THE OPTIMAL ARCHITECTURE FOR SIMULATING SPOKEN-WORD RECOGNITION

DENNIS NORRIS

MRC Cognition and Brain Sciences Unit, Cambridge, U.K.

ANNE CUTLER AND JAMES MCQUEEN

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands.

Simulations explored the inability of the TRACE model of spoken-word recognition to model the effects on human listening of subcategorical mismatch in word forms. The source of TRACE's failure lay not in interactive connectivity, not in the presence of inter-word competition, and not in the use of phonemic representations, but in the need for continuously optimised interpretation of the input. When an analogue of TRACE was allowed to cycle to asymptote on every slice of input, an acceptable simulation of the subcategorical mismatch data was achieved. Even then, however, the simulation was not as close as that produced by the Merge model, which has inter-word competition, phonemic representations and continuous optimisation (but no interactive connectivity).

A major distinction among models of the recognition of spoken words is whether or not a model allows feedback from logically later to logically earlier levels of processing. One of the leading current models is TRACE [1]. TRACE is an interactive model which allows feedback from words to phonemic representations. Other models (e.g. Shortlist; [4]) do not allow such feedback.

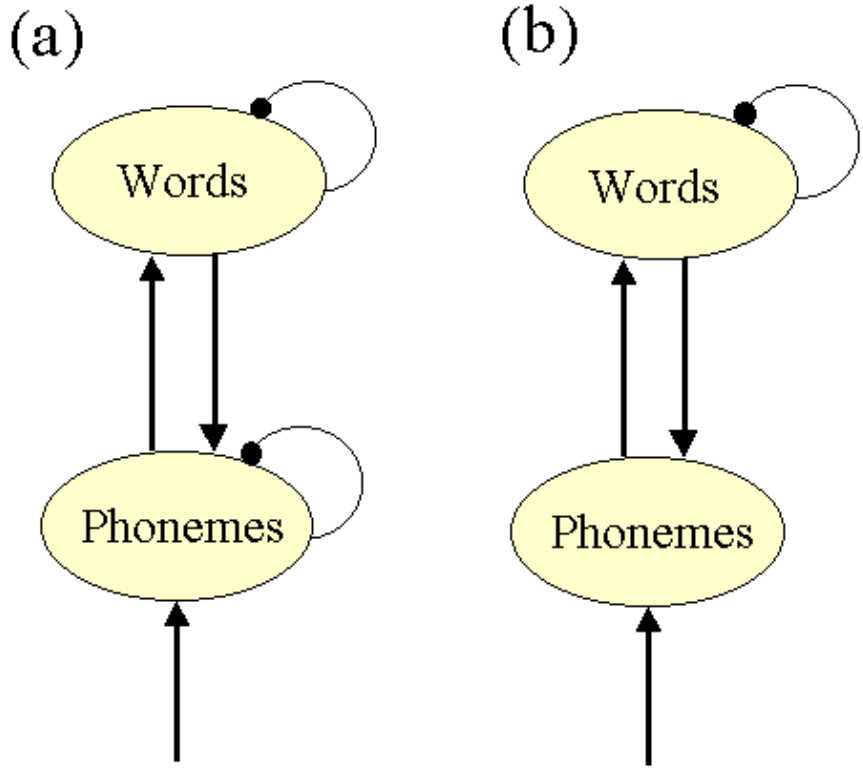

Figure 1. Connections between phoneme and word levels in TRACE (a) and Shortlist (b). Only TRACE has feedback from words to phonemes.

Recent research in spoken-word recognition ([1]; [3]) has posed a serious challenge to TRACE. In this research, listeners made phonetic decisions or lexical decisions on words and nonwords, some cross-spliced so that they contained acoustic-phonetic mismatches. For instance, the word job could have its initial portion jo- taken from the word jog or the nonword jod; the nonword smob could have its initial portion smo- taken from the word smog or the nonword smod. Figure 2 shows the relevant lexical decision data: "YES" responses to words cross-spliced with words and to words cross-spliced with nonwords were not significantly different, but "NO" responses were slower to nonwords cross-spliced with words than to nonwords cross-spliced with nonwords. (The same asymmetry appeared in the phoneme decision data.)

