# Representations of Serial Order

R. N. A. Henson
Department of Psychology, University College London, UK

N. Burgess
Department of Anatomy, University College London, UK

**Abstract**

Three means of representing serial order in connectionist models are identified: interitem associations (e.g., the recurrent network of Jordan [1]), ordinal representations (e.g., the activation gradient of Grossberg [2]), and positional representations (e.g., the control signal of Houghton [3]). Error data from studies of human short-term memory favour positional representations. Three types of positional representations are possible: those of temporal position (e.g., the OSCAR model of Brown et al. [4]), absolute position (e.g., the Articulatory Loop model of Burgess & Hitch [5]) and relative position (e.g., the Start-End Model of Henson [6]). Recent data [7] favour representations of relative position. A connectionist implementation of relative position is discussed.

## 1 Introduction

The problem of serial order is to explain how people store and retrieve a sequence of items in the correct order. This problem pervades many aspects of cognition, from the ordering of digits in a telephone number to the ordering of phonemes in a spoken word. However, since the importance of the problem was raised by Lashley [8], an agreed solution has proved surprisingly elusive.

This chapter begins by reviewing three main approaches to the problem. These approaches have their roots in psychological theories and are exemplified in several recent connectionist models. When tested within the domain of short-term memory (STM) however, only one of these approaches appears viable. This approach assumes that each item is coded for its position within a sequence. The question then becomes whether that position is represented temporally, absolutely or relatively. Connectionist models of STM tend to assume representations of temporal or absolute position, whereas recent psychological data [7] favour representations that code position relative to the start and end of a sequence. The chapter concludes with a new connectionist model that accommodates these data by assuming an array of oscillators of different frequencies that compete to best represent the input.

## 2 Three Approaches to Serial Order

Existing approaches to the problem of serial order can be categorised as interitem associations, ordinal representations or positional representations.

## 2.1 Interitem associations

This approach assumes that a sequence is stored by the formation of associations between representations of successive items. The order of items is retrieved by stepping along the chain of associations, such that each item becomes (part of) the cue for recall of its successor (also known as *chaining theory* [9]). Interitem associations are probably the oldest approach to serial order [10], being a simple extension of stimulus-response theory in which each response can become the stimulus for the next [8].

The simplest chaining theories assume only pairwise associations between representations of adjacent items [11] and cues that consist entirely of the preceding response. However, these theories face problems with a) repeated items, because the items following a repetition will share the same cue, and b) erroneous responses, because the cue for subsequent responses will be incorrect. More sophisticated theories overcome these problems by assuming remote associations as well as adjacent ones [12]. In such *compound chaining theories* (Figure 1A), the cue consists of a number of preceding items, providing additional context with which to disambiguate repeated items and to cater for occasional errors in recall.[1] Connectionist implementations of compound chaining theories include the recurrent neural networks of Jordan [1], Elman [13] and Taylor [14], which have successfully modelled some aspects of sequence production and recognition. Nonetheless, there are general arguments against chaining theory [8, 15] and it is argued later (Section 3.1) that there is no empirical support for interitem associations underlying STM for serial order.

## 2.2 Ordinal representations

Ordinal representations assume a single dimension along which order is defined, such as the relative strengths of item representations in memory. The connectionist models of Grossberg [2] for example assume that order is stored in a primacy gradient of strengths, such that the representation of each item is stronger than that of its successor. The order of items is retrieved by an iterative process of selecting the strongest item representation, and then suppressing it so that it is not selected again (Figure 1B). Other ordinal representations include the cyclic reactivations of Estes [16] and activation gradient in the Primacy Model of Page and Norris [17] (see Norris & Page, this volume).

Ordinal representations generally require token representations in order to handle repeated items: The order of sequences with repeated items can not be represented over type representations, each with a single strength. Because order is defined relationally, ordinal models also imply that errors will cooccur: If, for example, the representation of an item becomes stronger than that of its predecessor in the Primacy Model, owing to random noise, then the two items will transpose [17]. This is an attractive property, because such paired transpositions of adjacent items are common in people too (Section 2.2).

