Interactions of length and overlap in the TRACE model of spoken word recognition

James S. Magnuson (james.magnuson@uconn.edu)
Department of Psychological Sciences, University of Connecticut
Storrs, CT 06269-1020 USA

Elizabeth Schoen Simmons (elizabeth.a.simmons@uconn.edu)
Department of Psychological Sciences, University of Connecticut
Storrs, CT 06269-1020 USA

Abstract

What determines degree of competition among phonologically similar words? One proposal is that proportion of overlap predicts competition independently of word length. We argue that proportion of overlap may provide descriptive adequacy, but does not provide an explanation. We show that TRACE correctly predicts patterns previously attributed to proportion of overlap. In additional simulations, with independent manipulations of word length and proportion of overlap, proportion of overlap fails to predict the full pattern of results. We discuss how competition dynamics in TRACE modulate competition as word length and proportion of overlap change. These results have implications for theories of human spoken word recognition, and will motivate experiments to test these new TRACE predictions.

Keywords: spoken word recognition; computational models

Introduction

The nature of similarity mapping is a crucial question in human spoken word recognition. The time course of competition between different types of competitors provides important clues to the mechanisms underlying human speech recognition. Two types of competitors that have been important in the spoken word recognition literature are *cohorts* and rhymes. *Cohorts* overlap at onset. They are called "cohorts" because they (words overlapping in approximately the first 200 ms or first 2 phonemes) constitute the competitor set of the Cohort model (Marslen-Wilson & Welsh, 1978). *Rhymes* have the same number of syllables, and match in all positions except onset (for example, *beaker, speaker*. Thus, when multisyllabic, they need not correspond to linguistic *rime* (syllable nucleus and coda) or poetic rhyme (approximately, overlap from the final stressed vowel onward). Because the Cohort Model predicted that only items overlapping at onset could enter the recognition cohort, rhymes – words with complete overlap everywhere but onset – were a crucial test case. Marslen-Wilson and colleagues (e.g., Marslen-Wilson & Zwitserlood, 1989) tried to find evidence for rhyme activation in tasks like gating and cross-modal semantic priming. They generally found that, at best, a word might be slightly activated by a rhyme form *if* that form was not a real word, and it differed by no more than a single acoustic-phonetic feature at onset (e.g., *bleasant* can weakly activate *pleasant*).

However, there were other indications that rhymes might receive significant activation in spoken word recognition. There is similar priming from 4-phoneme words that overlap in 1-3 phonemes, with mismatch at offset (cohorts) or onset (rhymes, so long as the matching portion includes the vowel; Slowiaczek, Nusbaum, & Pisoni, 1987). According to the *Neighborhood Activation Model* (NAM; Luce & Pisoni, 1998), words differing by no more than a single phonemic deletion, addition, or substitution (the so-called "DAS rule") are sufficiently similar that the form for one should also activate the other. Notably, the DAS rule includes many items that Cohort excludes (e.g., rhymes), and excludes many items Cohort would include (cohort pairs differing by more than 1 phoneme, e.g., *cat, castle*, and *cadaver* are cohorts but not neighbors). Nonetheless, frequency-weighted neighborhoods based on NAM's DAS rule provide one of the best extant predictors of spoken word recognition facility for large sets of words. Thus, in the 1990s, the field faced a conundrum: using methods designed to investigate activations of word pairs, there was almost no evidence for (lexical) rhyme activation, but there was evidence for facilitory priming from rhymes (Slowiaczek et al., 1987) and indirect evidence that rhymes contribute to lexical competition (Luce & Pisoni, 1998) from other methods.

Allopenna, Magnuson, and Tanenhaus (1998) investigated these inconsistencies using the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). They gave participants spoken instructions to select items from a 4-alternative display (e.g., *click on the beaker*) as they tracked their eye movements. On critical trials, the display included a potential cohort competitor (e.g., *beetle*) and/or a rhyme (e.g., *speaker*), along with phonologically unrelated baseline items (e.g., *carriage*). Approximately 200 ms after word onset, fixations shifted to targets and cohorts (as expected, since both match the initial input). About 200 ms after the disambiguating phoneme (e.g., the /k/ in *beaker*), fixation proportions to cohorts began to decline, while fixations to the rhyme began to increase. Although the rhyme fixation proportion was significantly greater than that for the unrelated baseline, the rhyme peak was later and lower than the cohort peak.