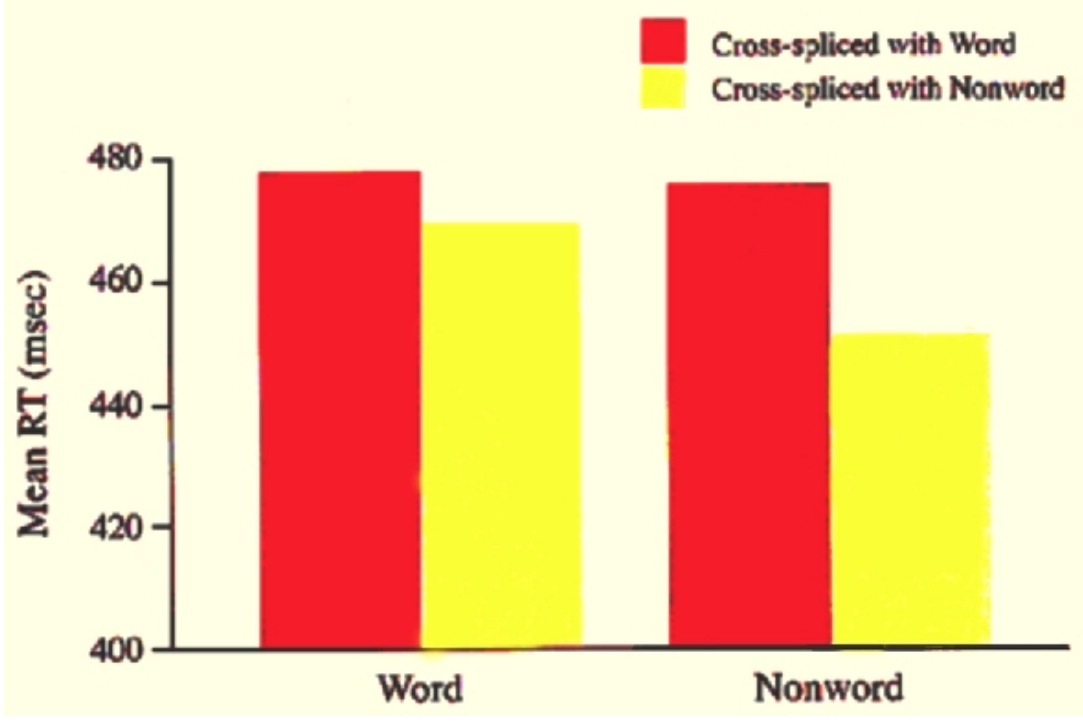

Figure 2. Mean "YES" reaction time (RT) to words and "NO" reaction time to nonwords in lexical decision task; data from [3].

The standard version of TRACE cannot simulate these findings. Marslen-Wilson and Warren [1] used this failure as a reason to reject the TRACE model, and they attributed the problems in particular to the inter-word competition process in the model and to TRACE's use of phonemic representations. However, these cannot be the reasons why TRACE failed to simulate the data. The Merge model, a model of phonemic decision-making which is integrated with Shortlist, can simulate the same data ([5]; [3]). Merge is an autonomous model, i.e. it does not allow the flow of information from the lexical to the phonemic level as shown in TRACE in Figure 1. In Merge, phonemic decisions are made by a dedicated decision-making process which accepts information from both phonetic processing and lexical activation. The model is implemented as a simple competitive network model, and several simulations with the model are reported by [5].