Unlike chaining models, models like the Primacy Model do not require feedback of responses, and a process like suppression can operate independently of errors occurring at later stages of output. Moreover, the processes of selection and suppression are simple to implement in connectionist models as winner-take-all networks (e.g., the competitive filter of Houghton [3]). Nonetheless, it is argued below (Section 3.2) that ordinal representations are not sufficient to

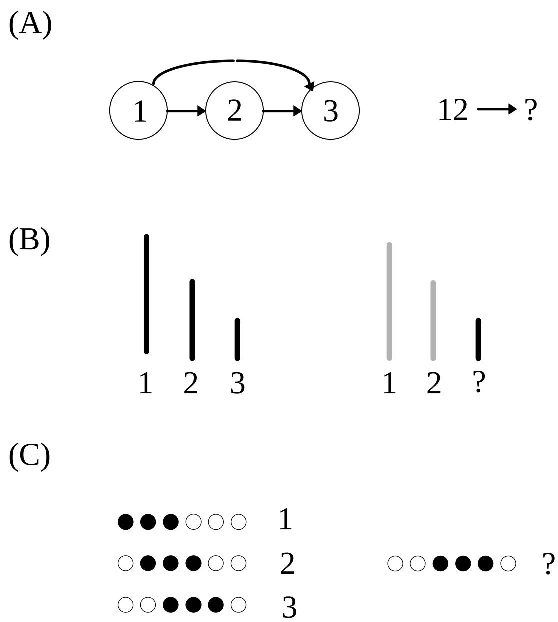

Figure 1: (A) Remote interitem associations (e.g., Jordan [1]); (B) an ordinal representation (e.g., Grossberg [2], suppression indicated by lighter lines); (C) a positional representation (e.g., Burgess & Hitch [5], filled circles represent active nodes in a connectionist network, unfilled nodes represent inactive nodes).

account for the errors people make in recall from STM.

## 2.3 Positional representations

This approach assumes that order is stored by associating each item with its position in a sequence, and that the order is retrieved by reinstating each positional code and cueing the associated item.

The extreme case of a positional representation is Conrad's "box" model [18]. Conrad suggested that people possess a number of boxes in STM in which item representations can be stored. The items can be retrieved in order by stepping through the boxes according to a predetermined routine. This model does not have a problem with repeated items, because they are stored in separate boxes, nor with erroneous responses, because the retrieval mechanism can continue to the next box irrespective of whether the contents of the previous box were retrieved correctly. This is of course the method by which conventional Von Neumann computers store and retrieve order, through routines accessing separate addresses in memory.

As a psychological theory, this approach is elaborated in the Perturbation Model of Lee and Estes [19, 20], in which of the positions of items are initially coded perfectly, but get perturbed over time such that nearby items are likely to exchange. An alternative proposal is that positional codes are not perfect, but overlap, in that the code for one position is similar to the codes for nearby positions. This is the approach taken in the connectionist model of Burgess and Hitch [5, 21, 22]. A "window" of activity moves from left to right across an array of nodes for each position in a sequence (Figure 2C), and is associated with other nodes (not shown) representing each item. However, because there is some overlap in the set of active nodes for nearby positions, items at these positions can be confused during retrieval.

The main question facing positional models is how the positional codes themselves are reinstated in the correct order. One suggestion is that the codes are derived from temporal oscillators in the brain [4, 21, 23]. Item representations can be associated with successive states of the oscillators, and these states reinstated simply by resetting the oscillators and letting them evolve under their own dynamics. However, though there is good evidence for positional representations in STM (Section 3.2), one goal of the present chapter is to argue for a modified interpretation of the oscillators assumed to underlie positional codes.

# 3 Evidence from Short-Term Memory

In spite of the strengths and weaknesses of the specific connectionist models mentioned above, important differences remain between the three general approaches to serial order. The difference between interitem associations and positional representations is obvious: The retrieval cue in the former is the previous item; the retrieval cue in the latter is some (abstract) positional code. The difference between positional and ordinal representations is less obvious, but relates to whether the position of an item can be defined independently of surrounding items. In positional representations, it can; in ordinal representations, it cannot. The consequence is that, with ordinal representations like that in the Primacy Model [17], the middle item in a sequence can only be retrieved after retrieval of its predecessors. With positional representations however, it is possible in principle to retrieve the middle item without retrieving its neighbours, by reinstating the appropriate positional code.

