University of Rochester Working Papers in the Language Sciences—Vol. Spring 2000, no. 1
Katherine M. Crosswhite and Joyce McDonough (eds.)

Simple Recurrent Networks and Competition Effects in Spoken Word Recognition

James S. Magnuson (magnuson@bcs.rochester.edu)
Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Richard N. Aslin (aslin@cvs.rochester.edu)
Department of Brain and Cognitive Sciences
University of Rochester, Meliora Hall, Rochester, NY 14627 USA

Abstract

Continuous mapping models of spoken word recognition such as TRACE (McClelland and Elman, 1986) make robust predictions about a wide variety of phenomena. However, most of these models are interactive activation models with preset weights, and do not provide an account of learning. Simple recurrent networks (SRNs, e.g., Elman, 1990) are continuous mapping models that can process sequential patterns and learn representations, and thus may provide an alternative to TRACE. However, it has been suggested that the features that allow SRNs to learn temporal dependencies lead them to work much like the Cohort model (e.g., Marslen-Wilson, 1987), such that items are activated by onset similarity to an input, but not by offset similarity (Norris, 1990). This would make them incompatible with TRACE and with recent results indicating that words that rhyme compete during spoken word recognition (Allopenna, Magnuson and Tanenhaus, 1998). We present simulations demonstrating that rhyme effects do emerge in SRNs, but this depends on how the training is carried out. We also find that SRN predictions provide a good fit to a series of recent studies of the time course of competition effects in spoken word recognition, including cohort, rhyme, and neighborhood density effects.

Introduction

Continuous mapping models of spoken word recognition (so called because lexical items are activated continuously as a function of their similarity to an input stimulus, without explicit consideration of word boundaries) such as TRACE (McClelland and Elman, 1986) have proven to be robust models of a wide range of spoken word recognition phenomena. However, most continuous mapping models are interactive activation models, in which the weights of connections

between units are preset on the basis of theoretical assumptions. While TRACE, for example, can be criticized as unrealistic in several respects (see Norris, 1994), we find the largest draw-back of interactive activation models to be their obvious inability to model learning and development.

Simple recurrent networks (SRNs) are another sort of continuous mapping model that also learn. SRNs are similar to standard feed-forward networks, but have an added component: a set of context units that contain a copy of the hidden unit activations from the previous time step (the SRN architecture is described in more detail below). This feature allows SRNs to learn a broad range of sequentially-dependent inputs.

Norris (1990) reported that, as one might expect given sequential stimuli, SRNs show a "left-to-right" bias: in Norris' simulations, words that overlapped at onset with an input became active, but words which overlapped at offset did not. Such performance would be consistent with the Cohort model (e.g., Marslen-Wilson, 1987), in which a "cohort" of possible matches to an input is winnowed down to a unique match by removing items as they mismatch with the input, in a left-to-right manner. Because onsets come first, the Cohort model predicts a large bias towards activation of items sharing onsets, and against activation of items mismatching at onset, even given later overlap. This means that rhymes, such as *beaker* and *speaker*, are not predicted to activate one another.

TRACE, on the other hand, predicts that rhymes should compete, which is consistent with recent empirical work: Allopenna, Magnuson and Tanenhaus (1998) reported evidence of robust rhyme competition during spoken word recognition. We are concerned about this reported discrepancy between SRNs and TRACE because the prediction of rhyme effects provides a critical distinction between continuous mapping models and "alignment" models like Cohort (so-called because of their emphasis on finding or assuming word boundaries).

In the next section, we discuss the differences between continuous mapping and alignment models with respect to rhyme competition, and briefly review the empirical evidence for rhyme competition. Then, we present SRN simulations using Norris' (1990) materials, followed by simulations of the results of Allopenna et al., as well as more recent work by Magnuson and colleagues (Magnuson, Dahan, Allopenna, Tanenhaus, & Aslin, 1998; Magnuson, Tanenhaus, Aslin & Dahan, 1999).

Cohort and Rhyme Effects and Models of Spoken Word Recognition

Theories of spoken word recognition agree on a broad set of basic assumptions: given a spoken word, multiple lexical candidates are activated and compete for recognition. The degree to which items become active depends on their similarity to the input, as well as other characteristics (e.g., their frequency of occurrence).

Where models tend to differ is in the set of candidate words predicted to become active. One division that can be made is between *alignment* and *continuous mapping* (or continuous *activation*) models. Alignment models (e.g., Cohort: Marslen-Wilson, 1987; and Shortlist: Norris, 1994) postulate mechanisms which actively seek (or assume) word boundaries. In the Cohort model, candidates are evaluated as to how well they match an input word beginning from word onset. Activations are greatly reduced given mismatches between input and candidate.

Continuous mapping models give no special consideration to word onsets. Instead, items become active as a function of their moment-to-moment similarity to the input, with no explicit penalty for mismatches. The term *continuous mapping* is potentially confusing. It does not simply mean the model continuously provides an output. For example, TRACE is a continuous mapping model, but effectively becomes an alignment model when its explicit end-of-word "silence phoneme" is used to mark word boundaries.[1] Similarly, while the interactive activation and competition decision level of Shortlist provides continuous output, Shortlist is very much an alignment model, since mismatches are explicitly penalized based on aligning a candidate word with a known word boundary.

