Zones of Proximal Development for Models of Spoken Word Recognition*

Daniel Mirman†
Department of Psychology, Drexel University and
Moss Rehabilitation Research Institute

Abstract

For over a quarter-century, the TRACE model and simple recurrent networks have dominated computational modeling of spoken word recognition, providing the broadest and deepest coverage of behavioral phenomena. This chapter will highlight a few of the more recent and currently-active phenomena. It will then describe challenges to these models – domains that are important for understanding spoken word recognition and where a substantial behavioral literature has accrued, but where computational accounts are still lacking. These zones of proximal development are promising domains for expansion toward a more complete model of spoken word recognition.

Computational accounts of spoken word recognition have largely relied on the TRACE model (McClelland & Elman, 1986) and Simple Recurrent Networks (SRNs; Elman, 1990). It is a testament to their explanatory power that, 25 years later, these models remain the dominant computational frameworks for spoken word recognition. These models have been so successful that they have largely defined the agenda for models of spoken word recognition, but there are now well-documented, important phenomena that are relevant to spoken word recognition and yet are (largely) not addressed by models of spoken word recognition. The goal of this chapter is to look forward from these models. Detailed reviews of spoken word recognition (e.g., Magnuson, Mirman, & Myers, 2013) and comprehensive historical reviews of models of spoken word recognition (e.g., Magnuson, Mirman, & Harris, 2012; Weber & Scharenborg, 2012) are available elsewhere; I will provide only a basic introduction to the computational problem of spoken word recognition and the architectures of the TRACE model and SRNs, then focus on a few important behavioral phenomena that these

*This chapter is meant to appear in In T. Harley (Series Ed.) & M.G. Gaskell & J. Mirkovic (Vol. Eds.), *Current Issues in the Psychology of Language: Speech Perception and Spoken Word Recognition*, Psychology Press. Preparation of this chapter was supported by grant R01DC010805 to D.M. from the National Institutes of Health.

†Address correspondence to Daniel Mirman, Department of Psychology, Drexel University, 3141 Chestnut St., Philadelphia, PA 19104, USA. Phone: 215-553-7169. Email: dan@danmirman.org

models have addressed and a few challenges for these models. The challenges take the form of neighboring domains that are important for understanding spoken word recognition and where a substantial behavioral literature has accrued, but where computational accounts are still lacking. Other models offer more focused accounts of phenomena either within spoken word recognition or in related domains, but none offer a substantive account of spoken word recognition that also addresses the challenges discussed in this chapter. I will conclude with some suggestions for expanding the scope of models of spoken word recognition through integration of computational accounts of spoken word recognition and other domains.

1 Spoken Word Recognition

The computational problem of spoken word recognition is typically defined as mapping a string of speech sounds to a lexical representation (e.g., Magnuson et al., 2013). The speech sounds are often represented abstractly, as distinct phonemes, which may be defined by articulatory features (voicing, place of articulation, etc.) to allow the model to capture effects of feature similarity between phonemes (e.g., that /b/ is more similar to /p/ than to /s/). The lexical representation is a critical intermediate step in the arbitrary mapping from the auditory (phonological) input to word meaning (semantics). That is, the lexical representations pull apart similar auditory/phonological representations in order to facilitate their mapping to very different semantic representations (for example, "bat" and "pat" sound similar but have very different meanings). Whether these non-semantic lexical representations are a theoretical claim or just a convenient simplification, "lexical access" is typically taken to mean activation of a representation of a unique word-form without necessarily activating its semantic, syntactic, or pragmatic relations.

Even within this narrow definition, spoken word recognition is a very challenging computational problem. The speech signal is transient and its rate is not controlled by the listener and, unlike printed words and letters, individual speech sounds are blended together due to coarticulation, lacking invariant cues to boundaries between speech sounds or spoken words. The word segmentation and recognition problem is further complicated by the fact that many longer words have shorter words embedded within them (e.g., "window" includes "win", "wind", "in", and "dough"/"doe"). Many of these issues are discussed in more detail in other chapters of this volume.

The computational difficulty of this problem is belied by the fact that typical adults seem to recognize spoken words effortlessly. There is broad agreement about the core aspects of how spoken word recognition is carried out:

- Incremental activation: activation of lexical candidates begins as soon as the initial sound is heard.

- Parallel activation: multiple lexical candidates are activated simultaneously, with activation strength proportional to their similarity to the input and their prior probability (frequency and fit with context).
- Competition: lexical candidates compete for recognition.
- Interactivity: activation of lexical candidates feeds back to influence speech perception and is influenced by top-down signals from semantic, syntactic, and pragmatic context.

2 The TRACE Model and Simple Recurrent Networks

The TRACE model (McClelland & Elman, 1986) and SRNs (Elman, 1990) have been the two most influential models of speech perception and spoken word recognition. At the time of their introduction, these models already captured many of the important behavioral phenomena and they have continued to be remarkably successful at making novel predictions and explaining new phenomena. Detailed descriptions of these models and their relations to other models of spoken word recognition are provided in the original articles and other recent reviews (Magnuson et al., 2012; Weber & Scharenborg, 2012).