1 COMPARATIVE SIMULATIONS

Figure 3 shows a Merge simulation of the lexical decision data depicted in Figure 2. Activation levels across time can be compared for the relevant lexical nodes associated with the four different types of mismatching cross-spliced item. That is, for word items, the activation level is shown for the word for which a "YES" response is made. It can be seen that the two different types of mismatching word in

fact reach the same level at asymptote, which is consistent with the result of the experiment (Figure 2), in which there was no significant difference in these "YES" responses. For the nonword items, activation is plotted for the lexical item which provided the word-onset which was cross-spliced - i.e., smog in the example given above. In the case of nonwords cross-spliced with other nonwords (e.g. smob in which smo- came from smod), there is no competing lexical activation, but in the case of nonwords cross-spliced with words (e.g. smob in which smo- came from smog), there is significant activation, again consistent with the result from the experiment (Figure 2), in which "NO" responses to nonwords cross-spliced with words were delayed relative to the other mismatched nonwords. The activation does not rise as high as that for the real words, however, consistent with the fact that the listeners did not erroneously classify these nonwords as real words. Greater detail of this simulation can be found in [5].

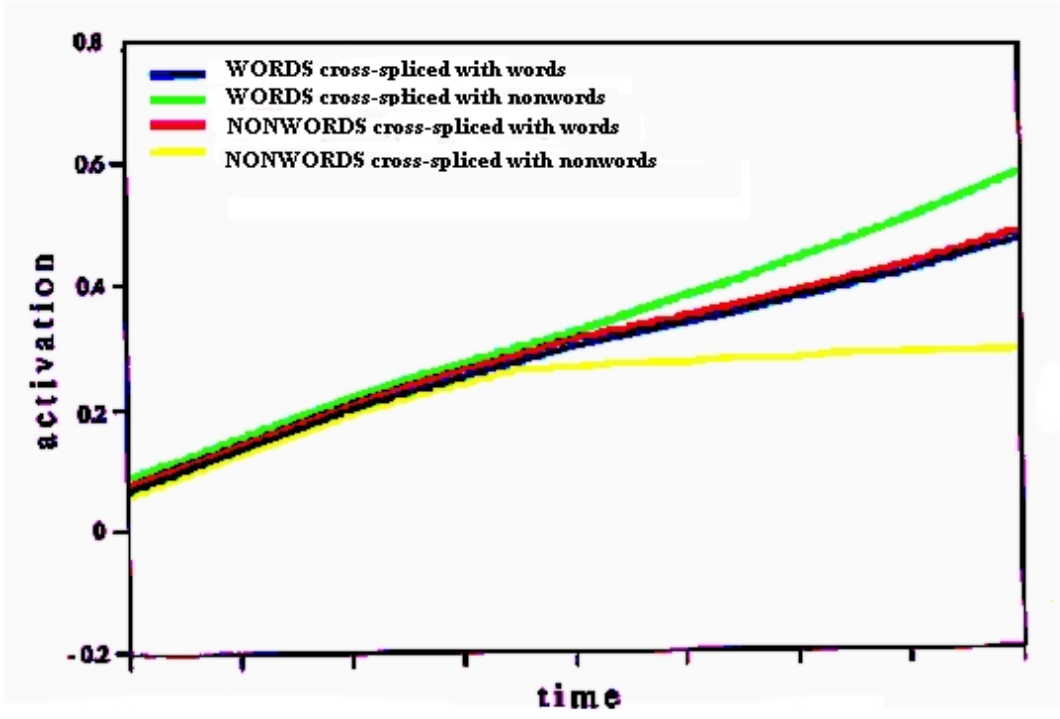

Figure 3. Activation levels in Merge simulations of the lexical decision data. Activation is shown for the relevant lexical nodes for each condition. Both types of word (green, blue lines) reach the same level at asymptote. In the case of nonwords cross-spliced with words (red line), a competing word is activated, making "NO" responses slow.

So why can Merge simulate the data while TRACE cannot? Merge and Shortlist differ from TRACE in a number of ways. (1) TRACE incorporates feedback between processing levels, Merge and Shortlist do not. (2) In Merge/Shortlist, the inter-word competition process produces a continuously optimal lexical parse of the input, but this is not the case in TRACE.