One might expect that explicitly searching for word boundaries would be an efficient or even optimal strategy. But consider the variability we experience in using spoken language. We recognize speech in countless circumstances where the acoustics of speech vary tremendously: outdoors, in stairwells, with different talkers, who might have different accents, or who might have just taken a bite out of a hamburger. A recognition mechanism optimized for clear speech (where word boundaries will still be difficult to find) may spend most of its time reanalyzing mis-segmented speech. A system which does not tie itself to word boundaries might prove more robust, since a wider range of possible matches to the input will be considered.

One result of the differences between continuous mapping and alignment models is a contrast in whether or not rhymes are predicted to compete. Both types of model predict that words sharing onsets will compete. Alignment models, because of the emphasis on mismatches, predict that candidates that mismatch at onset will compete weakly only if the initial mismatch is small (with evidence suggesting the mismatch can be no larger than one or two features; e.g., Connine, Blasko & Titone, 1993). However, because activation in continuous mapping models depends on overall similarity with no privilege given to onset or offset, rhymes are predicted to compete (although, e.g., in TRACE, they will be less strongly

[1] TRACE uses a brute force approach to solve the alignment problem: it employs multiple detectors for each word, aligned with different temporal positions.

activated than words sharing onsets, since onset competitors will inhibit the rhymes before the input begins to overlap with the rhymes).

Until recently, the empirical record favored alignment models; evidence for rhyme activation was weak at best. However, Allopenna, Magnuson and Tanenhaus (1998) reported robust rhyme effects using the recently developed "visual world paradigm" (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, participants respond to spoken instructions to move objects in a visual display, and their eye fixations are measured continuously. Fixations turn out to be tightly time-locked to speech -- at least given a task in which visually guided movements are required (which avoids problems of interpretation raised by Viviani, 1990, since the eye movements have a functional interpretation; see Allopenna et al. for more discussion). Fixations to objects whose names are similar to an input word begin as early as 200 ms after the onset of the input word. In very simple tasks, participants require approximately 150 msec to plan and launch a saccade (e.g., Matin, Shao, & Boff, 1993). Allowing for this planning time, the earliest eye movements are being planned approximately 100 msec after target onset. Thus, these fixations are indeed closely time-locked to the speech (compare them to minimal response times of about 200 ms after the *offset* of monosyllabic words in lexical decision).

Figure 1 shows the TRACE predictions and observed fixation patterns for critical trials from Allopenna et al. (1998). On most trials, subjects were asked to "click on" a target item (using the computer's mouse) displayed with three unrelated items. On critical trials, onset competitors (referred to as "cohorts" because they are the items predicted to compete in the Cohort model) and/or

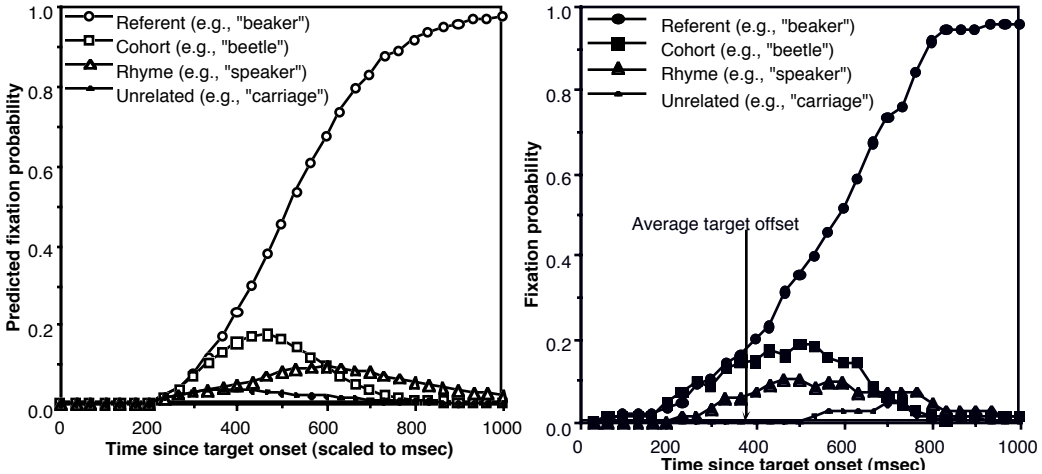

Figure 1: TRACE activations converted to predicted fixation probabilities (left panel) and observed probabilities of fixating a target referent, a cohort member, a rhyme, and an unrelated object (right panel) from Allopenna et al. (1998).

rhyme competitors were present. TRACE activations were transformed into predicted fixation probabilities using a variant of the Luce choice rule (see Allopenna et al.). As TRACE predicts (TRACE accounts for greater than 90% of the variance in each of the critical items), the data indicate that items compete for recognition as a function of their similarity to a stimulus over time, and even substantial initial mismatches do not block rhyme activation (since all of the rhymes differed by more than two features).

What do SRNs predict? Norris (1990) reported that the performance of SRNs is consistent with the Cohort model, since he found evidence of cohort (onset) competition, but not offset competition. We will begin our examination of SRNs by training an SRN to recognize the words in Norris' (1990) training lexicon.

Simulation 1: Norris (1990)

The basis for Norris' claim was a simulation using a 48-word lexicon consisting of phonemic transcriptions of 24 real words, plus 24 non-words created by reversing the transcriptions of the real words. This meant that for each pair overlapping at onset, there was a corresponding pair overlapping in the same segments (in reverse order) at offset.