The TRACE model (Figure 1, left panel) is an implementation of interactive activation and competition principles in the domain of speech perception and spoken word recognition. It consists of a feature layer, a phoneme layer, and a lexical layer; with each layer consisting of a set of simple processing units. Activation of each unit corresponds to the combined evidence that the corresponding linguistic unit occurred at a particular time within the spoken input. When input is presented at the feature layer, it is propagated in a cascading manner to the phoneme layer, then to the lexical layer through hand-coded excitatory connections between mutually consistent units. As phoneme and lexical units become active, mutually exclusive units within layers compete through inhibitory connections and send excitation back, so the excitatory process is bi-directional ("interactive") with both bottom-up (features to phonemes to words) and top-down (words to phonemes to features) signals. Processing proceeds incrementally with between-layer excitation and within-layer competitive inhibition. The units follow the standard interactive activation function: negative net input drives the unit towards its minimum activation level, positive net input drives the unit towards its maximum activation level, and activation decays toward a baseline rest activation level. A fully-documented Java implementation of TRACE is available at http://maglab.psy.uconn.edu/jtrace/ (Strauss, Harris, & Magnuson, 2007).

Unlike the TRACE model, which is a specific implementation of certain computational principles, simple recurrent networks constitute a diverse class of models that share certain implementational and computational features. A standard SRN (Elman, 1990) consists

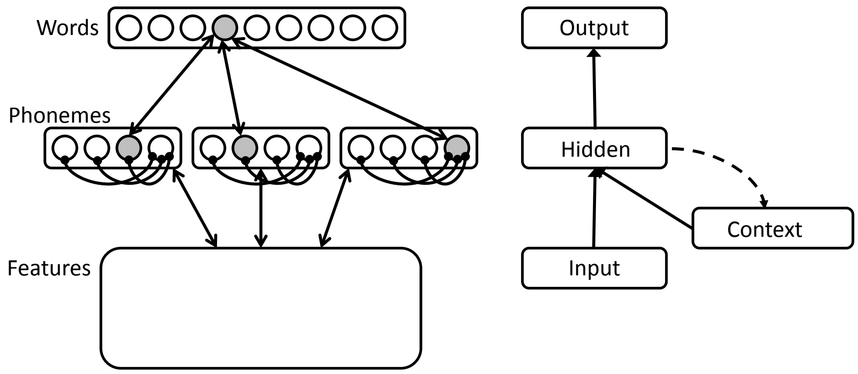

Figure 1: Left: Schematic diagram of the TRACE model of speech perception and spoken word recognition. Arrows represent hand-coded weights (only a subset is shown). Right: Schematic diagram of simple recurrent network. Solid arrows represent trainable weights, dashed arrow represents copy weights.

of simple processing units organized into four layers: input, hidden, output and context (Figure 1, right panel). There are feedforward connections from input to hidden units and from hidden to output units, as in a standard feedforward network. The context units contain an exact copy of the hidden units at the previous time step and are fully connected to the hidden units. These recurrent connections provide the network with the potential for graded memory of previous time steps. Except for the fixed "copy" hidden-to-context connections, all connections in the network are trained through backpropagation of error (Rumelhart, Hinton, & Williams, 1986), which is a generalization of the "delta rule" gradient-descent learning algorithm in which the observed network output is compared to a target pattern and connection weights are adjusted to reduce this error.

A typical approach to modeling spoken word recognition with SRNs is to present a sequence of input vectors corresponding to segments of auditory/phonological input (e.g., features, phonemes, or syllables) and set the desired output to be the next input segment or the current word. This is an important innovation over other supervised learning algorithms because it does not require an external teaching signal – the input signal itself provides the target activation information. The model is trained to predict the input, which forces it to learn the structure of the input (for a more detailed discussion of the importance of prediction in the related domain of sentence processing see Altmann & Mirković, 2009). Depending on the nature of the training corpus and the size of the network, SRNs can develop sensitivity to fairly long stretches of context. For the purposes of modeling spoken word recognition, an effective strategy (e.g., Magnuson, Tanenhaus, Aslin, & Dahan, 2003) is to use a feature-based representation of phonemes as input and localist lexical representations as output (like the TRACE model); however, one can use distributed output representations (e.g., Gaskell & Marslen-Wilson, 1997), or have multiple output layers for simulating simultaneous performance in different tasks (e.g., Gaskell & Marslen-Wilson, 1997; Mirman, Graf Estes, & Magnuson, 2010).

3 Model Successes

The TRACE model and SRNs can account for dozens of behavioral phenomena in speech perception and spoken word recognition; this section highlights several sets of behavioral phenomena selected because (1) they critically involve lexical representations, (2) they continue to be the subject of active behavioral research in spoken word recognition, and (3) they were demonstrated (largely) after the introduction of these models so the models were not designed to account for these phenomena. This last point is important because one of the primary strengths of computational models – that they make concrete testable predictions – typically leads them to be revised or even completely rejected. The fact that the TRACE model and SRNs have accounted for so many new phenomena represents their most impressive successes and is largely responsible for their continued relevance in current research on spoken word recognition.

3.1 Time Course of Activation and Competition

Allopenna, Magnuson, and Tanenhaus (1998) tracked participants' eye movements while they performed a spoken word-to-picture matching task and found that, as the input signal unfolds, fixation probabilities rise equivalently for words matching the input ("cohorts", such as "penny" and "pencil") until disambiguating information becomes available. Later in the time course, words that rhyme with the target (e.g., "carrot" and "parrot") are also fixated more than baseline unrelated words, though this effect is smaller than the cohort effect. Figure 2 (left panel) shows an example of data from such an experiment. Simulations of the TRACE model (Figure 2, middle panel) reveal the same pattern: strong initial activation of cohort competitors followed by later and (much) weaker activation of rhyme competitors. The rhyme activation is lower than cohort activation because it is being suppressed by the already-active target and cohorts. Together, these behavioral and simulation results demonstrate that the incremental and parallel activation principles must operate as a "continuous mapping" process with incoming input providing bottom-up support for all lexical representations that are consistent with it (for more discussion of different types of incrementality and continuous mapping see Allopenna et al., 1998).