Allopenna et al. (1998) conducted simulations with the TRACE model (McClelland & Elman, 1986). TRACE is an interactive activation neural network, where over-time phonetic feature nodes pass activation to phoneme nodes, which pass activation to word nodes, which send feedback

Table 1: Example critical items based on Simmons and Magnuson (under revision). One example (from nine pairs) is shown for each combination of conditions.

Type	Length	Item 1	Item 2	Unrelated
Cohort	Short	bat	bas	ril
Cohort	Long	kasdit	kaprʌd	gubluk
Rhyme	Short	kak	rak	lʌp
Rhyme	Long	sudʌls	pudʌls	gatikʃ

to constituent phonemes. There is lateral inhibition at each level, which governs competition. TRACE predicts a time course of activation and competition remarkably similar to patterns of eye movements. The crucial pattern (early, strong cohort competition, and later, weaker rhyme competition) emerges in TRACE due chiefly to lateral inhibition. Although rhymes have greater global similarity with a target, their activation is suppressed by lateral inhibition from the target and its cohorts, which get a literal head start due to earlier overlap with the bottom-up input.

McQueen and Viebahn (2007) extended Allopenna et al. (1998) in 4 ways: they used Dutch, they used text as visual referents rather than images, they manipulated word length (with short vs. long pairs), and they held proportion of overlap constant for cohorts and rhymes. Examples of each class include short cohorts, *tor, tol*; long cohorts, *geloof, geloop*; short rhymes, *rat, lat*; and long rhymes, *rotje, lotje*). They replicated the basic pattern found by Allopenna et al. (early, strong cohort competition, and later, weaker rhyme competition), but with an effect of length: there was stronger competition for longer cohort and rhyme pairs (though the latter was weak). McQueen and Viebahn proposed that *proportion of overlap* determines lexical competition. However, their results cannot distinguish between a word length (and total overlap) vs. proportion of overlap, as length, total overlap, and proportion of overlap were correlated in their design (due to their use of "maximal cohorts" that overlap in all but the final phoneme, in contrast to more conventional cohort definitiosn, such as overlap in the first 2 phonemes).

Recent work by Simmons and Magnuson (under revision) may be consistent with the proportion of overlap principle. We also manipulated word length. However, unlike McQueen

Table 2: Example critical items based on McQueen and Viebahn (2007). The only change compared to the examples in Table 1 is that long cohort pairs mismatch only in their final phonemes.

Type	Length	Item 1	Item 2	Unrelated
Cohort	Long	kasdit	kasdip	gubluk

Table 3: Example critical items from new proportion manipulation. Items have 67% or 75% proportion overlap.

Type	Length	Overlap	Item 1	Item 2	Unrelated
Cohort	Short	67%	kit	kib	sul
Cohort	Short	75%	guʃi	guʃʌ	kard
Cohort	Long	67%	tabula	tabuki	kipirg
Cohort	Long	75%	rulʃapak	rulʃapʌt	dirtʌsub
Rhyme	Short	67%	lab	ʃab	kus
Rhyme	Short	75%	iʃug	ʌʃug	kard
Rhyme	Long	67%	alubat	ikubat	kipirg
Rhyme	Long	75%	tʌpaʃlur	kapaʃlur	busʌtrid

and Viebahn (2007), we used the *conventional* cohort definition (overlap in first 2 phonemes). Thus, shorter pairs (e.g., *bus, bug*) had greater proportion of overlap than longer pairs (e.g., *bubbles, butter*). Proportion of overlap rather than total overlap predicted degree of competition: rhyme effects were larger for longer words, but cohort effects were larger for shorter pairs. Simmons and Magnuson confirmed that TRACE predicts this pattern, *and* that TRACE predicts the results of(McQueen & Viebahn, 2007), with analogs of their items. (They also make a case for focusing on conventional cohorts because they are known to compete strongly (Allopenna et al., 1998), and are much more common: 99.7% of English words have at least 1 such cohort, while "maximal" cohorts like those used by McQueen and Viebahn are much rarer; only 19% of English words have at least 1 such cohort.)

However, while proportion of overlap may *describe* some results, it does not provide an explanation. Simmons and Magnuson found greater competition for shorter cohort pairs even though shorter and longer pairs overlapped only in the first 2 phonemes in their materials. Either the system somehow "knows" how long the word will be, or it normalizes evidence by word length (as implied by McQueen and Viebahn, although they did not discuss potential mechanisms for such normalization). Simmons and Magnuson described how TRACE emergently normalizes by word length due to the dynamics of lateral inhibition. However, they did not address the challenge of *dissociating* proportion of overlap from total overlap as word length increases. We next report simulations with independent manipulations of length and proportion of overlap aimed at addressing this challenge.