The 24 real words were *baker*, *beat*, *boot*, *border*, *bounded*, *calm*, *cold*, *coroner*, *coronet*, *damp*, *delimit*, *deliver*, *dish*, *disk*, *door*, *fear*, *finish*, *flash*, *heap*, *hurt*, *pound*, *taker*, *trash*, and *tripe*. We used phonemic transcriptions of these, and reversals of the transcriptions, as our lexicon. Our phonemic transcriptions differed from Norris'. His were coded phoneme-by-phoneme, using a set of 11 features. Ours were coded similarly, but using a set of 18 features derived from O'Grady et al. (1989). Otherwise, our simulations were quite similar. Figure 2 shows a schematic of the SRN we used. Before describing the simulation further, we will briefly describe the network we used and some general properties of SRNs.

SRNs are nearly identical to standard feed-forward networks trained using backpropagation. The innovation that makes them good candidates for learning sequential patterns is the use of a *context layer*. In Figure 2, solid arrows between layers indicate full trainable connectivity (each unit in the lower layer has a weighted connection to each unit in the upper layer, and the weights on those connections can be modified during training). The dashed arrow indicates non-modifiable "copy-back" connections between the hidden layer and the context layer. At each time step, part of the input to the hidden layer is from the context layer. Context layer activation is a direct, one-to-one copy of the hidden layer activation from the *previous time step*. This allows the network to react not just to the current input, but also to its own state at the previous time step, *t*-1 -- and its state in multiple preceding time steps, since its hidden layer activation at the

previous time step would have been influenced by its input from the context layer containing copies of hidden unit activations at time *t*-2, and so on.

Initially, all trainable weights are set to small, random values. The weights are then modified as each input is presented using backpropagation. Activation from one layer is passed through weighted, trainable connections to the next layer; input and context activations are passed to the hidden layer, and hidden unit activations are passed through weighted, trainable connections to the output units. Output error is computed for each output unit as the difference between a desired output and the actual output. Hidden-to-output weights are changed according to how much of the error was contributed by each weighted connection. Error is propagated back to the hidden layer by assigning each hidden unit a proportion of responsibility for the output error, and changing the incoming weights from the input and context layers accordingly.

For the current simulation using Norris' word list, we proceeded as follows. The network consisted of 18 input units (one for each phonetic feature), 20 hidden and context units, and 48 outputs (one for each lexical item, using a localist representation). Bias units were used for both the hidden and output units, and bias activation was always set to 1. The network was trained for many *epochs*, with a learning rate of .05. At each epoch, the list of 48 items was randomly ordered. Then each item was presented phoneme-by-phoneme. The network's task at each time step was to indicate the lexical item that was being presented by activating that word's localist output unit, and setting all other lexical units to zero. Context activations were *not* reset to 0 between words, as is sometimes done with SRNs. Resetting the context weights would effectively make the SRN an alignment model, since an explicit cue to word boundaries would be given.

As in Norris' simulation, we found little co-activation at offset between the reversed cohort pairs which overlapped only in one or two final phonemes. (Note

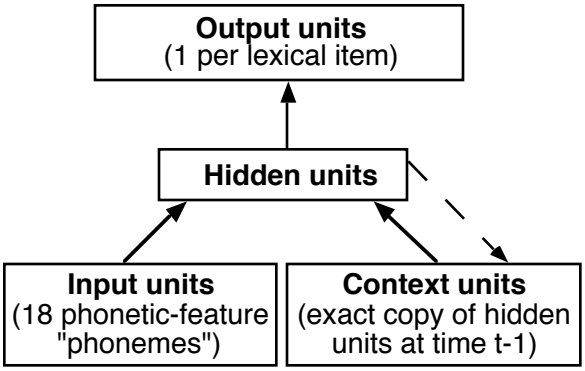

Figure 2: Schematic of SRNs used. Solid lines indicate fully connected, trainable, weighted connections. The dotted line indicates an exact 1-to-1 copy from each hidden unit to a corresponding context unit.