In a subsequent study, Magnuson et al. (2003) demonstrated that the same effect is produced by a SRN (Figure 2, right panel). Their SRN was trained to predict the current word based on sequential input of articulatory/auditory features. The cohort competition

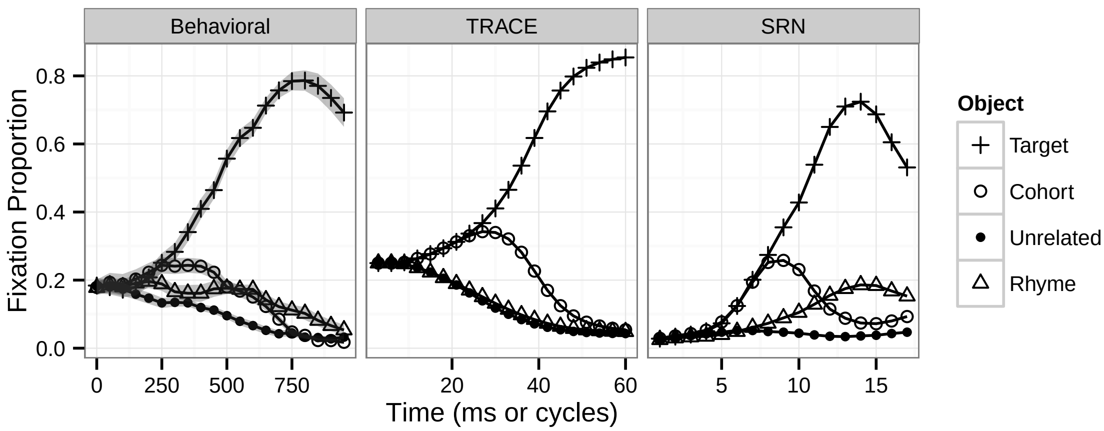

Figure 2: Cohort and rhyme competition effects. Left: example of behavioral (eye-tracking) data based on a pilot experiment conducted in the author's laboratory (gray ribbon indicates SE). Middle: TRACE model predicted fixation probabilities based on simulations reported by Mirman et al. (2011). Right: SRN predicted fixation probabilities, replotted with permission from Magnuson et al. (2003).

effect is straightforward: the model was trying to guess the incoming word, so given /bi/ it would partially active all words in its lexicon that started with /bi/ – beaker, beetle, etc. Similarly, later information (e.g., /-ikˆr/) would provide partial bottom-up support for words sharing that offset (e.g., beaker, speaker, etc.). The SRN did not have inhibitory weights between lexical units, so the magnitude of the rhyme effect was based entirely on the bottom-up weights from the hidden layer and, indeed, the rhyme effect evolved over the course of learning. Initially, the rhyme effect was as large as the cohort effect, indicating that the model had not yet learned its lexicon well enough to take complete advantage of the early input. As learning continued, the rhyme effect was gradually reduced, which is an important developmental progression (a point that is revisited below). Additional training led to the complete elimination of the rhyme effect as the model learned to take advantage of its simplified lexicon and perfectly clear input in order to rule out words that did not match at onset.

3.2 Statistical Learning

As mentioned above, the speech signal does not contain invariant cues to word boundaries, so word segmentation requires combining multiple imperfect cues (e.g., Christiansen, Allen, & Seidenberg, 1998). One of those cues is lexical information: even without marked word boundaries, it is sometimes possible to infer boundary locations based on the constraint that the resulting segments have to be real words. For example, there is only one way to segment the printed sequence "silencesbetweenwords" to produce a sensible sequence of words. The lexical competition dynamics of the TRACE model produce this sort of lexically-based word segmentation (a clear and detailed discussion with simulations is provided by McClelland & Elman, 1986, p. 61-69).

Another cue is the transitional probability between speech segments: the likelihood of a particular segment (e.g., syllable) given the preceding segment (syllable). Within-word transitional probabilities tend to be high and between-word transitional probabilities tend to be low. In a landmark study, Saffran and colleagues found that infants and adults are sensitive to such transitional probabilities and can use them to segment a speech stream that has no other cues to word boundaries (for a simple review see Saffran, 2003). Subsequent studies have shown that this kind of statistical learning facilitates learning of word meanings in infants (Graf Estes, Evans, Alibali, & Saffran, 2007) and adults (Mirman, Magnuson, Graf Estes, & Dixon, 2008). Although these behavioral observations had not yet been made, Elman's (1990) original description of SRNs demonstrated that SRNs are sensitive to transitional probabilities and that this information could be used to segment words. Later simulations specifically showed that SRNs also exhibit the facilitative effect of transitional probability on word learning and that this effect arises because syllables with higher transitional probabilities develop hidden-layer representations that are more distinctive and therefore easier to map to lexical-semantic representations (Mirman et al., 2010).

3.3 Lexically-Guided Tuning

One consequence of interactive processing in spoken word recognition is that lexical activation feeds back affect speech perception (for reviews see McClelland, Mirman, Bolger, & Khaitan, 2014; McClelland, Mirman, & Holt, 2006). A particularly important recent example is lexically-guided tuning (for a review see Samuel & Kraljic (2009) and for a broader discussion see Chapter 9 in this volume): if an ambiguous speech sound is consistently presented in disambiguating lexical contexts, it will come to be perceived in the lexically-consistent way even without the context. For example, an ambiguous sound between /f/ and /s/ that is presented in words like "sheri_" will come to be identified as /f/ even when presented in isolation (or will be identified as /s/ if it had been presented in words like "Pari_").

