

# Phonological Similarity Effects Without a Phonological Store: An Individual Differences Model

**C. Philip Beaman (c.p.beaman@reading.ac.uk)**

Department of Psychology, University of Reading  
Earley Gate, Whiteknights, Reading RG6 6AL UK

**Ian Neath (ineath@mun.ca)**

Department of Psychology, Memorial University of Newfoundland  
St John's, NL, Canada A1B 3X9

**Aimée M. Surprenant (asurprenant@mun.ca)**

Department of Psychology, Memorial University of Newfoundland  
St John's, NL, Canada A1B 3X9

## Abstract

The feature model of immediate memory (Nairne, 1990) is applied to an experiment testing individual differences in phonological confusions amongst a group (N=100) of participants performing a verbal memory test. By simulating the performance of an equivalent number of “pseudo-participants” the model fits both the mean performance and the variability within the group. Experimental data show that high-performing individuals are significantly more likely to demonstrate phonological confusions than low performance individuals and this is also true of the model, despite the model's lack of either an explicit phonological store or a performance-linked strategy shift away from phonological storage. It is concluded that a dedicated phonological store is not necessary to explain the basic phonological confusion effect, and the reduction in such an effect can also be explained without requiring a change in encoding or rehearsal strategy or the deployment of a different storage buffer.

**Keywords:** Short-term memory; Cognitive modeling; Individual differences.

## Individual Differences in Immediate Memory

Research into immediate verbal memory is now sufficiently well-advanced that a number of phenomena are taken to typify the functioning of immediate memory. In particular, the *phonological similarity effect* (PSE), the simple empirical observation that verbal items, whether visually or auditorily presented, are more likely to become confused on immediate recall tests if they sound alike (Conrad, 1964), is taken as prima facie evidence for the existence and utilization of a short-term phonological store (Baddeley, 1986, 2003). By extension, the lack of a phonological similarity effect is frequently interpreted as the absence of a phonological store, for example in neuropsychological case studies of brain-damaged individuals (Della Sala & Logie, 1997; Vallar & Baddeley, 1984; Vallar & Shallice, 1990). In less severe cases the appearance of a reduced PSE is interpreted as the failure to make use of the store, for example in developmental

studies of young children's memory, (Hitch & Halliday, 1983; Hulme, Thomson, Muir & Lawrence, 1984), or in recall of supra-span lists which might be interpreted as requiring longer-term memory storage (Baddeley, 2000a) or implicating the existence of an alternative form of episodic short-term storage (Baddeley, 2000b). However, as noted by Logie and colleagues (Della Sala & Logie, 1997; Logie, Della Sala, Laiacona, Chalmers & Wynn, 1996) although the effect is robust at the group-level of analysis, it is routinely absent from the recall protocols of a minority of normally-functioning individuals. The current study aims to examine whether a particular model of immediate memory, the feature model of Nairne (1990) can account for variation in immediate memory performance and in the appearance of the PSE.

The feature model was chosen for this study for a number of reasons: Firstly, it is an extant model of immediate memory that has not previously been applied to individual differences data. In this it is not unique. To our knowledge, only the ACT-R list memory simulation (Anderson, Reader & Lebiere, 1996) has been applied to model individual data, although models have been applied to neuropsychological data (e.g., Brown, Della Sala, Foster & Vousden, in press). Secondly, the feature model, unlike other models eschews a dedicated phonological store and instead represents information in terms of abstract “features” including, but not restricted to, phonological features. Thirdly, the model has no strategic rehearsal mechanism. Any individual differences predicted by the model must therefore be a straightforward effect of mnemonic efficiency, rather than a by-product of the efficacy of one particular rehearsal or other maintenance strategy. Fourthly, the predictions of the model are based upon running multiple trials (typically 2-5000) and reporting the average performance as “the” prediction of the model. Like many models in which performance is simulated rather than calculated, the ontological status of individual trials or runs of the model has always been ambiguous. Here we suggest that individual simulation runs can be examined as if they were individual experimental trials. Averaging over a

series of such runs will provide data representing the performance of a “pseudo-participant” in a computational experiment (cf., Cooper, Yule & Fox, 2003; Hintzman, 1986).