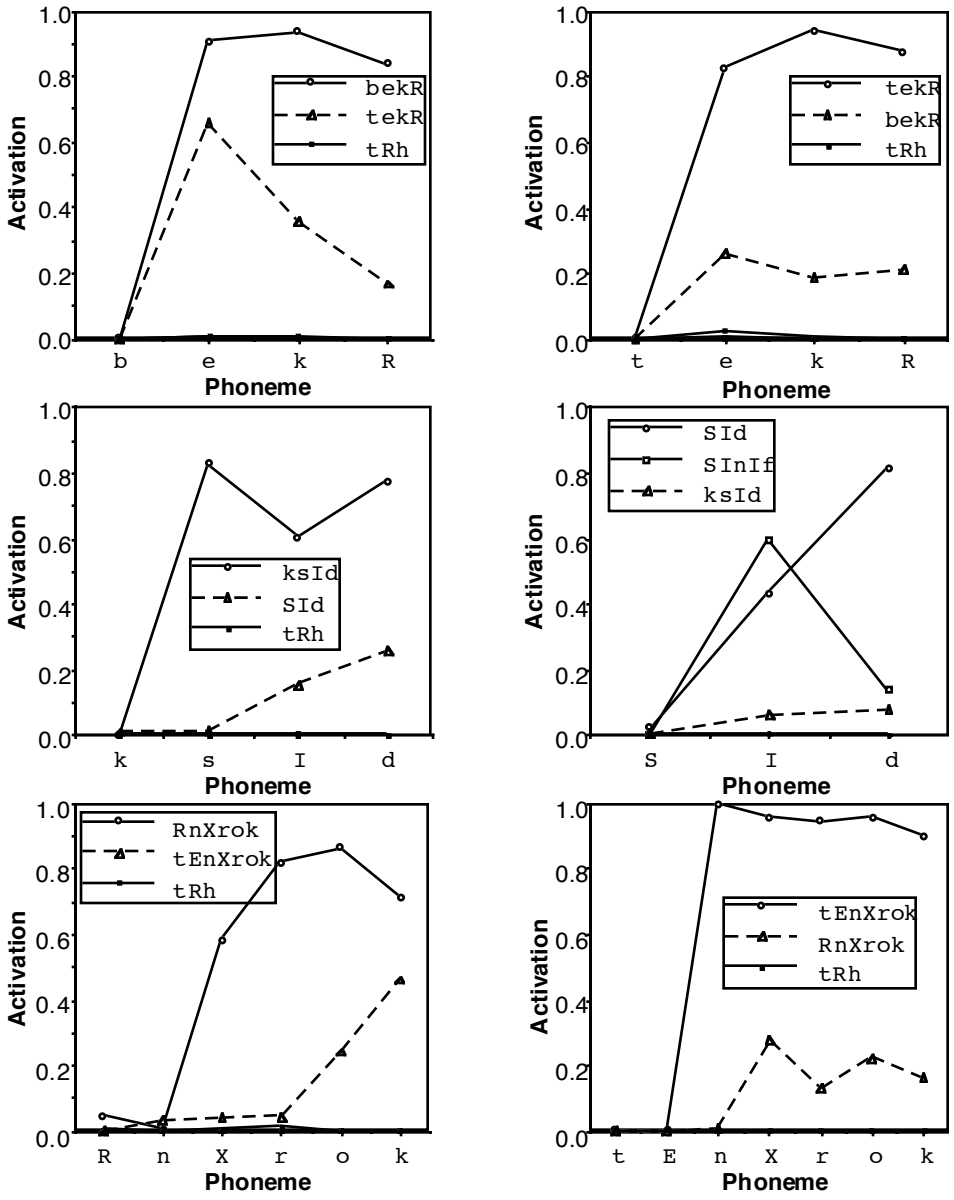

Figure 3: Rhyme effects using a variant of Norris' (1990) lexicon. The three most active items are shown in each case. The top two panels show the effects for *baker* and *taker*. The middle panels shows the effects for *ksid* and *hsid*. Note the cohort effect for /SId/. The bottom panels show the effects for *ronoroc* and *tenoroc*.

that for now, we will talk about co-activation and not competition; later, we will discuss whether activations in these simulations indicate competition.) However, other models, such as TRACE, would not predict much competition between these items, either, because they overlap so little.

If we consider items with more complete rhymes, the results are quite different. Of the 48 items, there were 7 rhyme pairs (given in their orthographic forms here): *baker*/*taker*, *renoroc*/*tenoroc* (*coroner*/*coronet* reversed), *reviled/timiled* (*delimit* reversed), *dish*/*finish*, *hsid*/*ksid* (*dish*/*disk* reversed), *hsinif*/*raef* (*finish*/*fear* reversed), and *flash*/*trash*. We examined the performance of the network after 10,000 epochs. By this point, the most activated word unit was always the correct item by the last phoneme.

Strong rhyme co-activation was observed for three of the pairs after 10,000 epochs of training (*baker*/*taker*, *renoroc*/*tenoroc,* and *hsid*/*ksid*), and weak activation was observed for *trash*/*flash*, *dish*/*finish*. The two pairs which did not show even weak co-activation overlapped only slightly in the last syllable, and so the lack of activation is not surprising. Also, there were co-activation effects for these items earlier in training, with rhymes more active than unrelated items. However, prior to the 10,000 epoch mark, not all items were being "correctly identified" by the last phoneme. We will return to this in the discussion section. The results for the three strong rhyme pairs after 10,000 epochs of training are presented in Figure 3.

There is an asymmetry in each of the pairs in Figure 3. This is correlated with the density for the initial two segments in the items. For example, while there were more words beginning with /t/ than with /b/ (8 vs. 5) there were 7 initial CV or CC sequences beginning with /t/, as compared to 4 for /b/. Only one /t/-word overlapped by more than one onset phoneme with another word. The transitional probability for /te/ was .125, and .2 for /be/. It appears that with this lexicon, the network has learned to minimize error by "reserving judgment" for some initial sequences. There is less error associated with no response than with activating all members of a large onset cohort when the cohort can be narrowed significantly by waiting for one more segment. For the other pairs, the asymmetry is due to the misalignment of the rhymes; in each case, the item with an initial disadvantage has one more segment preceding the overlapping offset. That we still find rhyme effects given this misalignment bodes well for handling the rhyme effects Allopenna et al. (1998) reported. Not only do we find rhyme activation for items differing by more than one feature at onset, we find it for mis-aligned rhymes similar to those among the Allopenna et al. stimuli (e.g., *beaker*/*speaker*).

