
5 Feedback is never necessary. *Behavioral & Brain Sciences*, 23, 299–370.
- 6
7 Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive*
8 *Psychology*, 47, 204–238.
- 9
10 Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cogni-
11 tive process. In: K. Johnson, & J. W. Mullennix (Eds), *Talker variability in speech processing* (pp.
12 109–132). San Diego, CA: Academic Press.
- 13
14 Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In: Y.
15 Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds), *Speech perception, speech production, and*
16 *linguistic structure* (pp. 113–134). Tokyo: OHM.
- 17
18 Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief
19 auditory experience. *Cognition*, 83, B13–B23.
- 20
21 Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and
22 recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory,*
23 *& Cognition*, 19, 309–328.
- 24
25 Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the*
26 *Acoustical Society of America*, 24, 175–184.
- 27
28 Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition.
29 *Cognition*, 25, 21–52.
- 30
31 Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon?
32 *Journal of Memory & Language*, 39, 347–370.
- 33
34 Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech com-
35 prehension and production: A distributed connectionist approach. In: B. MacWhinney (Ed.), *The*
36 *emergence of language* (pp. 381–416). Mahwah, NJ: Lawrence Erlbaum Associates.
- 37
38 Protopapas, A. (1999). Connectionist modeling of speech perception. *Psychological Bulletin*, 125,
39 410–436.
- 40
41 Quené, H. (1992). Durational cues for word segmentation in dutch. *Journal of Phonetics*, 20,
42 331–350.
- 43
44 Quené H. (1993). Segment durations and accent as cues to word segmentation in dutch. *Journal of*
45 *the Acoustical Society of America*, 94, 2027–2035.
- 46
47 Remez, R. E. (2005). The perceptual organization of speech. In: D. B. Pisoni, R. E. Remez (Eds),
48 *The Handbook of Speech Perception*, (pp. 28–50). Oxford: Blackwell.

- 1 Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken
2 words than in spoken nonwords. *Perception & Psychophysics*, *19*, 394–398.
- 3
4 Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the res-
5 olution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.
- 6
7 Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of*
8 *Experimental Psychology: General*, *110*, 474–494.
- 9
10 Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive*
11 *Psychology*, *32*, 97–127.
- 12
13 Samuel, A. G. (2001a). Knowing a word affects the fundamental perception of the sounds within
14 it. *Psychological Science*, *12*, 348–351.
- 15
16 Samuel, A. G. (2001b). Some empirical tests of merge’s architecture. *Language & Cognitive*
17 *Processes*, *16*, 709–714.
- 18
19 Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensa-
20 tion for coarticulation. *Journal of Memory & Language*, *48*, 416–434.
- 21
22 Savin, H.B. (1963). Word frequency effects and errors in the perception of speech. *Journal of the*
23 *Acoustical Society of America*, *35*, 200–206.
- 24
25 Scharenborg, O., ten Bosch, L., Boves, L., & Norris, D. (2003). Bridging automatic speech recog-
26 nition and psycholinguistics: Extending shortlist to an end-to-end model of human speech recog-
27 nition (I). *Journal of the Acoustical Society of America*, *114*, 3032–3035.
- 28
29 Shankweiler, D., & Fowler, A. E. (2004). Questions people ask about the role of phonological
30 processes in learning to read. *Reading and Writing: An Interdisciplinary Journal*, *17*, 483–515.
- 31
32 Shillcock, R. C., & Bard, E. G. (1993). Modularity and the processing of closed-class words. In:
33 G. T. M. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The second*
34 *sperlonga meeting* (pp. 163–183). Mahwah, NJ: Lawrence Erlbaum Associates.
- 35
36 Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word
37 recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of the*
38 *Acoustical Society of America*, *96*, 1314–1324.
- 39
40 Speer, S., & Blodgett, A. (2006). Prosody. In: M. A. Gernsbacher, & M. Traxler (Eds), *Handbook*
41 *of Psycholinguistics*. Amsterdam: Elsevier.
- 42
43 Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case
44 of final /t/. *Journal of Memory and Language*, *52*, 322–338.
- 45
46 Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of con-
47 text effects. *Journal of Verbal Learning & Verbal Behavior*, *18*, 645–659.

- 1 Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the
2 processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal*
3 *Behavior, 18*, 427–440.
- 4
5 Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration
6 of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.
- 7
8 Trueswell, J.C., & Kim, A.E. (1998). How to prune a garden-path by nipping it in the bud: Fast-
9 priming of verb argument structures. *Journal of Memory and Language, 39*, 102–123.
- 10
11 Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, compet-
12 ing accumulator model. *Psychological Review, 108*, 550–592.
- 13
14 Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable struc-
15 ture variation on word recognition. *Perception and Psychophysics, 62*, 1297–1311.
- 16
17 van Ooijen, B. (1996). Vowel mutability and lexical selection in english: Evidence from a word
18 reconstruction task. *Memory and Cognition, 24*(5), 573–583.
- 19
20 Warren, R. M. (1970). Restoration of missing speech sounds. *Science, 167*, 392–393.
- 21
22 Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception*
23 *& Psychophysics, 35*, 49–64.
- 24
25 Zwislerlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word pro-
26 cessing. *Cognition, 32*, 25–64.
- 27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45



AUTHOR QUERY FORM

ELSEVIER

Handbook of Psycholinguistics

JOURNAL TITLE:	HP-TRAXLER
ARTICLE NO:	Ch008

Queries and / or remarks

Query No	Details required	Author's response
AQ1	In Hintzman et al. (1972), Hintzman is spelt as Hintzman and Hintzman. Please confirm the correct spelling.	
AQ2	In Hintzman (1986), Hintzman is spelt as Hintzman and Hintzman. Please confirm the correct spelling.	
AQ3	Please confirm whether Dahan et al. (2001) corresponds to 2001a or 2001b.	
AQ4	Nygaard et al. (1994), Mann & Repp (1981), Mirman (in press), Tanenhaus, Seidenberg, and Leiman (1979) are not listed in Reference list.	
AQ5	Please provide the year of publication in Mirman et al. (in press). Please check whether Mirman et al. (in press) refers to Mirman et al. (2005).	
AQ6	Please check whether any symbol is missing in between 'γ' and 'μ' in the sentence "By contrast, the presentation ... to this candidate (e.g., attire)."	
AQ7	Bagley (1900), Cole & Rudnicky (1983), Kuhn (1962), Mann & Repp (1980), Mirman et al. (2005), Tanenhaus, Leiman, and Seidenberg (1979) are not cited in the text.	
AQ8	Please provide the place of publication in Cairns et al. (1995).	
AQ9	Please provide the vol. no. and page range in Hahn & Bailey (2004).	

AUTHOR QUERY FORM

ELSEVIER

**Handbook of
Psycholinguistics**

JOURNAL TITLE:	HP-TRAXLER
ARTICLE NO:	Ch008

Queries and / or remarks

Query No	Details required	Author's response