### Background to the feature model.

The basic idea behind the feature model is that recall is guided by a set of “Primary Memory” (PM) cues of varying effectiveness in identifying the target item from a search set defined within “Secondary Memory” (SM), or memory proper. The major function of PM is to hold a variety of cues indicating which items were recently presented. Cues do not decay but are subject to a process of interference that renders the cues less accurate and comprises the basic mechanism for memory limitations. Formally, items in PM and SM comprise sets of internally-generated modality-independent (MI) and externally generated modality-dependent (MD) features organized as row vectors. For simulation purposes, feature values are randomly generated. The main source of forgetting in the model is retroactive interference: If feature  $x$  of item  $n+1$  is identical to feature  $x$  of item  $n$ , then the value representing feature  $x$  of item  $n$  is lost and cannot be used as a recall cue. A simplifying assumption is that only immediately adjacent items interfere.

In the feature model, order information is represented as a point in multidimensional space and this point can perturb (or drift) along the relevant dimension as described by Estes (1972). The probability that a cue’s encoded representation will perturb along the position dimension during a particular time interval is given by the parameter  $\theta$ , which is held constant at .05. The probability that an item,  $i$ , will occupy a particular position  $p$  during the next time interval is therefore given by the probability that no perturbation occurred ( $1 - \theta$ ) and the item was already in position  $p$  plus the probability that a perturbation occurred and the item was one position away from position  $p$ . Recall begins by determining, for each cue in primary memory, which was most likely to occupy position 1 originally. To recall the second item, the cue that was most likely in position 2 is used, and so on (see Neath, 1999a for a full explanation). The relative number of features available to cue the item in SM then dictates recall performance. The probability that a particular SM trace  $SM_j$  will be retrieved for a particular PM cue  $PM_i$  is calculated according to Equation 1. The value  $M_k$  is equal to 1 if the feature at position  $k$  of PM cue  $i$  does not match the feature at the corresponding position of SM representation  $j$ , and is 0 otherwise. The number of feature mismatches (the numerator of Equation 1) is divided by the value  $N$  (the number of features in each of the vectors) and the results summed. The parameter  $a$  is a scaling parameter representing overall level of attention.

$$d_{ij} = a \sum \frac{M_k}{N} \quad (1)$$

Next, the difference between the PM and SM items is transformed to provide a similarity metric (Equation 2).

$$s(i, j) = e^{-d_{ij}} \quad (2)$$

Equation 2’s similarity metric is used to calculate the probability that a particular secondary memory trace  $SM_j$  will be “sampled” given a particular primary memory cue  $PM_i$ . The probability of sampling a particular item is given by a similarity-based choice rule (Equation 3).

$$P(SM_j | PM_i) = \frac{s(i, j)}{\sum_{l=1}^{\text{no of vectors}} s(i, l)} \quad (3)$$

Next, the probability of recovering a sampled item is given by Equation 4, where  $c$  is a constant and  $r$  is the number of times the sampled item has already been recalled on this trial. This equation, and the  $r$  parameter, are used to reduce the likelihood of recalling the same item on multiple occasions, which avoid doing even when the same item is repeated within the to-be-recalled list (the “Ranschberg effect”, Jahnke, 1969). If two attempts at recovery are unsuccessful, an omission error is recorded.

$$P_r = e^{-cr} \quad (4)$$

### Calculating the PSE

The PSE is a difference score between recall levels for phonologically dissimilar items and recall levels for phonologically similar items, calculated (following Logie et al., 1996 and Neath et al., 2003) as a proportion of the mean performance in the dissimilar condition according to the following equation:

$$\text{PSE} = (D-S) / D \quad (5)$$

Where  $D$  is the mean performance in the dissimilar condition and  $S$  is the mean performance in the similar condition. Performance of the model will be compared to experimental data using this metric.