The addition of a simple Hebbian learning algorithm was sufficient to produce this lexically-guided tuning effect in the TRACE model (Mirman, McClelland, & Holt, 2006): an ambiguous input initially activates both possible phoneme units, but lexical feedback provides direct excitatory input to the lexically-consistent phoneme unit. The Hebbian learning algorithm implemented by Mirman et al. strengthened the connection between concurrently active units; since the lexically-consistent phoneme unit was more active (due to lexical feedback) it developed stronger connections to the input pattern. As a result, after the exposure/learning period, the previously ambiguous input pattern activated the lexically-consistent phoneme more strongly even when there was no lexical context.

3.4 Individual Differences

Understanding individual differences in speech perception and spoken word recognition has important theoretical and practical implications, but these efforts are often marginalized. On the practical side, understanding individual differences may help to develop more effective treatment strategies for developmental and acquired language deficits and may help adults learn a second language. On the theoretical side, individual differences reveal the structure of the cognitive system under investigation. The individual differences approach has deep roots, from the work of 19th century neurologists, which formed the foundation of cognitive neuroscience, to research on chess and memory experts at the dawn of the cognitive revolution, though its appeal may have waned with the advent of functional neuroimaging and undergraduate participant pools. Nevertheless, the combination of computational modeling and case series studies of individuals with acquired language deficits (Patterson & Plaut, 2009; Schwartz & Dell, 2010) has had tremendous impact on theories of semantic memory (e.g., Lambon Ralph, 2014; Rogers et al., 2004) and spoken word production (Dell, Schwartz, Nozari, Faseyitan, & Coslett, 2013; Schwartz, Dell, Martin, Gahl, & Sobel, 2006). Note that these computational models are instantiations of general theories of semantic memory and spoken word production, not theories of deficits per se. The neuropsychological data provide additional constraints on those theories and a potential link between cognitive theories and the neural implementation (e.g., Dell et al., 2013; Lambon Ralph, 2014). There have been far fewer systematic efforts to use computational models to account for individual differences in spoken word recognition. However, there have been a few important and promising successes.

Yee (2005) found that individuals with Broca's aphasia exhibited reduced cohort competition and increased rhyme competition, whereas individuals with Wernicke's aphasia exhibited the opposite pattern – increased cohort competition and reduced rhyme competition. Mirman et al. (2011) re-analyzed these data and found that, in addition to the apparent double dissociation, the participants with aphasia also exhibited a negative correlation between cohort and rhyme competition effect size. Mirman et al. then manipulated TRACE model parameters to implement six different competing accounts of spoken word recognition deficits in aphasia. Only one implementation accounted for both the double dissociation and the negative correlation between cohort and rhyme effect sizes: a response selectivity parameter that controls how activation of lexical representations drives responses. Low selectivity meant less sensitivity to activation differences, which predicted relatively more selection (fixation) of weak competitors (rhymes) and less selection (fixation) of strong competitors (cohorts), as observed for individuals with Broca's aphasia. High selectivity meant increased sensitivity to activation differences, which predicted relatively more selection (fixation) of strong competitors (cohorts) and less selection (fixation) of weak competitors (rhymes), as observed for individuals with Wernicke's aphasia. Regardless of whether this specific claim holds as more data are collected, computational models provide a common ground for testing theories. Creating concrete instantiations of the competing accounts makes it easier to make predictions from individual theories, to test theories, and to improve them. Similar approaches have also been applied to individual differences within the relatively homogenous population of college students (Mirman, Dixon, & Magnuson, 2008) and in children with specific language impairment (McMurray, Samelson, Lee, & Tomblin, 2010).

These successes represent just first steps toward computational models of individual differences in spoken word recognition. There are far more behavioral data about individual differences in speech perception and spoken word recognition than have been addressed computationally, from bilingualism (Chapter 6 of this volume) to atypical development (specific language impairment, autism spectrum disorders, etc.; Chapter 8 of this volume) to acquired language disorders. It is not uncommon to attribute such individual differences to constructs, such as inhibitory control or language experience, that have relatively direct computational implementations in TRACE and/or SRNs (e.g., lexical inhibition parameter, number of training epochs). Therefore, such theoretical claims are readily testable using these computational models.

3.5 Development

An extremely important and very common source of individual differences in spoken word recognition is simply development. Mayor and Plunkett (2014) conducted simulations of the TRACE model using lexicons based on published norms of infant vocabulary development and provided new insights into how lexicon size and composition should affect spoken word recognition during development. Examining the specific case of a proposed shift from holistic to segmental word representations (Walley, 1993), SRN simulations (Magnuson et al., 2003) offered a new theoretical perspective: a naïve (untrained) model does not know whether word onsets or offsets are going to be more informative about the identity of the word and it takes some time for the model to learn the phonological structure of its lexicon. As a result, early in learning, weak representations produce equal sensitivity to onset (cohort) and offset (rhyme) overlap. As the model's knowledge of the lexical structure develops, the model learns to take more and more advantage of the onset information, leading to larger cohort effects and smaller rhyme effects.

As these studies demonstrate, a biologically plausible model is not strictly necessary for computational investigations of spoken language development and individual differences. By providing direct access to the parameters that govern its processing dynamics, the TRACE model is a powerful framework for testing and developing computational accounts of individual differences and development. Developmental and acquired language impairments can be modeled using adjustments of these parameters; typical development can be modeled as increases in connection strengths and lexicon size. SRNs offer another framework for computational investigation of development (and individual differences), one in which both development and the parameters that govern processing dynamics are more intrinsic to the model and the results are emergent properties. Like the domain of individual differences, there are far more behavioral data on development of spoken word recognition than have concrete computational accounts (see, e.g., Chapters 3 and 7 of this volume); however, the existing successes suggest that TRACE and SRNs are promising frameworks for future computational efforts.