We explored the reasons why TRACE could not simulate the crucial data. Using a small-scale version of TRACE, we systematically altered features of the model and compared the performance of each version with the successful Merge simulation. This small-scale TRACE analogue was, essentially, a modified version of the Merge model. We eliminated Merge's decision nodes, and we added feedback from the word to the phoneme layer and within-level inhibition at the phoneme layer.
By these means we transformed a Merge network into an interactive model with the same connectivity pattern as TRACE. This enabled us to investigate whether some simple manipulation might enable this TRACE analogue to simulate the subcategorical mismatch data after all. Clearly there must be some reason why [1] simulation with TRACE had not worked; equally clearly, given Merge's successful simulation of the same data, the reason could not be any of the ones proposed by Marslen-Wilson and Warren.

The simplest kind of interactive model we could try was one with dynamics like the original TRACE model: one network cycle per time slice and no phoneme-word inhibition and no resetting of activations. For these simulations, as described, there was word-to-phoneme feedback and within-level inhibition at the phoneme level. In TRACE the between-phoneme inhibition is required in order to reach an unambiguous decision as to which phoneme is in the input. (For exactly the same reason, there is between-unit inhibition in the decision units in Merge.) Because there was between-phoneme inhibition in this TRACE analogue, the phoneme level cycled in synchrony with the word level.

Figure 4 shows the result of a simulation which yielded the same poor fit as that obtained by Marslen-Wilson and Warren with the full TRACE model. The model incorrectly predicted a difference between the two types of mismatching cross-spliced word: it can be seen that the lines representing the two word types do not reach the same level of activation. The model also exaggerated the difference between the two types of mismatching cross-spliced nornword, to the extent that nonwords cross-spliced with words produced as much lexical activation as words cross-spliced with words. This would predict the same pattern of response for both item types, i.e. a very high error rate for the nonwords cross-spliced with words (i.e. many incorrect "YES" responses to smob when the smo- came from smog). There was no such effect in the experimental data, as pointed out above; but both Marslen-Wilson and Warren's simulations, and the present ones with this small-scale TRACE analogue, produced this unwanted effect.