## Experiment

Experimental data were obtained from an undergraduate sample asked to reconstruct the serial order in which sequences of phonologically dissimilar, or similar, items were presented. Phonological similarity is implemented in the model by varying the number of modality-independent features with common values for each of the list items and keeping all other parameters constant.

### Method.

**Participants.** One hundred undergraduates of Purdue University participated in exchange for course credit. All were native US English speakers. For the modeling analog of this sample, results were generated for one hundred pseudo-participants, using the parameter values given in the Appendix.

**Stimuli.** The to-be-remembered items were 64 one-syllable words previously employed in Surprenant, Neath &

LeCompte's (1999) study of PSE. An example similar list is *vote, boat, float, note, coat*. An example dissimilar list is *break, sick, vote, greet, rat, fun*. All participants saw the same words in different random orders.

**Procedure.** The to-be-remembered stimuli were shown, in 20-point upper-case Helvetica, for 1.5 s each. Following the presentation of the final stimulus word, 6 response buttons became active, labeled with the 6 words in alphabetical order. Participants were asked to indicate the order in which the words had been presented by clicking on the appropriately labeled buttons using a mouse. Participants were presented with 20 lists, half with dissimilar items and half with similar items. The order of the dissimilar and similar items was separately and randomly determined per participant. Participants were tested individually and the experimenter remained within the room throughout. Pseudo-participant performance was based on 20 trials per pseudo-participant, half with similar and half with dissimilar items exactly as with the experimental participants.

**Results.**

Participants correctly recalled more dissimilar than similar items ( $F(1,99) = 101.89, MSE = 0.052, p < .05$ ) and there was also a main effect of serial position ( $F(5,495) = 156.88, MSE = 0.015, p < .05$ ) which interacted with similarity, ( $F(5,495) = 5.94, MSE = 0.014, p < .05$ ). A basic fit to the data was also obtained by running 100 pseudo-participants and calculating the results for this group in the same manner as for the experimental participants. Figure 1 shows the serial position curves obtained for participants (upper panel) and pseudo-participants (lower panel).

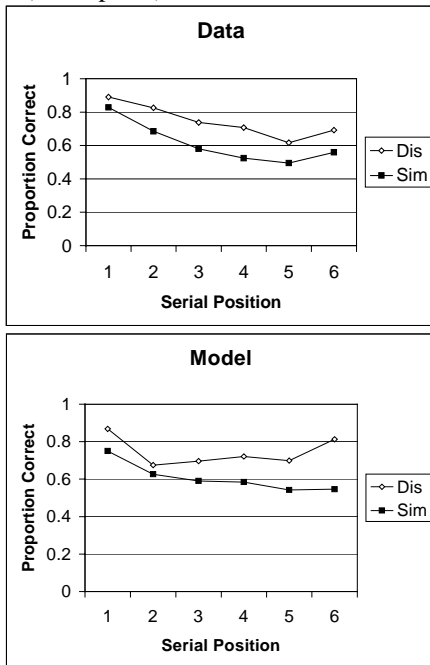


Figure 1: Proportion correct per serial position (averaged over participants or pseudo-participants) for data and model

Distributions of effect sizes for the PSE were also calculated. The mean PSE was 0.173 (lower quartile = 0.091; median = 0.198; upper quartile = 0.262; S.D. = 0.128; range = -0.143 to 0.442) and the distribution did not differ from normal (Kolmogorov-Smirnov  $Z = 0.813, n.s.$ ). However, in this basic fit, only 3 pseudo-participants showed greater recall for similar than dissimilar words, fewer than in the human sample (Figure 2).