4 Model Challenges

The previous section highlighted some of the most important successes of the TRACE model and SRNs for current research in spoken word recognition. However, these successes are within the specific domain of spoken word recognition; the weaknesses are in addressing how spoken word recognition interacts with and is influenced by other cognitive domains such as semantic memory, cognitive control, and learning. Because these issues have been defined as outside the scope of models of spoken word recognition, the existing models have fallen farther and farther behind contemporary behavioral and neuroscience research on spoken word recognition. As the final section will discuss, these domains have their own computational models, so there is potential for integration, but current models of spoken word recognition have not done so.

4.1 Semantic, Syntactic, and Pragmatic Constraints

Models of spoken word recognition typically stop at a simple localist lexical representation that does not capture any semantic information. Semantic information is activated very quickly from speech input, even before the word is complete (e.g., Gaskell & Marslen-Wilson, 2002) and top-down semantic influences reach down to affect speech perception (e.g., Borsky, Tuller, & Shapiro, 1998; Warren & Warren, 1970). That is, cascading and interactive processing should produce lexical activations that reflect the combination of top-down influences and bottom-up phonological influences. Indeed, like the cohort and rhyme competition effects, eye-tracking studies have revealed incidental activation of semantic competitors during spoken word-to-picture matching (e.g., Huettig & Altmann, 2005; Mirman & Graziano, 2012; Mirman & Magnuson, 2009; Yee & Sedivy, 2006), including time course differences between distinct types of semantic features or relations (e.g., Kalénine, Mirman, Middleton, & Buxbaum, 2012).

Failing to consider semantic influences when modeling spoken word recognition could lead to substantial misrepresentation or misunderstanding of the processing dynamics, especially when the core experimental paradigms often involve semantic processing (e.g., spoken word-to-picture matching requires accessing the semantic content of the word and the pictures). For example, although competition between phonological neighbors typically delays recognition of spoken words from "dense" phonological neighborhoods, Chen and Mirman (2014) found that semantic input (implemented as picture preview) could reverse this effect. This surprising facilitative effect was predicted by their model (see also Chen & Mirman, 2012) in large part because their model was designed to capture the effects of both phonological and semantic input.

Models of spoken word recognition have also ignored a growing literature showing that listeners anticipate upcoming words based on preceding semantic context (for a review see Altmann & Mirković, 2009), which affects both target recognition and degree of competitor activation (e.g., Altmann & Kamide, 1999; Kalénine et al., 2012; Kukona, Fang, Aicher, Chen, & Magnuson, 2011). This rapidly growing literature suggests that the effects of anticipation on lexical competition are not straightforward and there are open questions that may be fruitfully addressed by computational accounts (see also Chapter 4). For example, syntactic and pragmatic expectations about grammatical class (e.g., whether the speaker's next word is going to be an adjective or a noun) seem to constrain lexical competition such that nouns compete with nouns but not with adjectives (Magnuson, Tanenhaus, & Aslin, 2008). On the other hand, pragmatic "common ground" expectations seem to produce target anticipation without reducing competition from incompatible competitors, whereas semantic expectations produce both target anticipation and reduced competition from incompatible cohorts (Barr, 2008). Without a computational implementation, it is not clear how common ground constraints could be strong enough to produce robust anticipation effects without at least reducing (if not necessarily eliminating) competition from incompatible cohorts. There may be not-yet-discovered aspects of spoken word recognition dynamics that explain this, but investigating them requires a model that implements semantic, syntactic, and pragmatic constraints.

These phenomena are outside the scope of current models of spoken word recognition, but they do not necessarily require a radical overhaul of computational frameworks. For example, SRN models of sentence processing readily produce anticipation/prediction effects (Altmann & Mirković, 2009) and such effects can be captured using other approaches within the parallel distributed processing (PDP) framework (e.g., Kukona, Cho, Magnuson, & Tabor, 2014). The challenge may be integrating SRN models of spoken word recognition and sentence processing rather than developing a novel computational framework.

4.2 Auditory and Acoustic Realities

Just as the simplified output representations ignore the influence of higher-level factors, the simplified input to models of spoken word recognition also underestimates the way the realities of the speech signal affect spoken word recognition. Much like semantic, syntactic, and pragmatic context, phonological context also produces anticipation (e.g., DeLong, Urbach, & Kutas, 2005) and may similarly constrain the activation and competition dynamics (e.g., Tobin, Cho, Jennett, & Magnuson, 2010).

Models of spoken word recognition typically assume clear input and a cognitive-perceptual system uniquely engaged in spoken word recognition. In stark contrast, natural spoken word recognition takes place under a variety of "adverse conditions" (for a comprehensive review see Mattys, Davis, Bradlow, & Scott, 2012): word reductions and deletions (typical of casual speech, e.g., Dilley & Pitt, 2010), accented speech, disfluencies, energetic masking (background noise) and filtering (e.g., telephone transmission filters out low and high frequencies), and listener attention/distraction (cognitive load) or imperfect hearing. Generalization of the lexically-guided tuning effect described above appears to be influenced by allophonic similarity (Kraljic & Samuel, 2006; Mitterer, Scharenborg, & McQueen, 2013), suggesting that speech sound representations preserve aspects of auditory input beyond just abstract phoneme identity (see also Chapter 1). Sensitivity to systematic sub-phonemic differences can also simplify the spoken word recognition task – the association between vowel length and word length can help to distinguish embedded and carrier words such as "ham" from "hamster" (e.g., Salverda, Dahan, & McQueen, 2003). Addressing these issues from a computational perspective requires grappling with core theoretical questions such as: What are the representational units of speech perception? How do abstract and episodic representations jointly contribute to spoken word recognition? If immediate speech sound recognition is not possible in adverse conditions, how is the time course of lexical activation and competition affected?