These differences can be tested empirically within the domain of STM by using the memory span task, in which subjects must recall a novel list of items in the correct order. It is particularly fruitful to examine the errors people make when they misrecall a list. It will be argued that these errors necessitate some type of positional representation in STM.

## 3.1 Evidence against interitem associations

The main prediction of chaining models is that recall of an item will depend on the properties of its predecessor. In particular, "...errors are more likely when discriminations must be made between similar states..." (Jordan, p.37 [1]). For example, it is well-established that lists of similar-sounding items (e.g., *BTGPDV*) are harder to recall in order than lists of dissimilar items (e.g., *HRMQJY*), even when presented visually (e.g., [24, 25]). This suggests that items are represented in STM in a phonological form. If so, chaining models predict that (part of) the difficulty people face with lists of phonologically similar items stems from the similarity between the cues for each item. However, using lists in which similar and dissimilar items alternated in order (e.g., *BRGQDY*), Henson et al. [25] found no effect of whether or not the previous item was phonologically similar to other items in the list. This is problematic for any model that chains along associations between phonological representations. Furthermore, Henson et al. found that the probability of recalling a dissimilar item appeared independent of whether or not the previous similar item was recalled correctly. This is troublesome for any closed-loop chaining model that assumes responses are fed back to cue subsequent items.

An extreme case of similarity is of course identity. Even compound chaining models with remote associations predict that there should be a greater probability of errors following repeated items (assuming they employ type representations of items). For example, in a list *HRMRJY*, there should be more errors in recalling the items following the repeated item (i.e., *M* and *J*) than for items at corresponding positions in control lists with no repeated items. In fact, chaining theory predicts that these errors are likely to be exchanges between the following items themselves (e.g., *HRMRJY* recalled as *HRJRMY*), given that they share the same cue. Preliminary support for this hypothesis was reported by Wickelgren [26]. However, several recent experiments [9] found only a small effect of repetition on recall of subsequent items, which failed to reach significance. Furthermore, any small difference that is found may well have alternative explanations, given that the sheer presence of repeated items has several effects on recall of a list [27]. For example, because there are fewer different items to guess from in a list with a repeated item than one without, a simple guessing hypothesis also predicts a higher incidence of exchanges following repeated items when compared with control lists. Thus there does not appear to be any conclusive evidence for an effect of repetition on cueing either.

The failure to find reliable evidence for an effect of phonological similarity, errors or repetition on cueing is problematic for chaining models. It may be possible to construct a specific chaining model that is consistent with the above data (such a model might chain along associations between nonphonological, token representations for example, independently of response feedback). However, given that there is not, as yet, any positive evidence for chaining, and that there is positive evidence for positional representations (below), it seems reasonable to argue against interitem associations on the grounds of parsimony.

### 3.2 Evidence for positional representations

The most common errors in serial recall from STM are order errors, or *transpositions*. The most striking aspect of these errors is their distribution: Erroneous items are clustered around their correct position, rather than being randomly distributed (e.g., [16, 25]). This finding is often taken as evidence for positional representations, by suggesting that there is some generalisation across positional codes that occasionally causes errors between items at nearby positions. However, the finding does not force this conclusion, because similar distributions of transpositions can be produced by models with ordinal representations (see [17]): Errors in the relative order of nearby items also produce the appropriate clustering of transpositions, without any coding of the position of those items. In fact, the same pattern can even be produced by compound chaining theories [28].

However, there are two types of error that do imply the presence of positional representations. The first of these occurs when lists are grouped. Grouping items by their rhythm of presentation, as common with telephone numbers for example, is well known to improve recall (e.g., [29]). Though grouping reduces the overall incidence of errors, one type of error actually increases [30]. These *interpositions* [9] are transpositions between groups that maintain their position within groups (and are not simply the result of whole groups swapping [9, 20]). These errors imply that items can be coded for their position within a group independently of surrounding items.

The second type of positional error is found between recall of lists on successive trials. Conrad [31] showed that an erroneous item in one trial is more likely than chance to have occurred at the same position in the previous trial. Henson [9] called the errors caused by such proactive interference of positional information *protrusions*. These errors imply that items can be coded for their position within a trial.