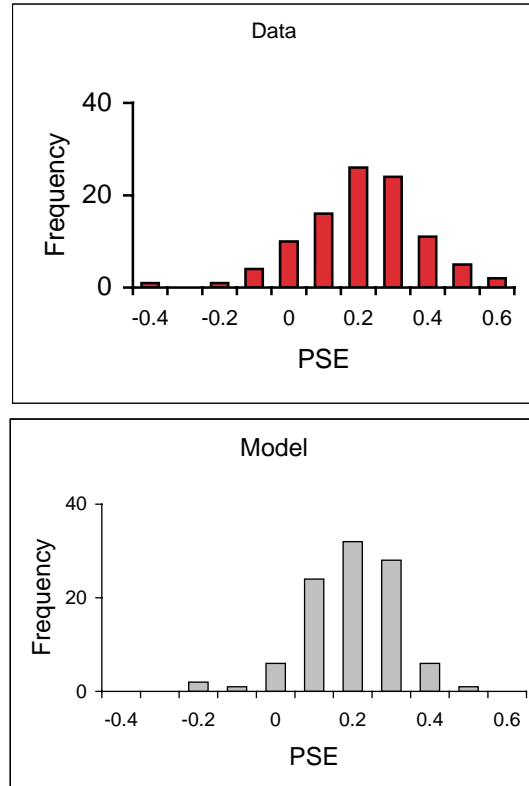


Figure 2: Distribution of effect sizes for the experimental sample and for a randomly generated sample of pseudo-participants.

For the basic fit shown in Figures 1 and 2, no parameters were varied between pseudo-participants. These Figures indicate that chance, rather than strategy, variation may be sufficient to account for the individual differences in displaying the effect. Figure 3 shows the results of a more principled approach to modeling the distribution within the experimental sample. Whereas Figures 1 and 2 show how random variation within the pseudo-participant sample affects the serial position function (Figure 1) and the distribution of effect sizes (Figure 2), Figure 3 shows fits of the model to the mean and variance of the experimental sample, again employing a sample of pseudo-participants but this time randomly varying the attentional scaling parameter,  $a$ , across pseudo-participants. For each pseudo-participant,  $a$  was increased (or decreased) from its standard setting by sampling randomly from a normal distribution with mean 0 and s.d. 4. This parameter was varied as is the only parameter (not linked to stimulus characteristics) that affects overall performance levels (Beaman, Neath & Surprenant, 2007). As

the figure demonstrates, attempts to fit either mean or variance for a PSE experiment produce a more accurate estimate of the number of reversals in the sample (phonologically similar lists recalled better than dissimilar lists) but have little effect on other features of the data. The upper panel of Figure 3 shows a fit to the mean values from the experimental data, the lower panel represents the values obtained when the simulation run represents a fit to the variances from the same data-set.

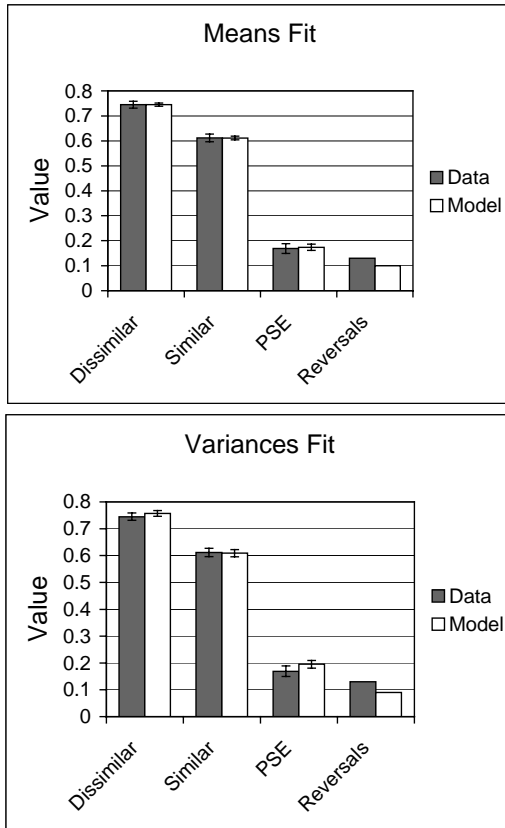


Figure 3: Fits of the model to means (upper panel) and variances (lower panel). Values on the y-axes represent the proportion correct for dissimilar and similar lists, the proportion value of the PSE calculated according to equation 5 and the proportion of the sample showing “reversals”; superior performance on phonologically similar lists. Error bars are standard error.