Simulations

Simulations were conducted with the TRACE (McClelland & Elman, 1986) C code distributed by Jay McClelland, modified for compatibility with current compilers. Our emphasis is on a new manipulation of proportion of overlap, but we begin with simulations using materials like those of Simmons and Magnuson (under revision) and McQueen and Viebahn (2007) to establish a basis for comparison.

Materials

The 3 sets of simulations used critical items added to the original 212-word TRACE lexicon, expanding it to 300 words. For examples of critical items based on Simmons and Magnuson, see Table 1, and Table 2 for items based on McQueen and Viebahn. For these items, there were 9 item sets in each combination of length and competitor type. Cohorts conformed to a conventional definition (overlap in first 2 phonemes). For simulations of McQueen and Viebahn, long cohort pairs were altered to differ only in the final phoneme. Finally, in Table 3, we present examples of new materials: cohorts and "pseudo-rhymes" with length and proportion of overlap manipulated separately. Note that although the original TRACE lexicon is based on real English words, the words in the tables were constructed with a focus on precise control of similarity, *without* consideration of whether the items have real English word analogs.

Defining phonological competitors *Cohorts* overlap at onset, and are called "cohorts" because they constitute the *recognition cohort* in the Cohort Model (Marslen-Wilson & Welsh, 1978). Cohorts are limited to items overlapping at onset due to a theoretical commitment (*the algorithm for word recognition should make maximal use of bottom-up information*) grounded in empirical findings (strong evidence for mutual activation of items overlapping in approximately the first 200 ms at a typical speaking rate [first 2 phonemes], and virtually no evidence of activation for items that mismatch at onset, even if they match at all subsequent positions; (Marslen-Wilson & Zwitserlood, 1989)). *Rhymes*, again, as operationally defined in the human spoken word recognition literature, are items that overlap everywhere but word onset (vs. overlap in a single syllabic rime [though they overlap from the first rime onward] or poetic rhymes, which can mismatch in length so long as they overlap from the final stressed vowel onward). We revisit these definitions because our materials include 3 cases where we deviate from these conventional operational definitions.

First, McQueen and Viebahn created unconventional cohorts that were matched to rhymes in terms of amount of overlap (and amount of mismatch): their Dutch words matched everywhere but the final position (e.g., short cohort pairs such as *tor, tol*, and long cohort pairs like *geloof, geloop*. Even though previous studies (e.g., (Allopenna et al., 1998)) had demonstrated stronger competition for cohorts than rhymes in the VWP, this was an important and useful control that complements other work on phonological competition in the VWP. However, by allowing a maximum of one phonemic mismatch in cohorts (and rhymes), it entails that *proportion of overlap* is confounded with word length in such items. Note that they claim that proportion of overlap is a determining principle of degree of competition (McQueen & Viebahn, 2007); while their results are consistent with that claim, they cannot provide a critical test since their materials confound length and proportion of overlap. The second and third deviations are in the new materials in Table 3, where we have cohorts and "pseudo"-rhymes that vary in proportion of overlap at different word lengths. The items overlap by 2/3 (literally 2 of 3 phonemes in short words, and in 4 of 6 phonemes in longer words) or 3/4 (3 of 4 phonemes in shorter words, or 6 of 8 phonemes in longer words).

Predictions

Critically, the new items in Table 3 allow us to examine interactions of proportion of overlap and length. If proportion of overlap truly predicts degree of competition (McQueen & Viebahn, 2007), we should see greater competition for 75% overlap than 67% overlap for *both* shorter and longer pairs. If instead, as suggested in (Simmons & Magnuson, under revision), the time course of lexical activation and competition follows possible non-intuitive trajectories due to complex interactions of lateral inhibition and degree of bottom-up support, we may observe interactions of word length and proportion of overlap.

Procedure

We conducted 1 simulation for every target. For Simulations 1 and 2, there were 72 targets: 9 pairs (2) x length (short, long) x phonological competitor type (cohort, rhyme). For Simulation 3, there were 64 targets: 4 pairs x length x competitor type x proportion of overlap (67%, 75%). For each simulation, we tracked target, competitor, and unrelated baseline activations. Results from all simulations within a condition were averaged to provide summary plots (Figures 1-4).