Why did we find substantial rhyme co-activation when Norris reported no co-activation due to offset overlap? Some of the examples he presented had rhymes, but in his simulation, rhymes were never active. The differences between our simulations might have been due to learning rate, our phonological representations (ours were richer than those used by Norris), or the *amount* of training. The most likely explanation is that the amount of learning Norris allowed before examining the SRN's performance led to the elimination of rhyme effects. We replicated our simulation, but accelerated learning with a learning rate of .5. After 200 epochs,

the SRN's output was similar to that of our original simulation, albeit somewhat noisier. However, by 1000 epochs, rhyme effects disappeared, presumably because the model learned to give more weight to context activations for rhymes, and cohort co-activations nearly mirrored transitional probabilities. This means the SRN had learned the lexicon nearly perfectly, which we will argue later provides a poor analog to the human language processor.

Simulation 2: Allopenna et al. (1998)

Simulation 1 demonstrated that SRNs do predict rhyme activation under certain circumstances. We now turn to the question of how well those predictions match human data, specifically the data from Allopenna et al. (1998) shown in Figure 1.

We used an SRN similar to the one described for the previous simulation, except that it had 23 localist outputs (one for each possible response), 40 hidden and context units, and we used a learning rate of .1.[2] The items we used were phonemic transcriptions of the words *beaker*, *beetle*, *speaker*, *carrot*, *carriage*, *parrot*, *candle*, *candy*, *handle*, *pickle*, *picture*, *nickel*, *casket*, *castle*, *basket*, *paddle*, *padlock*, *saddle*, *dollar*, *dolphin*, *collar*, *sandal* and *sandwich*. The training procedure was identical to that for the previous simulation. For each epoch of training, the words were randomly ordered, and then presented phoneme-by-phoneme (using the 18-feature vector representation) to the SRN. The desired output was the current word, and context unit activations were not reset between words.

We chose to examine the model after 1500 epochs of training, because at that point, the correct output node was always the most highly active by the last phoneme of each word, but rhyme and cohort effects were still present. In order to compare the model's performance to the data in Figure 1, we chose all of the target-cohort-rhyme sets in which the target was four phonemes in length (five of the eight sets, with the targets *beaker*, *dollar*, *pickle*, *paddle*, and *sandal*). Nearly identical effects were found for the other targets (*carrot* and *candle*, of length 5, and *casket*, of length 6), but we restricted our analyses to 4-phoneme targets because it is not clear how responses to phonemes of different lengths should be combined. We averaged cohort and rhyme conditions for each of the 4-phoneme targets. The average output is shown in the top panel of Figure 4. Target, cohort, and rhyme activations represent the averages across all 4-phoneme sets. The unrelated activation is the *maximum* value found at each phoneme from any set.

[2] Note that the a wide range of parameters (number of hidden and context units, learning rate, and training epochs) lead to the same result (as evidenced by our successful replication of Simulation 1 with a larger learning rate and smaller number of trianing epochs). For training sets like the one used for this simulation, increasing the number of hidden units allows a smaller learning rate to arrive at the desired performance threshold ("recognition" of targets, i.e., target units having the highest activation at word offset).

A weakness of the current input representations is that entire phonemes are presented in a single time-step. An input representation which allowed more continuous input presentation would clearly provide a better comparison to the human data. In order to compare the current simulation output to the data, we used linear interpolation and extrapolation to fit the simulation output to the data.

There were 30 frames of human data (each frame corresponding to a 33 msec video frame). In order to stretch our four simulation data points, we aligned point 1 to the fifth frame of the human data, which was the frame before any of the fixation probabilities were greater than .01. Then, 5 frames were inserted between each simulation data point. This took us intentionally to frame 20. At the last simulation data point, the rhyme activation has decreased from its peak value. Frame 20 corresponds to a similar point in the human data. From frame 21 to frame 30, we assumed the target probability should rise to 1, and the other values should decrease to 0, as is true for the human data. We then computed correlations between the interpolated simulation predictions and the human data. The r^2 values for the targets, cohorts and rhymes were .87, .93 and .81, respectively. The interpolated simulation output is presented in the lower panel of Figure 4.

This simulation demonstrates that not only can SRNs predict rhyme effects, the predictions map quite well onto human data. At the same time, we do not wish to present this as an adequate model of spoken word recognition. Better input representations are needed, and the lexicon for this simulation was rather odd, given that every item had a cohort and/or rhyme. A replication with a more naturalistic lexicon is needed, and we are currently working on assembling such a lexicon. However, the results are still quite promising, and we need not alter SRNs