In addition to fitting the means and distributions of the PSE it is also possible, using the pseudo-participant technique, to examine how the PSE varies as a function of overall performance. Figure 4 shows the means and incidence of PSE and “reversals” amongst the upper and lower quartiles of the human participants and a pseudo-participant sample. These values came from the same simulation run that resulted in the best match to the observed variance. As this figure demonstrates, participants scoring in the upper quartile, based upon their performance in the dissimilar condition, are also showing the greatest PSE in both the data and the model with very few reversals in the data and none in the simulation. Surprisingly, more reversals appear in the human data in the

lower quartile of overall performance, and there is a lower mean PSE score although also more variability amongst this group. A Pearson’s correlation confirms a significant positive association between overall performance and size of the PSE,  $r = 0.259$ ,  $N = 100$ ,  $p = 0.01$ , consistent with Logie et al. (1996). It is also significant that both larger PSE amongst the upper quartile and more reversals amongst the lower quartile are predicted by the pseudo-participant simulation (as illustrated in Figure 4). The model also provides a significant positive association between overall performance and size of the PSE,  $r = .589$ ,  $N = 100$ ,  $p = .001$ . Both data and model also show the same patterns when the absolute size (D-S) of the PSE is calculated rather than a proportion ( $r = .360$  and  $.648$  respectively,  $N = 100$ ,  $p < .001$  in both cases).

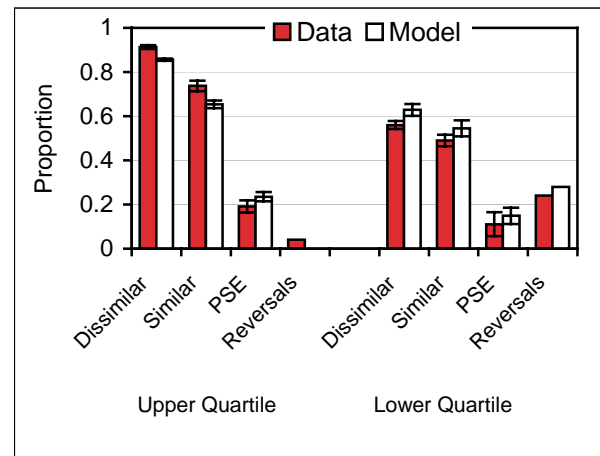


Figure 4: Sample performance divided into highest- and lowest-scoring quartiles on the dissimilar list condition for both participants and pseudo-participants. The values on the y-axis represent proportion correct for dissimilar and similar lists, proportion value of the PSE and proportion of the sample showing “reversals”. Error bars are standard error.

Overall, the model slightly under-predicts performance amongst the most able human participants and slightly over-predicts performance amongst the least able. It is noteworthy that neither the successful fit to PSE and reversals nor this potential failing of the model would be apparent from attempts to fit only mean performance data.

## Discussion.

The experiment reported here confirms the general pattern of results reported by Logie et al. (1996). A sizeable minority of participants either failed to show a PSE (9% or the total sample) or showed a reverse PSE (7%). These zero and negative effects were replicated using a model designed to show detrimental effects of phonological similarity on average. Furthermore, the model successfully accounts for these data, and other measures of variability, without the need to implement any qualitative changes in its operation. Previously, it has been suggested that a diminished effect of phonological similarity might reflect a change in strategy or type of encoding, for example from shorter- to longer-term

memory (Baddeley, 2000a; Salamé & Baddeley, 1986) or to a different type of short-term memory (Baddeley, 2000b). The simulations reported here demonstrate that a merely quantitative change is sufficient to recreate such variation in effect sizes without the need to hypothesize alternative memory systems. The simulations also show that the feature model is capable of simulating group variation by means of computational experiments employing “pseudo-participants”.