Results

The key results are shown in Figures 1-4. In Figure 1, where amount of mismatch is held to one phoneme, and so proportion of overlap increases with word length, we see greater competition for longer words, which have greater proportion of overlap (the CminusU lines are higher in lower than upper panels), although the difference is quite small for rhymes. This is consistent with the human performance data of (McQueen & Viebahn, 2007), which was the basis for their claim that degree of competition depends on proportion of overlap. In Figure 2, which plots the results from simulations where the materials used the conventional cohort definition of match in the first two phonemes, the rhyme results are identical (they are the same results shown for rhymes in the previous figure), but the cohort trends have flipped: now there is greater competition for *shorter* pairs, consistent with proportion of overlap (2 of 3 phonemes vs. 2 of 5 for longer words).

In Figure 3, the results are from new simulations with cohort items varying in length and proportion of overlap (see Table 3). There are clear main effects of length, with greater competition for longer words (compare top to lower panels), and greater competition with greater proportion of overlap (compare left to right panels). However, note that there proportion of overlap does not dictate the *absolute* amount of competition: there is slightly greater competition for long

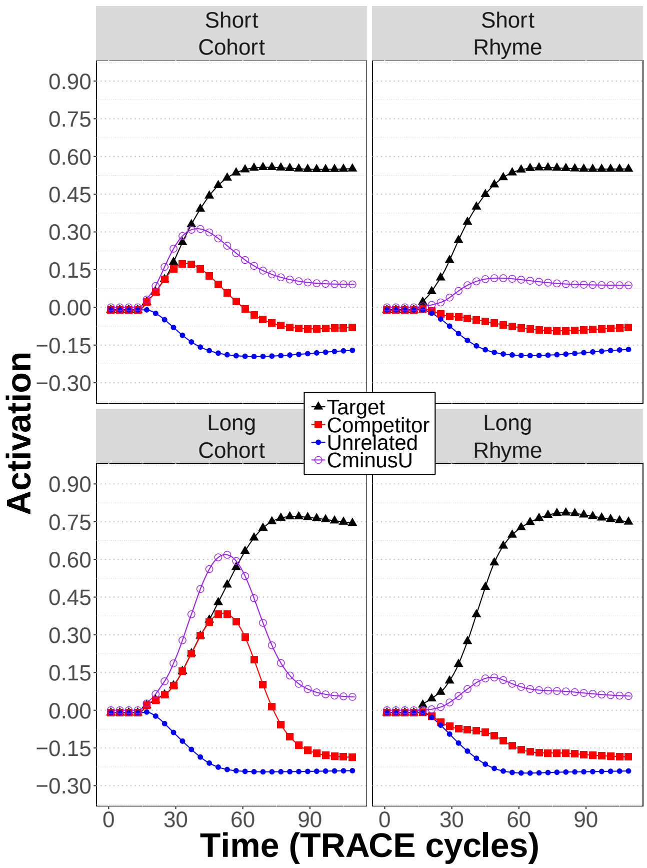

Figure 1: Simulations of (McQueen & Viebahn, 2007), with mismatch constant (1 phoneme, and thus higher proportion of overlap for longer pairs). "CminusU" = competitor - unrelated baseline. Effects are stronger for longer items (compare CminusU in top vs. bottom panels), consistent with the claim that proportion of overlap determines degree of competition (McQueen & Viebahn, 2007). *For examples of items, see Tables 2 (for Long Cohort items) and 1.*