As with top-down effects of semantic, syntactic, and pragmatic context, the SRN framework is likely to be able to capture at least some of these effects. SRNs would certainly produce phonological anticipation effects and the SRN model developed by Magnuson et al. (2003) would likely exhibit sensitivity to vowel length as a signal of word length if its input representation included this property of natural spoken words. In simulations of the TRACE model, addition of input noise increases the size of the rhyme competition effect (Mirman et al., 2011), providing a computational prediction regarding the effects of noise in the input speech signal or peripheral auditory perceptual impairment.

4.3 Cognitive Control

In models of spoken word recognition, "inhibition" is usually implemented as a domain-specific process such as inhibition between lexical representations, either by hand-coded inhibitory connections or emergent from representational overlap. In cognitive science more broadly, "inhibition" is typically framed as a higher-level executive or cognitive control function ("inhibitory control"). Studies that explore inhibitory control have used a mixture of non-linguistic tasks (e.g., the flanker task) and tasks that have a verbal component (e.g., the Stroop task), but have rarely made any meaningful contact with mechanisms of word processing. Instead, they seem to view inhibitory control as something that takes place outside the language system and modulates processing within the language system.

One might imagine that this is merely coincidental terminological overlap with no mechanistic implications, were it not for the fact that cognitive control is often invoked to explain aspects of spoken word processing, particularly individual differences in spoken word recognition. For example, larger inhibitory effects of lexical neighborhood density have been attributed to decreases in inhibitory control in normal aging (Sommers & Danielson, 1999; Sommers, 1996; Taler, Aaron, Steinmetz, & Pisoni, 2010), Alzheimer's type (Sommers, 1998), and aphasia (Botezatu & Mirman, 2014; Mirman & Graziano, 2013).

The role of cognitive control in spoken word recognition is also apparent in aphasic lexical-semantic "access deficits", in which word (or concept) knowledge is intact but access of this knowledge appears to be ineffective, inefficient, and inconsistent (for a detailed review see Mirman & Britt, 2014). These effects have been demonstrated in spoken word-to-picture matching and spoken-to-written word matching (and other tasks), but substantial extensions of existing models of spoken word recognition would be required to account for all of the relevant phenomena and there is no single account of access deficits to be integrated with models of spoken word recognition. On one view, access deficits arise from abnormal activation and deactivation (refractoriness) due to damage to neuromodulatory systems (Gotts & Plaut, 2002), but this model did not attempt to capture even basic aspects of spoken word recognition. Combining existing models of spoken word recognition with the computational model of neuromodulatory damage may provide a way toward a comprehensive account of access deficits and the role of cognitive control in spoken word recognition. As a starting point, Mirman et al. (2011) found that modulating the gain of the TRACE model's response selection mechanism (the slope of the non-linear relationship between relative activation and response probability) provided the best account for the trade-off between rhyme and cohort competition effects in 8 individuals with aphasia, consistent with the hypothesis that the deficit is related to cognitive control mechanisms outside the spoken word recognition system.

The contrast between the models of Gotts and Plaut (2002) and Mirman et al. (2011) is instructive. Gotts and Plaut proposed a sophisticated, biologically-based model of neuromodulatory cognitive control, but a highly simplified model of word recognition. Mirman et al. used a sophisticated, comprehensive model of spoken word recognition (TRACE), but a highly simplified model of cognitive control and response selection. Further efforts to bridge this gap and connect spoken word recognition with cognitive control may lead to substantial advances in our understanding of how spoken words are recognized.

4.4 Learning & Memory

Although models of spoken word recognition have captured some very important aspects of the interactions between spoken word recognition and learning and memory, these accounts focus on individual phenomena and are still a long way from capturing the complexity of how learning and memory impact spoken word recognition (see Chapter 9, this volume). One particularly interesting set of phenomena involve the effect of sleep-related memory consolidation of word learning and lexical structure. The importance of sleep for memory consolidation has been known for some time (Stickgold, 2013), but recent studies have demonstrated its particular role in word learning (for a review and theoretical perspective see Davis & Gaskell, 2009). Specifically, novel words can be learned quite quickly (e.g., Shtyrov, Nikulin, & Pulvermüller, 2010), but they do not begin to compete with known words until a period of consolidation has passed, typically overnight (Dumay & Gaskell, 2007, 2012; Henderson, Weighall, Brown, & Gaskell, 2012; Lindsay & Gaskell, 2013). Complementary learning systems (e.g., O'Reilly, Bhattacharyya, Howard, & Ketz, 2014) accounts of memory offer a framework for understanding this pattern: newly-learned words may be initially stored in a fast-learning hippocampal-dependent memory system, then more slowly integrated into the neocortical lexical-semantic system. Implementing this mechanism within models of spoken word recognition would require a more complex notion of the "mental lexicon" and would open the door to new predictions based on the complementary learning systems framework that could help to refine our understanding of how spoken word recognition interacts with learning and memory systems.

5 Toward a More Complete Model of Spoken Word Recognition

The TRACE model and SRNs have served as very powerful computational frameworks for understanding spoken word recognition. They account for dozens of behavioral phenomena, including some that were documented after the development of these models. These successes have relied, in part, on detailed investigations of a relatively narrowly-defined computational problem: identifying a lexical target from phonological input. After more than a quarter-century of highly productive behavioral, computational, and neuroscience research, it may be time to expand what we mean by "spoken word recognition". This chapter has identified four particularly promising domains for expansion based on their wealth of behavioral evidence and computational tractability.