**Post hoc fits versus theoretical predictions.** The current study demonstrates that the phonological similarity effect, previously considered the empirical “signature” of the phonological store of working memory (Jones, Macken & Nicholls, 2004; Jones, Hughes & Macken, 2006) might be expected to vary as a function of quantitative changes in participants abilities regardless of their strategy choice. The question arises, however, whether the simulations reported here represent post-hoc fits between the model and the data or a genuine, and novel, prediction. The serial position functions and distributions reported in Figures 1 and 2 reflect neither a post-hoc parameter fitting exercise nor a theoretically motivated prediction as they act simply as an existence-proof that the model can, when sampled using a random selection of pseudo-participants, reproduce approximately the pattern of performance obtained when a group of experimental participants are likewise recruited using an opportunity-sampling recruitment procedure. In contrast, the fits to the mean and variance data shown in Figure 3 are the result of post-hoc parameter fitting as employed by numerous cognitive modeling studies and, although they again demonstrate the capability of the model to match the data, these post-hoc fits also lack something in the way of explanatory power. The subsequent splitting of the samples into high and low performance groups, however, was carried out after the parameters had been determined and thus comprises a novel and genuinely theory-driven prediction of how the data from the experimental participants should appear: a larger PSE amongst the higher-performing individuals is a necessary prediction of the model.

The reason for this observed pattern of performance, which was not obvious *a priori*, lies with the attention parameter. Aside from a few randomly varying features, which cannot plausibly account for the positive Pearson’s correlation between the appearance of the PSE and overall performance, the only parameter that varies between low and high performance pseudo-participants is the attention parameter. Close inspection suggests that it is the role of the attention parameter in magnifying the distance scores between mismatches and potential matches (Equation 1) that magnifies the appearance of the PSE in high-performance pseudo-participants relative to low-performance pseudo-participants. Thus, a single, potentially quantifiable, resource, the value of which is represented in the feature model by the attentional parameter, necessarily alters the pattern of immediate memory effects observed. This has a number of consequences for within-participant experimental designs as well as for between-participant differences. For example, the

diminishment of the PSE as memory-load increases might, contrary to previous suggestions (Baddeley, 2000a; Salamé & Baddeley, 1986) reflect a drop in attentional capability rather than a tendency to switch strategies as list length increases. Preliminary evidence already exists that the feature model is capable of accurately reproducing the effects of manipulating list-length (Beaman, 2006).

**Conclusions.** Given that the attention parameter is the only parameter explicitly varied to obtain our variance fits, it also seems that this is sufficient to shift from a high-performing (pseudo-) participant showing phonological similarity effects to a lower-ability participant for whom such effects are both smaller in size and more variable in appearance. This observation has a number of implications.

Firstly, and most generally, it suggests that purely quantitative variations in a single performance factor could appear, within the experimental data, as qualitative shifts in performance characteristics. The example explored here is that of the phonological similarity effect, but there is no a priori reason why the same might not hold true for other well-known immediate memory effects (e.g., word-length, concurrent articulation or irrelevant sound). This observation provides support for the claim that quantitative computational simulations are a productive method for testing (often implicit) theoretical assumptions that might otherwise go untested (Lewandowsky, 1993; Neath, 1999b).

Secondly, more specifically, the results imply that the deployment of particular storage buffers (Baddeley, 2000a, b), is not necessary to explain reductions in phonological similarity effects amongst individuals undertaking immediate memory tasks. The feature model predicts this pattern of results using only a single primary memory “store” and no maintenance rehearsal.