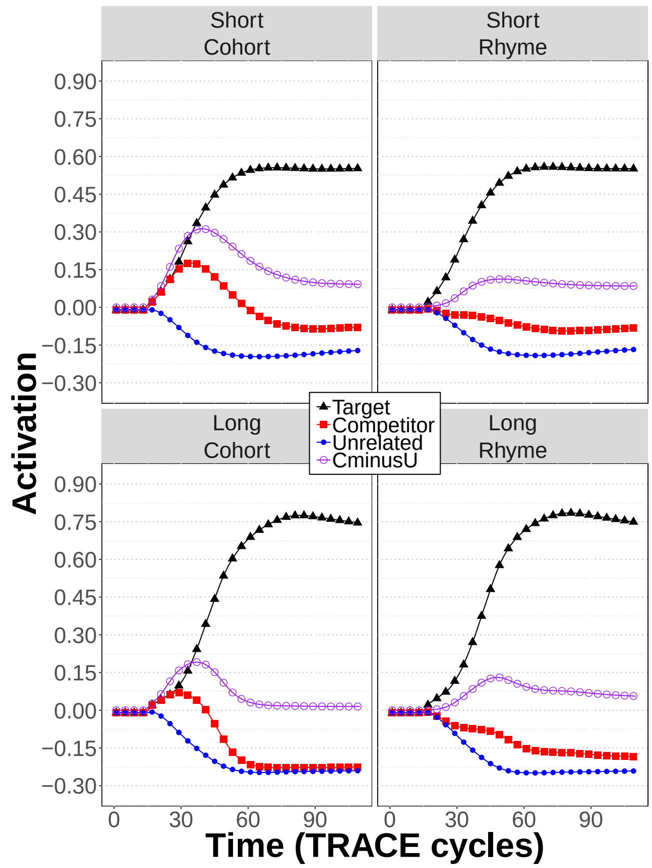

Figure 2: Simulations of (Simmons & Magnuson, under revision). Effects are stronger for longer rhymes and shorter cohorts. Because degree of **match** was constant for cohorts, shorter pairs had higher proportion of overlap. Thus, this pattern is also consistent with the claim that degree of competition follows from proportion of overlap (McQueen & Viebahn, 2007). *For examples of items, see Table 1.*

pairs with 67% overlap than short pairs with 75% overlap. In Figure 4, the results are from the new simulations with "pseudo"-rhymes varying in length and proportion of overlap. Here, the pattern is different. There is *less* competition for longer than shorter pairs (compare top to bottom panels), but still greater competition for greater proportion of overlap (though there is virtually no difference for longer pairs).

Discussion and conclusions

As suggested by Simmons and Magnuson (under revision), and contra McQueen and Viebahn (2007), proportion of overlap does not determine amount of competition – in TRACE. Of course, these results are thus far restricted to the TRACE model, and TRACE's predictions must be tested against human performance. Nonetheless, it is notable that TRACE predicts the correct pattern for both previous experiments with human subjects discussed above where degree of competition *was* consistent with the proportion of overlap principle (Simmons & Magnuson, under revision; McQueen & Viebahn, 2007; see Figures 1 and 2, respectively). Thus, the principle of proportion of overlap cannot *describe* the full range of outcomes we observe (with TRACE). Furthermore, as discussed by Simmons and Magnuson (under revision), even when TRACE's results are consistent with the principle of proportion of overlap, proportion of overlap does not, by itself, constitute an *explanation*. This is true in particular because we are considering an input signal that emerges over time. Consider cohort pairs that are matched on *amount* of overlap (2 phonemes) rather than proportion of overlap, as in Figure 2. In these cases, cohorts match targets through the second position. The model does not "know" how long the words will be, and yet there is greater competition for shorter pairs where two phonemes represents a greater proportion of overlap. Why?

In TRACE, the answer follows from lateral inhibition.

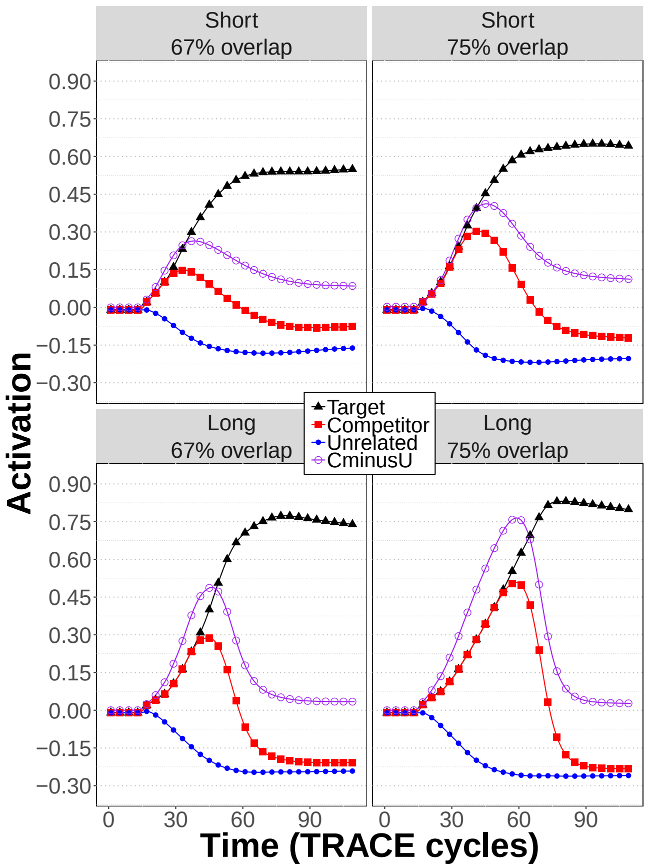

Figure 3: Simulation 3, interaction of word length and proportion of overlap for **cohort** pairs. *For example items, see Table 3.*