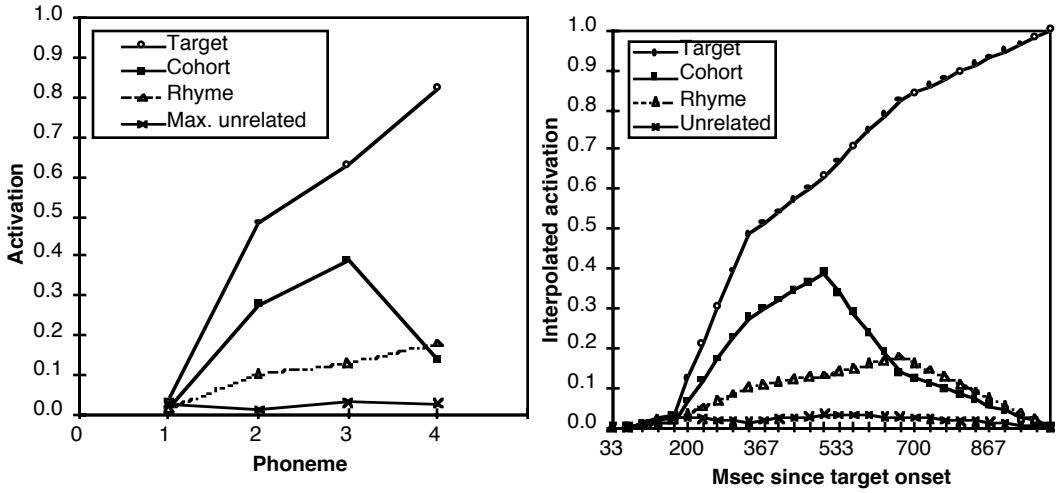

Figure 4: Simulations of the cohort and rhyme effects from Allopenna et al. (1998). Top: activations. Bottom: activations interpolated for comparison with the data.

in order to provide a coarse account of rhyme effects.

Simulation 3: Magnuson et al. (1998, 1999)

Magnuson et al. (1998, 1999) replicated and extended the Allopenna et al. (1998) results. In order to have precise control over distributional characteristics of the input, we constructed an artificial lexicon with very specific properties. The lexicon consisted of 16 words. The novel words were randomly mapped onto distinct, novel geometric objects. Subjects learned the names of the objects over two days of training (either two or four objects would be displayed, and the subject would hear an instruction such as, "click on the pibu"; after the subject

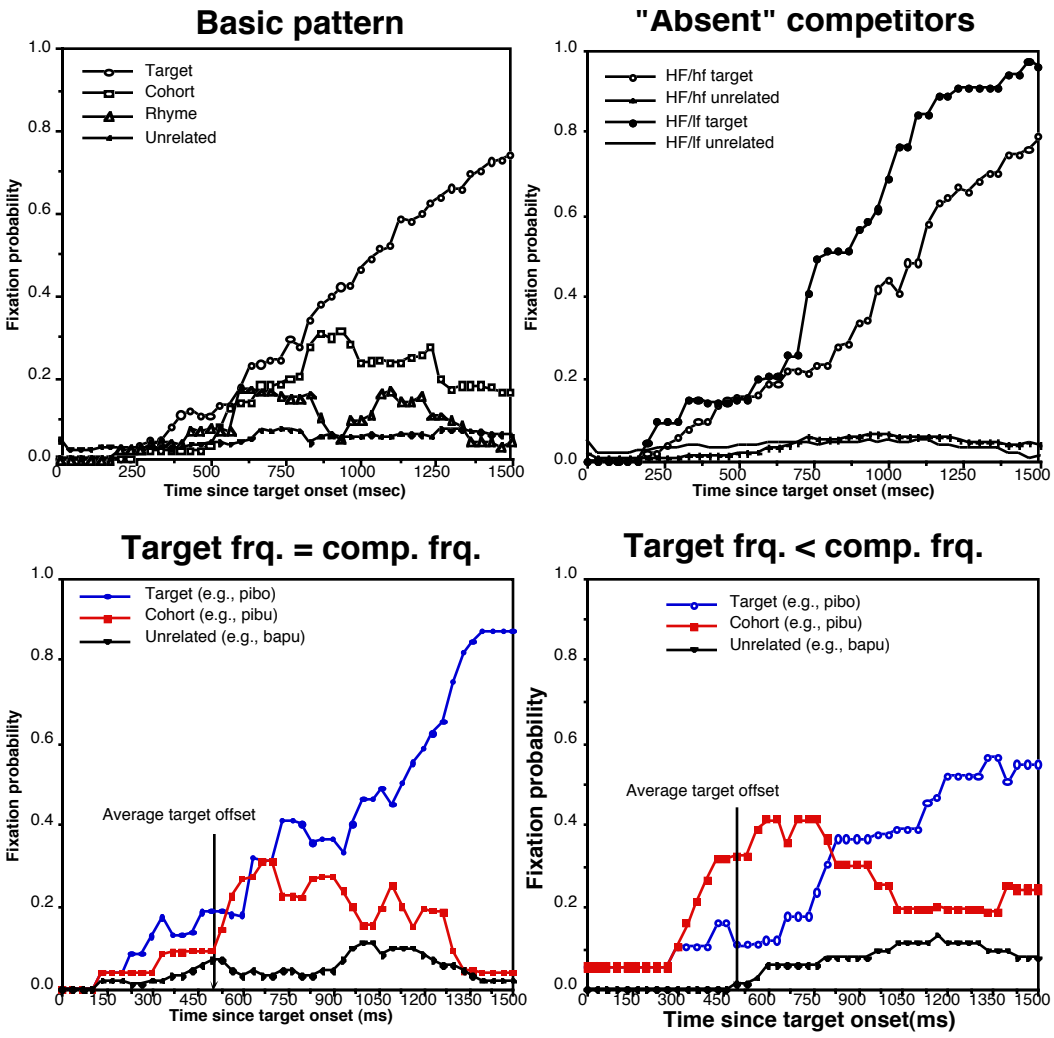

Figure 5: Major trends from artificial lexicon studies (Magnuson et al., 1998, 1999).

clicked on one object, feedback was given by removing all the incorrect choices from the display and repeating the target name).