Thirdly, since the feature model is essentially a model of the recall process that views memory as a discrimination activity (e.g., Nairne, 2002), it is possible that individual variation amongst human participants might likewise be a consequence of variation in capability to discriminate at recall rather than a difference in either encoding strength or maintenance activity.

## Acknowledgments

Portions of this research were carried out whilst the second and third authors were at Purdue University, Indiana, USA, and also whilst they were visiting professors at City University, London, UK.

## References

- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30, 221-256.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. D. (2000a). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin & Review*, 7, 403-423.

- Baddeley, A. D. (2000b). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience*, 4, 829-239.
- Beaman, C. P. (2006). The relationship between absolute and proportion scores of serial order memory: Simulation predictions and empirical data. *Psychonomic Bulletin & Review*, 13, 92-98.
- Beaman, C. P., Neath, I., & Surprenant, A. M. (2007). *On abandoning the phonological store*. Paper presented to the joint meeting of the Experimental Psychology Society and the Psychonomics Society, Edinburgh, UK.
- Brown, G. D. A., Della Sala, S., Foster, J. K., & Vousden, J. I. (in press). Classical amnesia within a unitary model of memory. *Psychonomic Bulletin & Review*.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, 55, 75-84.
- Cooper, R. P., Yule, P., & Fox, J. (2003). Cue selection and category learning: A systematic comparison of three theories. *Cognitive Science Quarterly*, 3, 143-182.
- Della Sala, S., & Logie, R. H. (1997). Impairments of methodology and theory in cognitive neuropsychology: A case for rehabilitation? *Neuropsychological Rehabilitation*, 7, 367-385.
- Estes, W. K. (1972). An associative basis for coding and organization in memory. In: A. W. Melton & E. Martin (Ed.s), *Coding processes in human memory* (pp. 161-190). Washington, DC: Winston.
- Hintzmann, D. L. (1986). "Schema abstraction" in a multiple-trace model. *Psychological Review*, 93, 411-428.
- Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical transactions of the Royal Society of London, Series B*, 302, 325-340.
- Hulme, C., Thomson, N., Muir, C., & Lawrence, A. L. (1984). Speech rate and the development of short-term memory. *Journal of Experimental Child Psychology*, 38, 241-253.
- Jahnke, J. C. (1969). Output interference and the Ranschberg effect. *Journal of Verbal Learning & Verbal Behavior*, 8, 614-621.
- Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory & Language*, 54, 265-281.
- Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 656-674.
- Lewandowsky, S. (1993). The rewards and hazards of computer simulation. *Psychological Science*, 4, 236-243.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251-269.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53-81.
- Neath, I. (1999a). Modeling the effects of irrelevant speech on order. *International Journal of Psychology*, 34, 410-418.
- Neath, I. (1999b). Computer simulations of global memory models. *Behavior Research Methods, Instruments & Computers*, 31, 74-80.
- Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, 10, 430-434.
- Salamé, P., & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, 24, 263-265.
- Surprenant, A. M., Neath, I., & LeCompte, D. C. (1999). Irrelevant speech, phonological similarity and presentation modality. *Memory*, 7, 405-420.
- Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning & Verbal Behavior*, 23, 151-161.
- Vallar, G., & Shallice, T. (1990). *Neuropsychological impairments of short-term memory*. Cambridge: Cambridge University Press.

## Appendix

Table: Parameter values for reported simulations. Values are shown in the "Similar" column only when they differed from those for the "Dissimilar" condition. For the "variance" simulation, attention (*a*) was set to 13.5.

	Dissimilar	Similar
MI Features		
<i>Number</i>	20	
<i>Range</i>	4	
<i>No. guaranteed similar</i>	0	8
MD Features		
<i>Number</i>	2	
<i>Range</i>	4	
<i>No. guaranteed similar</i>	0	1
Attention ( <i>a</i> )	10	
No. of perturbation opportunities	5	
Recovery Constant ( <i>c</i> )	2	
No. of recovery attempts ( <i>r</i> )	2	