Protrusions and interpositions are examples of a general tendency for errors between sequences to maintain their position within a sequence. Such errors cannot be attributed to errors of relative order within a sequence, and are therefore inexplicable by ordinal representations. Nor can they be explained by interitem associations. Positional errors can only be explained if STM for serial order utilises positional representations.

# 4 Positional Representations

Three types of positional representations can be distinguished: representations of temporal position, absolute position and relative position.

## 4.1 Temporal Position

In models that code temporal position, items are associated with a representation of their time of occurrence (often relative to the start of a sequence). Such representations have been proposed by both Burgess and Hitch [21, 22] and Brown et al. [4]. In particular, the OSCAR model [4] (see also Vousden & Brown, this volume) is a connectionist model that assumes an array of free-running oscillators of different frequencies, the phases of which can be combined to form a smooth "timing signal". Items are associated with the states of this signal at their time of presentation, and recalled in order simply by resetting the oscillators to their states at the start of presentation.

A simple formalisation of such a timing signal uses an array of harmonically-related sinusoidal oscillators $s = 1..N$, such that the phase of oscillator $s$ at time $t$, $\phi_s(t)$, is:

$$\phi_s(t) = sin(\frac{2\pi t}{2^{s-1}B} + \Phi_s)$$

where $B$ is the period of the *base* (fastest) oscillator $s = 1$ and $\Phi_s$ is a random starting phase for oscillator $s$ at $t = 0$. The timing signal can then be represented by a dynamic vector $\mathbf{v}(t)$ generated from a weighted combination of the oscillator phases, with more weight given to slower oscillators, such that:

$$v_s(t) = W^{N-s}\phi_s(t)$$

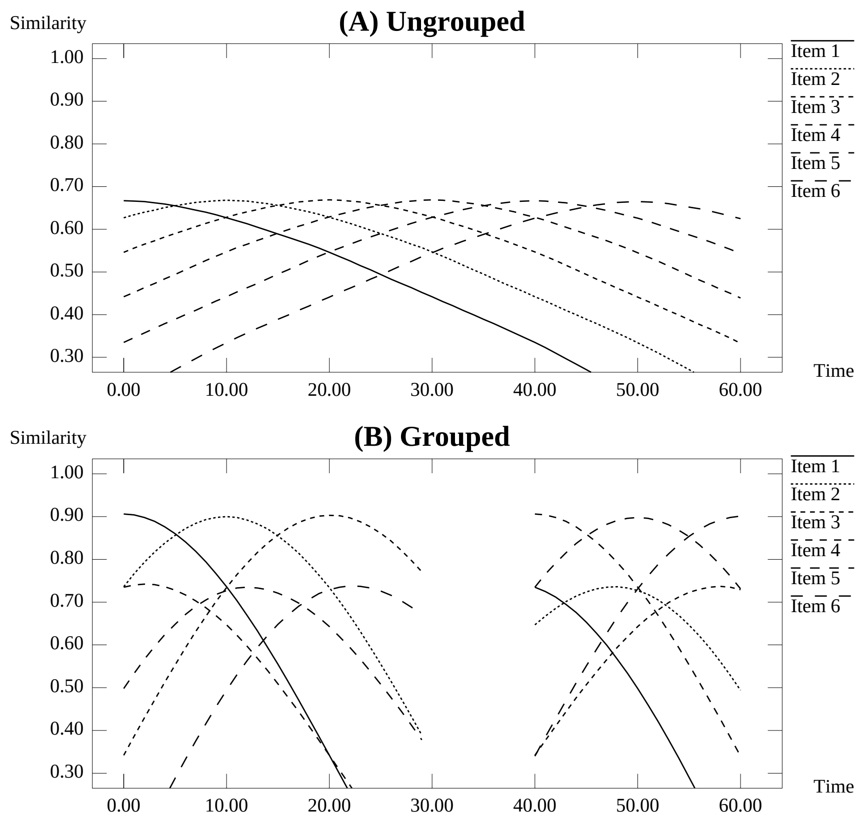


Figure 2: Similarity gradients for six items presented every 10 time units (A) and for six items presented as two groups of three (B) via an extra 10 time units between the third and fourth item (*B=10, N=6, W=0.5*, weighting of oscillator $s{=}4$ increased to *1.0* for the grouped list). Gradients are averaged over 10,000 runs with random initial phases for each oscillator.

where $W < 1$ is a weighting parameter. Providing there are sufficient oscillators, the period of the slowest oscillator can be made long enough to ensure that the timing signal does not repeat within the timescale of interest.