The lexicon could be divided into four sets of four words. For example, one set was /pibu/, /pibo/, /dibu/, and /dibo/. Each item has one cohort (/pibu/ and /pibo/, /dibu/ and /dibo/) and one rhyme (/pibu/ and /dibu/, /pibo/ and /dibo/). The real advantage of these subsets was the frequency manipulation they allowed. For example, if /pibu/ and /dibo/ (which were not predicted to compete significantly) were presented with high frequency in the learning phase, and /pibo/ and /dibu/ were low-frequency, we would have two different frequency conditions: high-frequency items with low-frequency competitors, and vice-versa. In Magnuson et al. (1998), items were either high- or low-frequency, such that there were four target/competitor frequency conditions: high/high, low/low, high/low and low/high. In Magnuson et al. (1999), a third, "medium" level of frequency was added, which allowed a crucial test. On some trials, high-frequency targets which had either high- or low-frequency competitors were presented among three unrelated, medium-frequency distractors. If competition effects in the paradigm were driven only by the characteristics of items displayed on a given trial, there should have been no difference in the fixation probabilities to these items, since their high- or low-frequency competitors were absent from the display. However, we found robust differences; fixation probabilities rose much more slowly for high-frequency targets with high-frequency distractors than for high-frequency targets with low-frequency distractors.

The major trends from these experiments are shown in Figure 5. In the top left panel, target, cohort, rhyme, and unrelated probabilities averaged over all conditions are shown. In the top right panel, target and average unrelated distractor probabilities are shown from the "absent competitor" condition just described. The lower panels show cohort effects when (bottom left) target and competitor frequencies are equal and (bottom right) target frequency is less than competitor frequency. These results provided the first fine-grained measures of the time course of activation and competition among items as functions of phonological similarity and experience (see Dahan, Magnuson and Tanenhaus, submitted, for similar results using real words).

To verify that an SRN would capture the major trends of these studies, we conducted a simulation with the same 18-feature input units from Simulations 1 and 2, 20 hidden and context units, 16-localist lexical outputs, and a learning rate of .01. High-frequency items were presented twice as often as low-frequency items. A small high/low ratio, a small learning rate, and many training epochs (528,000) were required to fit the data (although again, a range of parameters

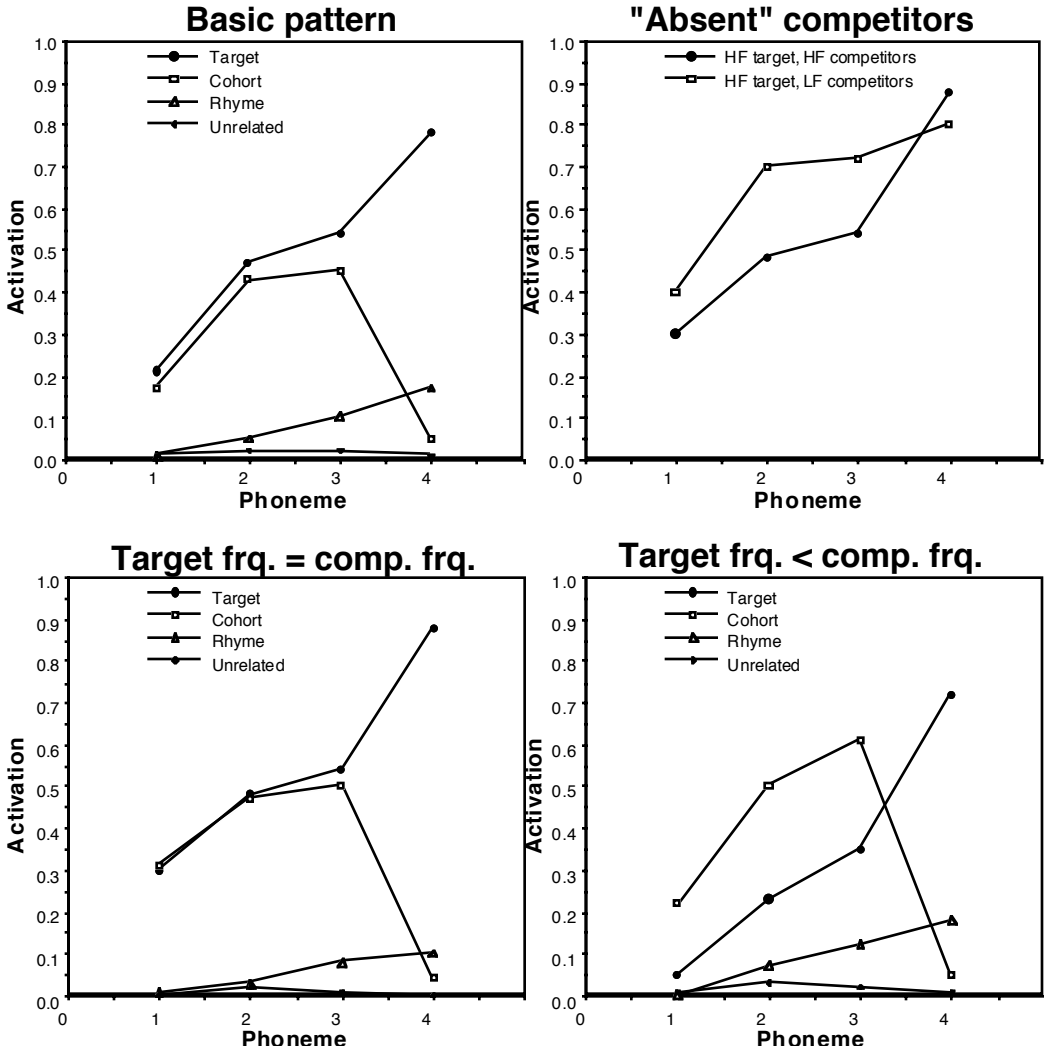

Figure 6: Simulations of the major trends from the artificial lexicon studies.

would provide good fits). The simulation results are shown in Figure 6. For all four panels, activations are based on simulations using the entire lexicon. Rather than using a variant of the Luce choice rule, as Allopenna et al. (1998) did, to capture the constraints of the subjects' task, we present these results as evidence that the SRN provides a basis for the major trends of the artificial lexicon studies.