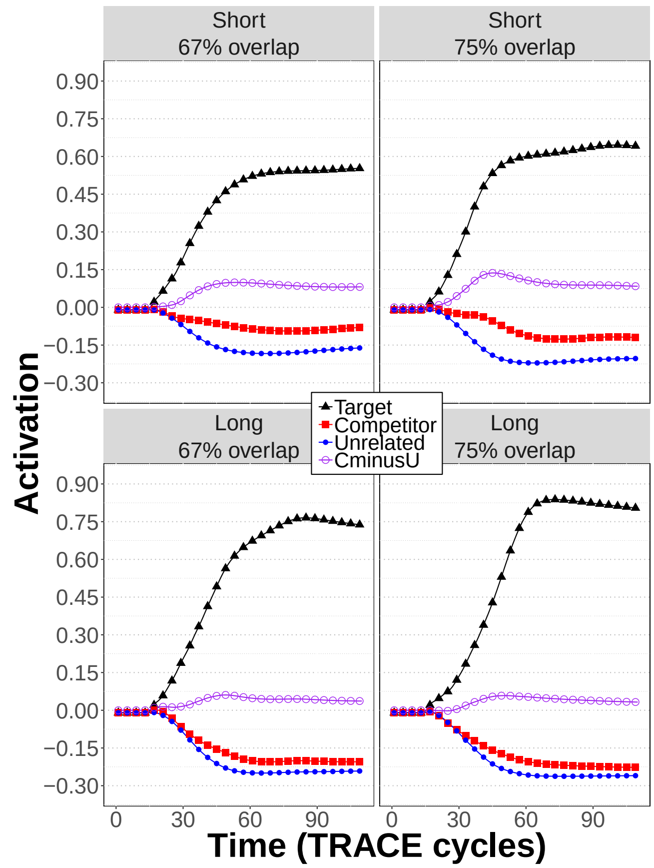

Figure 4: Simulation 3, interaction of word length and proportion of overlap for **rhyme** pairs. *For example items, see Table 3.*

Word units have a literal temporal extent in TRACE's memory, and words have inhibitory connections to all words they overlap with in "time" (position) in TRACE's memory. Thus, longer words have more incoming inhibitory connections (because they span more positions and therefore overlap with more words). Even tiny amounts of activation in the sending units can have non-trivial impact on a word's activation. This leads to longer cohorts being less able to compete, because they are receiving proportionally more inhibition than shorter cohorts (not just from the target, but from any word with any non-zero activation). When length and proportion of overlap are independently manipulated, as in Figure 3, similar or greater competition can occur with less proportion of overlap for longer words (e.g., the greater competition for 67% overlap long cohorts vs. 75% overlap short cohorts).

Inhibition is also crucial for understanding the dynamics of rhyme competition. In Figures 1 and 2 (where the same rhyme simulations are presented in the right column of each figure), there is only a slight increase in competition for longer rhymes, despite the fact that there is a large increase in proportion of overlap, as shorter pairs share only 2/3 phonemes, while longer pairs share 5/6. This is because as word length increases, several interactions occur. In addition to the one mentioned above (longer words have more inhibition sites in TRACE), the other major impact is that the amount of inhibition target words can *send* increases with word length, because longer words simply receive more bottom-up support and reach higher levels of activation.

Thus, *proportion of overlap* cannot explain, nor even describe, the full range of outcomes that follow from word length, type of phonological similarity, and degree of similarity in lexical access. In many cases, a model with a simple architecture but many simple interacting elements (like TRACE) will defy intuition, and exhibit complex emergent behaviors. Simulations are *often* ***required*** to elucidate the actual behavior of the model.

Of course, now that we have new predictions in hand from TRACE (for independent manipulations of onset and offset overlap, as in Figures 3 and 4), the next step will be to test these predictions with human subjects. We are developing an experiment based on these simulations currently.

Acknowledgments

This project was supported by U.S. National Science Foundation grant 1754284 (*Computational Approaches to Human Spoken Word Recognition*) to JSM and U.S. National Institutes of Health award F31 DC018220 to ESS. We thank Eleanor Magnuson for graphic design advice.

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 and Language*, *38*, 419–439.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

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

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance*, *15*, 576–585.

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

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*(5), 661–671.

Simmons, E. S., & Magnuson, J. S. (under revision). Word length, proportion of overlap, and the time course of phonological competition in spoken word recognition: An empirical and computational investigation.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *13*(1), 64–75.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.