First, whether isolated lexical representations exist or not, they are certainly not the end result of spoken word recognition, which is influenced by semantic, syntactic, and pragmatic factors. The TRACE model and SRNs are part of the PDP framework and other PDP models have already been developed for semantic (e.g., Rabovsky & McRae, 2014; Rogers & McClelland, 2004; Rogers et al., 2004) and syntactic (e.g., Altmann & Mirković, 2009; McClelland, St. John, & Taraban, 1989) processing, so integrating these models to develop a model of spoken word comprehension in sentence and discourse contexts is a natural next step.

Second, modeling speech input as a clear and unvarying signal has misrepresented the dynamics of real-world spoken word recognition and may have distorted our theories of spoken word recognition. There has been tremendous progress in automatic speech recognition (e.g., Cutajar, Micallef, Casha, Grech, & Gatt, 2013) and in understanding the relationship between speech perception and speech production (Hickok, 2012). A few computational modeling efforts have attempted to take advantage of perception-production links (Plaut & Kello, 1999) and insights from automatic speech recognition (Scharenborg, Norris, Bosch, & McQueen, 2005) and machine learning (Hannagan, Magnuson, & Grainger, 2013), but these approaches have not yet had wide-spread impact on models of spoken word recognition. The combined effort to include higher-level factors (semantic, syntactic, pragmatic) and lower-level (acoustic/auditory) factors will also move models of spoken word recognition away from stipulated representations and toward emergent representations, providing greater insights into the mechanisms involved.

Third, cognitive control functions are often invoked to explain phenomena in spoken word recognition with relatively little effort to specify these mechanisms or how they interact with spoken word recognition mechanisms. A prime example is “inhibition”, which can be used to refer to a property of the spoken word recognition system (e.g., inhibitory weights in the TRACE model), a cognitive control function (i.e., inhibitory control), or a general property of cognitive processing (i.e., “interactive activation and competition”). Perhaps inhibition is all three of these (a general property of cognitive processing, modulated by a cognitive control system, and playing an important role in spoken word recognition), but appeals to underspecified cognitive control mechanisms are generally not falsifiable and hold back development of theories of spoken word recognition. There are already quite detailed cognitive and neural models of cognitive control, and inhibitory control in particular (e.g., Munakata et al., 2011), which could potentially be integrated with models of spoken word recognition.

Fourth, models of spoken word recognition have been remarkably successful at capturing some aspects of learning in spoken word recognition but these efforts have been largely divorced from the broader cognitive, neurobiological, and computational research on learning and memory. TRACE models and SRN simulations have given us new insights into how learning new words affects spoken word recognition, but little insight into what it means to learn new words in relation to other kinds of learning and memory.

These four "zones of proximal development" are the most promising domains for expanding models of spoken word recognition, providing new insights into the mechanisms and representations that support spoken word recognition and making connections with other domains of perceptual and cognitive processing.

6 References

Allopenna, 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*(4), 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583–609. doi:10.1111/j.1551-6709.2009.01022.x

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*(1), 18–40. doi:10.1016/j.cognition.2008.07.005

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103*(5), 2670–2676.

Botezatu, M. R., & Mirman, D. (2014). A Link Between Lexical Competition and Fluency in Aphasia. *Frontiers in Psychology Conference Abstract: Academy of Aphasia – 52nd Annual Meeting.* doi:10.3389/conf.fpsyg.2014.64.00022

Chen, Q., & Mirman, D. (2012). Competition and Cooperation Among Similar Representations: Toward a Unified Account of Facilitative and Inhibitory Effects of Lexical Neighbors. *Psychological Review, 119*(2), 417–430. doi:10.1037/a0027175

Chen, Q., & Mirman, D. (2014). Interaction between phonological and semantic representations: Time matters. *Cognitive Science*, Manuscript in press.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes, 13*(2-3), 221–268. doi:10.1080/016909698386528

Cutajar, M., Micallef, J., Casha, O., Grech, I., & Gatt, E. (2013). Comparative study of automatic speech recognition techniques. *IET Signal Processing, 7*(1), 25–46. doi:10.1049/iet-spr.2012.0151

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*(1536), 3773–800. doi:10.1098/rstb.2009.0111

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition, 128*(3), 380–396. doi:10.1016/j.cognition.2013.05.007

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*(11), 1664–1670. doi:10.1177/0956797610384743

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35–39. doi:10.1111/j.1467-9280.2007.01845.x

Dumay, N., & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition, 123*(1), 119–32. doi:10.1016/j.cognition.2011.12.009

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5), 613–656. doi:10.1080/016909697386646

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45*(2), 220–266. doi:10.1016/S0010-0285(02)00003-8

Gotts, S. J., & Plaut, D. C. (2002). The impact of synaptic depression following brain damage: A connectionist account of “access/refractory” and “degraded-store” semantic impairments. *Cognitive, Affective & Behavioral Neuroscience, 2*(3), 187–213.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science, 18*(3), 254–260. doi:10.1111/j.1467-9280.2007.01885.x

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology, 4*(563), 1–17. doi:10.3389/fpsyg.2013.00563

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674–87. doi:10.1111/j.1467-7687.2012.01172.x

Hickok, G. S. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135–145. doi:10.1038/nrn3158

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), B23–B32. doi:10.1016/j.cognition.2004.10.003

Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional action knowledge during auditory comprehension of artifact words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1274–1295. doi:10.1037/a0027626

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262–268.