Assuming the initial state of the timing signal during presentation can be reinstated at recall, the similarity (inner product) between the state of the timing signal for each position in a list of six items and its state at each time point during recall is shown in Figure 2A (averaging over random initial phases, $\Phi_s$; equivalent to the combination of multiple timing signals in OSCAR). These similarity gradients indicate the degree of temporal generalisation across positions. In a model where items were output when this similarity exceeded a threshold, any noise in the temporal codes or in the thresholding process would produce the clustered pattern of transposition errors seen in the data (the greater similarity between Positions 1 and 2 than between Positions 1 and 3 making Item 1 more likely to transpose with Item 2 than with Item 3).

Temporal grouping can be simulated by giving more weight to the oscillator whose half-period is closest to the group period. This might arise for example if the rhythm of presentation caused people to focus attention on the timescale of groups. The further assumption that this group oscillator can be reset between groups (so that all oscillators are reset at the start of recall, but only the group oscillator is reset at the start of each group) produces the similarity gradients shown for two groups of three in Figure 2B. In this example, the oscillator $s = 4$ with a half-period equal to the group period (40 time units) had its weighting increased from 0.25 to 1.00 (and the weightings of other oscillators were normalised so that the sum total of weightings remained constant). The effect of this resetting and increased relative weighting is to improve the temporal resolution of positions within groups, by producing taller and sharper similarity gradients, but impair the resolution of positions between groups, by increasing the similarity between the same positions within groups (e.g., increasing the similarity between the Positions 1 and 4). Both these properties are true of the data [9], the latter accounting for the interpositions between groups that support a positional representation of serial order (Section 3.2).

If it is also assumed that the oscillators are yoked such that an increase in the frequency of the base oscillator produces a proportionate increase the frequency of the other oscillators, then the recall rate can be uncoupled from the presentation rate. A tonic signal to increase the frequency of the base oscillator prior to recall for example would allow the items to be recalled faster than they were presented. However, unless this tonic signal can also vary *during* presentation and recall, a model using temporal representations always predicts that the relative timing of recall of each item will match that of presentation of each item. As discussed later (Section 5.1), this property proves problematic.

## 4.2 Absolute Position

Rather than coding each item with its temporal position, items can be coded with their absolute (ordinal) position (e.g., first, second, third, etc.). This is the positional representation adopted by event-driven models, such as the original connectionist model of Burgess and Hitch [5]. In this model, the context signal (Figure 1C) only changes as each item occurs, irrespective of the temporal spacing between items. Thus the positional codes for a list of items presented slowly are identical to those for the same list presented rapidly. However, it will be argued later (Section 5.2) that even this level of abstraction is not sufficient to explain the pattern of errors in recall from STM.

## 4.3 Relative Position

The most abstract representation of position is a relative one, in which items are coded for their position relative to the start and end of a sequence. This is the representation chosen in the computational model of Henson [6, 9] and in the connectionist model of Houghton [3]. These models assume a start marker, which is strongest at the start of a sequence but decreases in strength with each subsequent item, and an end marker, which increases in strength with each item towards its maximum strength at the end of a sequence. The relative strengths of the start and end markers therefore provide an approximate two-dimensional code for each position within a sequence.

One problem facing these models is to specify how items are coded relative to the end of the sequence when the end of the sequence has not yet occurred. Henson [7] suggested that the strength of the end marker might correspond to the degree of expectation for the end of a sequence. Houghton [3] assumed that the end marker was only triggered at the very end of presentation, upon which it was associated with a recency gradient of decaying activations of item representations. By growing in strength more gradually during recall, these associations allowed the end marker to exert an influence backwards in time. Both these models are consistent with the evidence for relative position discussed below (Section 5.2). Nonetheless, Section 6 describes a new model that codes relative position without the problems associated with an end marker.

# 5 Evidence from Short-Term Memory

There is also evidence from STM experiments concerning the three types of positional representation. Though the evidence is not clear cut, the pattern of positional errors between sequences certainly favours a representation of relative position over representations of purely temporal or absolute position.