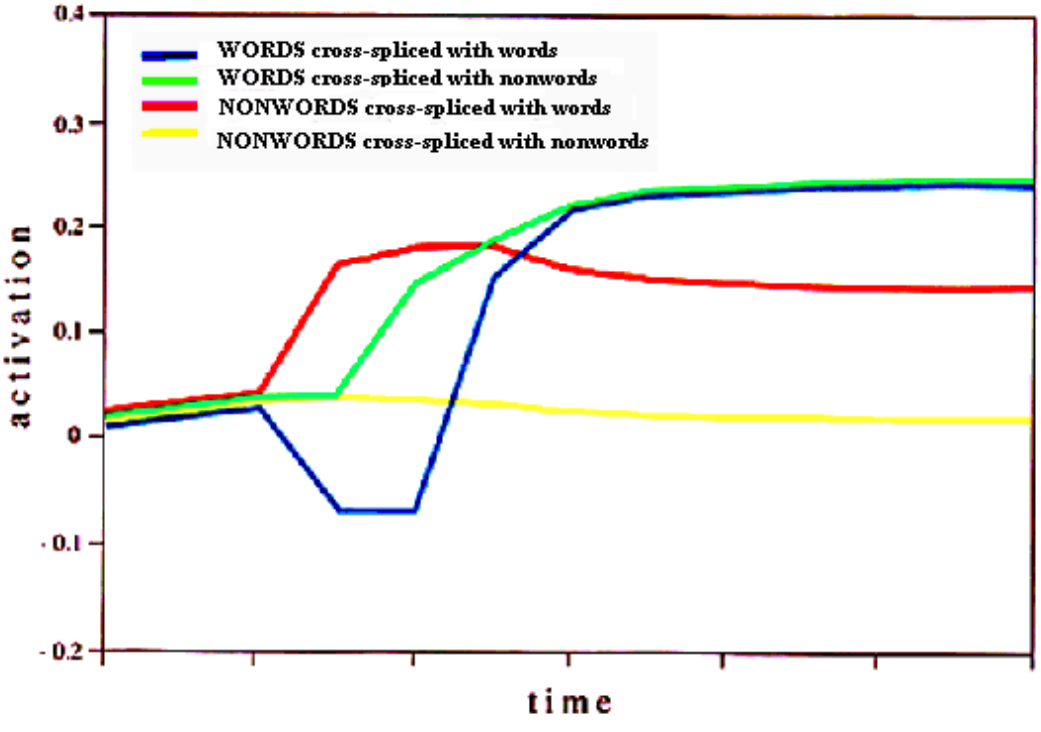

Figure 4. Activation levels in TRACE simulations of the lexical decision data. Activation is shown for the relevant lexical nodes for each condition. The two types of word (green, blue lines) do not reach the same level of activation. Nonwords cross-spliced with words (red line) produce as much activation as words cross-spliced with words (blue line).

We made extensive attempts to improve this model's fit to the data by setting parameters by hand, but these attempts were completely unsuccessful. In fact, it was hard to be sure exactly why it was proving so difficult to discover a suitable set of parameters. Parameter setting in an interactive model is, by the nature of the model, very difficult to do, since adjustments to the phoneme parameters alter lexical behavior and vice versa. We therefore decided to investigate use of an optimization procedure to set the parameters of the interactive model automatically in order to reproduce the same activation pattern as the autonomous model, Merge (see Figure 3). To achieve this we used Powell's conjugate gradient descent method ([7]) to fit the parameters of the interactive model.

In these optimizations, the phoneme activations of the TRACE analogue were targeted to reproduce the decision unit activations of Merge. Correlations of lexical and phonemic activations were computed independently and we attempted to maximize the sum of those two correlations. Note that even this was unsuccessful when we used the best Merge model (see Figure 3) as our target. Better results were obtained using another set of Merge parameters which was not optimal but which nevertheless still allowed Merge to give a plausible account of the data. The TRACE simulation was required to produce the same pattern of activation as Merge did regardless of any differences in absolute activation levels. These optimization procedures did improve the model's fit to the lexical decision data, but it was quite impossible to find a single set of parameters which enabled the model to fit both the lexical decision and the phoneme decision data.

We next constructed a range of networks, each successively more similar to Merge, but all with TRACE's architecture (single-outlet, i.e. phoneme decisions only from the phoneme nodes, and feedback from lexical to phonetic processing). We increased the number of cycles per slice and added resetting activation after each time slice. In brief, we found that the more closely the model resembled Merge, the better it was able to simulate the data. Eventually we found a version of the model that produced an acceptable simulation. However, the model was very unstable and even the very best version never produced as close a simulation of the data as Merge did.

The subcategorical mismatch lexical decision simulation of our best TRACE analogue is shown in Figure 5. In this simulation, the model used 15 cycles per slice, a momentum term at both the word and phoneme levels, reset and no bottom-up inhibition. For the momentum term, some proportion of the final activation level at the end of the previous time slice was added to the node's input at each cycle. The parameters used in this simulation are given in an appendix to [5]. This model clearly provides a reasonable fit to the lexical decision data, quite comparable with the Merge simulation results plotted in Figure 3 above. However, it should be noted that again this TRACE analogue could not give as good a representation of the phonemic decision data as the Merge simulation did, since the TRACE analogue showed very large word-nonword differences, differences which were not reliable in the human data.