Discussion

The simulations described here show that rhyme effects are predicted by SRNs, but that the amount of co-activation observed depends on training parameters and the similarity density at each segment. Among training parameters, the most

important one would appear simply to be the amount of training; we replicated Simulation 1 even after increasing the learning rate by an order of magnitude. Whether or not we observed rhyme effects depended on when training was stopped. If an SRN is trained until virtually no changes occur in connection weights, and if it has a sufficient number of hidden and context units to represent the temporal dependencies of the input, its outputs will mirror the statistics of the lexicon perfectly. This does not provide a good analog to the human language processor.

There are obviously many differences between the learning situations of our SRNs and a human learner. One is that our SRN always received an input of perfect fidelity (with the exception of context activations at word onsets, which will contain arbitrary information about the ending of the preceding, randomly selected word), whereas the human must adapt to differences in acoustics, talker, dialect, and rate, among others. One coarse way to approximate this is to end the learning phase for the SRN before it has learned its input perfectly, i.e., while its *representations* are still noisy.

A simple way to improve our inputs might be to add noise to them, or to simulate coarticulation by allowing adjacent phonemic representations to blend with one another. A better alternative would be to use more realistic input. The phonetic features used here could be replaced by pseudo-acoustic representations, like those used for TRACE, or those developed by Plaut and Kello (1999), although noise to simulate natural variability would still be called for.

However, the current results do show that cohort, rhyme, frequency, and neighborhood density effects can be expected from generic SRNs, which paves the way to further explorations of SRNs as models of spoken word recognition. One issue that must be explored is the role of competition at the output level. McQueen and his colleagues (e.g., McQueen, Cutler, Briscoe & Norris, 1995), among others, have argued that competition is required to model some phenomena in spoken word recognition (e.g., the temporal dynamics of processing embedded words). It is possible that an explicit competition mechanism would not be required for a trained SRN to model these phenomena.

Although SRNs are often described as purely bottom up processing models (e.g., Cairns, Shillcock, Chater & Levy, 1997), there are two respects in which they are top-down. The first is that hidden unit activations are passed recurrently through the network via the context units. The context unit activations are not a bottom-up source of information, in that they are directly related to the state and output of the network at the previous time step. Second, learning in an SRN is driven in a top-down fashion. Weights are updated based on an error signal propagated back from an explicit comparison of the output to a desired output. Two co-active output units effectively reflect competition among subsets of the weighted connections for the reward of weight increases. It is possible that this

competition during learning may produce a processing system which is functionally equivalent to one feeding into an explicit competition mechanism. While reports of SRNs used to examine, e.g., embedded word effects (Davis, Gaskell, & Marslen-Wilson, 1997) do not support this claim, this may be due to specific training parameters and lexical characteristics used. We are currently performing lexical analyses of English with the goal of assembling a reasonably small lexicon with a good approximation of the lexical statistics of English. Such a "reasonably" small approximation to a real lexicon will allow simulations to be carried out realistically quickly, with a minimum of effects due to unexpected or unrealistic lexical characteristics.

Acknowledgments

Supported by NIH HD27206 to MKT, NSF SBR-9729095 to MKT and RNA, and an NSF GRF to JSM.

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.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33, 111-153.

Connine, C.M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193-210.

Davis, M. H., Gaskell, M. G., & Marslen-Wilson, W. (1997). Recognising embedded words in connected speech: Context and competition. In J. Bullinaria & G. Houghton (Eds.), *Proceedings of the 4th Neural Computation in Psychology Workshop.*

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

Magnuson, J. S., Dahan, D., Allopenna, P. D., Tanenhaus, M. K., & Aslin, R. N. (1998). Using an artificial lexicon and eye movements to examine the development and microstructure of lexcical dynamics. In Gernsbacher, M. A., and Derry, S. J. (Eds.), Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, 651-656.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (1999). Spoken word recognition in the visual world paradigm reflects the structure of the entire

lexicon. Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society, 331-336.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.

Matin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53, 372-380.

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

McQueen, J. M., Cutler, A., Briscoe, T. & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, 10, 309-331.

Norris, D. (1990). A dynamic-net model of human speech recognition. In G.T.M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Persepectives*, 87-104. Cambridge: MIT.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.

O'Grady, W., Dobrovolsky, M., & Aronoff, M. (1989). *Contemporary Linguistics*. New York: St. Martin's.

Plaut, D. C., and Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist model. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 381-415). Mahwah, NJ: Erlbaum.

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

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Eye Movements and Their Role in Visual and Cognitive Processes. Reviews of Oculomotor Research V4*. Amsterdam: Elsevier.