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 326–347.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119*(1), 23–42. doi:10.1016/j.cognition.2010.12.002

Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120392. doi:10.1098/rstb.2012.0392

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39*(2), 608–22. doi:10.1037/a0029243

Magnuson, J. S., Mirman, D., & Harris, H. D. (2012). Computational models of spoken word recognition. In M. J. Spivey, M. F. Joanisse, & K. McRae (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 76–103). New York, USA: Cambridge University Press.

Magnuson, J. S., Mirman, D., & Myers, E. B. (2013). Spoken Word Recognition. In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (pp. 412–441). New York, USA: Oxford University Press.

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition, 108*(3), 866–873.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953–978.

Mayor, J., & Plunkett, K. (2014). Infant word recognition: Insights from TRACE simulations. *Journal of Memory and Language, 71*(1), 89–123. doi:10.1016/j.jml.2013.09.009

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

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive Activation and Mutual Constraint Satisfaction in Perception and Cognition. *Cognitive Science, 38*(6), 1139–1189. doi:10.1111/cogs.12146

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(8), 363–369. doi:10.1016/j.tics.2006.06.007

McClelland, J. L., St. John, M. F., & Taraban, R. (1989). Sentence Comprehension: A Parallel Distributed Processing Approach. *Language and Cognitive Processes, 4*(3/4), 287–335.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60*(1), 1–39.

Mirman, D., & Britt, A. E. (2014). What we talk about when we talk about access deficits. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634). doi:10.1098/rstb.2012.0388

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory & Language, 59*(4), 475–494. doi:10.1016/j.jml.2007.11.006

Mirman, D., Graf Estes, K., & Magnuson, J. S. (2010). Computational Modeling of Statistical Learning: Effects of Transitional Probability Versus Frequency and Links to Word Learning. *Infancy, 15*(5), 471–486. doi:10.1111/j.1532-7078.2009.00023.x

Mirman, D., & Graziano, K. M. (2012). Individual differences in the strength of taxonomic versus thematic relations. *Journal of Experimental Psychology: General, 141*(4), 601–609. doi:10.1037/a0026451

Mirman, D., & Graziano, K. M. (2013). The Neural Basis of Inhibitory Effects of Semantic and Phonological Neighbors in Spoken Word Production. *Journal of Cognitive*

Neuroscience, 25(9), 1504–1516. doi:10.1162/jocn

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition, 37*(7), 1026–1039. doi:10.3758/MC.37.7.1026

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271–280. doi:10.1016/j.cognition.2008.02.003

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13*(6), 958–965. doi:10.3758/BF03213909

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language, 117*(2), 53–68. doi:10.1016/j.bandl.2011.01.004

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition, 129*(2), 356–61. doi:10.1016/j.cognition.2013.07.011

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences, 15*(10), 453–459. doi:10.1016/j.tics.2011.07.011

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary Learning Systems. *Cognitive Science, 38*(6), 1229–1248. doi:10.1111/j.1551-6709.2011.01214.x

Patterson, K. E., & Plaut, D. C. (2009). "Shallow Draughts Intoxicate the Brain": Lessons from Cognitive Science for Cognitive Neuropsychology. *Topics in Cognitive Science*, 1, 39–58.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 381–416). Mahweh, NJ, USA: Erlbaum.

Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition, 132*(1), 68–89. doi:10.1016/j.cognition.2014.03.010

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. E. (2004). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review, 111*(1), 205–235.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. Vol. 1, 318–362). Cambridge, MA: MIT Press.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110–114.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*(1), 51–89.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*(6), 1207–1218. doi:10.3758/APP

Scharenborg, O., Norris, D., Bosch, L., & McQueen, J. M. (2005). How should a speech recognizer work? *Cognitive Science, 29*(6), 867–918. doi:10.1207/s15516709cog0000_37

Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology, 27*(6), 477–494. doi:10.1080/02643294.2011.574111

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory & Language, 54*(2), 228–264. doi:10.1016/j.jml.2005.10.001

Shtyrov, Y., Nikulin, V. V, & Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *The Journal of Neuroscience, 30*(50), 16864–7. doi:10.1523/JNEUROSCI.1376-10.2010

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11*(2), 333–341.

Sommers, M. S. (1998). Spoken word recognition in individuals with dementia of the Alzheimer's type: changes in talker normalization and lexical discrimination. *Psychology and Aging, 13*(4), 631–646.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14*(3), 458–472.

Stickgold, R. (2013). Parsing the role of sleep in memory processing. *Current Opinion in Neurobiology, 23*(5), 847–53. doi:10.1016/j.conb.2013.04.002

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39*(1), 19–30.

Taler, V., Aaron, G. P., Steinmetz, L. G., & Pisoni, D. B. (2010). Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 65*(5), 551–60. doi:10.1093/geronb/gbq039

Tobin, S. J., Cho, P. W., Jennett, P. M., & Magnuson, J. S. (2010). Effects of anticipatory coarticulation on lexical access. *Proceedings of the 33rd Annual Cognitive Science Society Meeting*. Retrieved from http://palm.mindmodeling.org/cogsci2010/papers/0517/paper0517.pdf

Walley, A. (1993). The Role of Vocabulary Development in Children's Spoken Word Recognition and Segmentation Ability. *Developmental Review, 13*(3), 286–350. doi:10.1006/drev.1993.1015

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American, 223*(6), 30–36.

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *WIREs Cognitive Science, 3*(3), 387–401. doi:10.1002/wcs.1178

Yee, E. (2005). *The time course of lexical activation during spoken word recognition: Evidence from unimpaired and aphasic individuals.* Brown University, Providence, RI.

Yee, E., & Sedivy, J. C. (2006). Eye Movements to Pictures Reveal Transient Semantic Activation During Spoken Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 1–14.