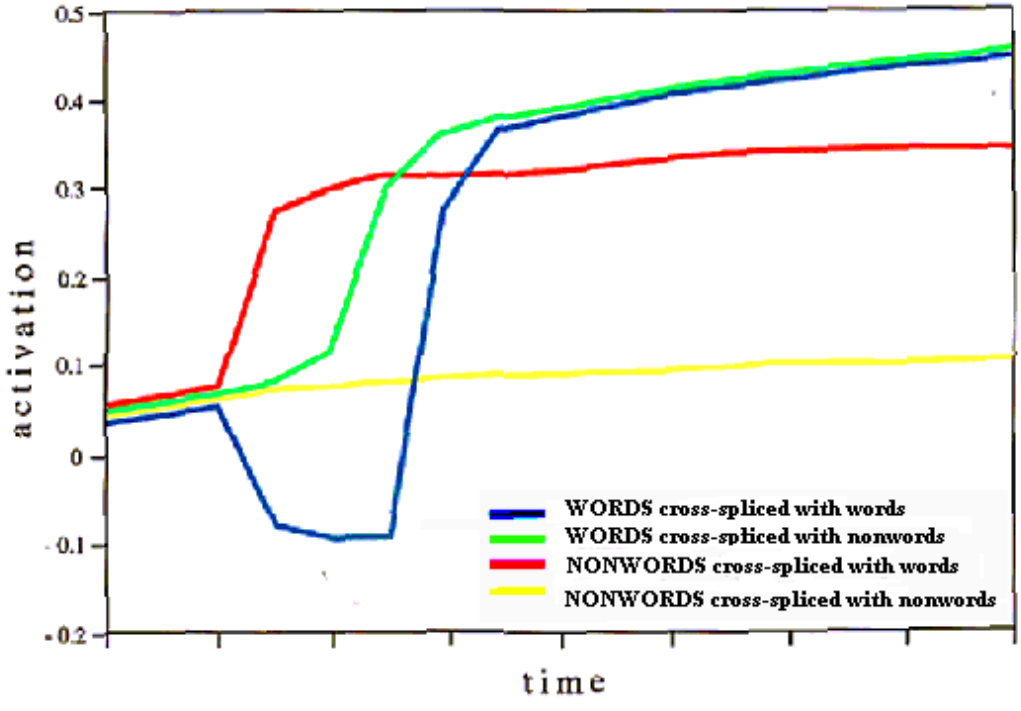

Figure 5. Activation levels in simulations of the lexical decision data with the best adapted TRACE-like model. Activation is shown for the relevant lexical nodes for each condition. The results are very similar to Figure 3.

2 CONCLUSION

The best-performing small-scale TRACE-like model differed from Merge principally by having feedback. Thus since this TRACE-like model, like Merge, can simulate the data, the presence of feedback could not be the reason for the failure of the simulation by the standard version of TRACE. Instead, the crucial feature which the TRACE adaptation required was the addition of the Shortlist continuous optimisation procedure, involving 15 cycles of interactive activation per time slice. That is, the primary reason why TRACE is unable to account for the subcategorical mismatch findings is that it does not allow lexical level processes to cycle to asymptote on a small enough time scale. The model therefore incorrectly predicts competition effects in the words cross-spliced with words, and it is probable that this also causes the model to overestimate the inhibitory effect in the nonwords cross-spliced with words.

Of course, there is no reason to prefer the model with feedback given that the model without feedback performs rather better, and given also that there are other data which crucially challenge the feedback assumption ([6];[5]).

NOTE

An expanded version of this report can be found as:
Cutler, A., Norris, D.G. & McQueen, J.M. Tracking TRACE's troubles. Proceedings of the Workshop on Spoken Word Access Processes, Nijmegen, May 2000.

REFERENCES

1. Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. Psychological Review, 101, 653-675.
2. McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.
3. McQueen, J.M., Norris, D.G. & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. Journal of Experimental Psychology: Human Perception and Performance, 25, 1363-1389.
4. Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52, 189-234.
5. Norris, D.G., McQueen, J.M. & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23.
6. Pitt, M.A. & McQueen, J.M. (1998) Is compensation for coarticulation mediated by the lexicon? Journal of Memory and Language, 39, 347-370.
7. Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1986). Numerical Recipes: The Art of Scientific Computing. Cambridge: Cambridge University Press.