## 5.1 Evidence for and against temporal position

One clear prediction of representations of temporal position is that the positions of items closer in time will be harder to discriminate. In STM experiments, this implies that slower presentation rates should produce better recall. However, one problem with this prediction is that subjects will often rehearse items subjectively, at a rate that may differ to the objective presentation rate. Faced with a slow presentation rate for example, rehearsal rate is likely to exceed presentation rate. When covert rehearsal was prevented by concurrent articulatory suppression, Baddeley and Lewis [32] found that serial recall was worse for slow presentation rates, contrary to the prediction of temporal position. Likewise, when covert rehearsal was minimised by articulation of distractors between presentation of each item, Neath and Crowder [33] found that overall performance decreased as the number of interitem distractors increased, in spite of the greater temporal separation entailed.[2]

One objection to the above evidence is that slower presentation rates and greater numbers of interitem distractors also entail a greater delay between presentation and recall of each item. Given the transient nature of information in STM, longer delays are likely to produce a greater loss of information, which may override any advantage from more distinctive temporal coding. What is required is a comparison of performance when the temporal spacing between items is varied but the temporal spacing between presentation and recall of the first item is kept constant (and rehearsal is prevented). The authors are not aware any such STM experiment to date.[3]

Some evidence that supports a role for temporal factors comes from other findings by Neath and Crowder [33]. When comparing presentation schedules in which the temporal spacing between items either increased or decreased over successive positions (with the total presentation time equated at six seconds), free recall was better with the decreasing schedule than the increasing schedule. This is explicable if subjects adopted a backward temporal perspective, for

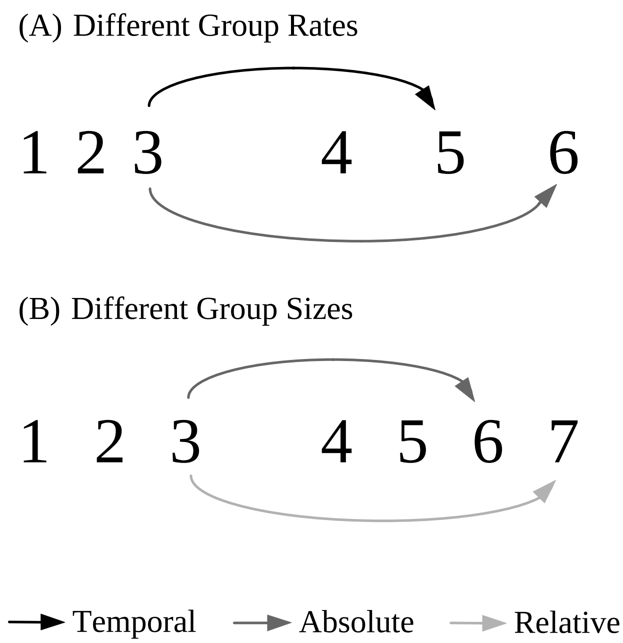


Figure 3: Examples of groupings used by (A) Ng [35] and (B) Henson [7] and the corresponding transposition errors predicted by representations of temporal, absolute and relative position.

which a greater separation of early items is more advantageous than a greater separation of late items (so-called *temporal distinctiveness theory*, cf. a line of telegraph poles receding from the observer). However, in experiments with a total presentation time closer to one second, Neath and Crowder [34] found that performance was better with the increasing schedule. The authors suggest that with such fast presentation rates (and a tendency to recall serially), subjects adopt a forward perspective, in which case it is more advantageous for later items to be widely separated. These data are returned to in Section 6, in relation to a connectionist model that incorporates some influence of temporal factors in the coding of relative position.

Perhaps more difficult for representations of purely temporal position are the error data from an experiment by Ng [35]. Ng used nine items grouped as three groups of three, such that the middle group was presented either faster or slower relative to the first and last groups (a two-group example is shown in Figure 3A). The interest was in whether transpositions between groups were more likely between the same temporal positions within groups (e.g., Positions 3 and 5) or the same absolute positions within groups (e.g., Positions 3 and 6). Ng found that transpositions were more likely between the same absolute position within groups, and argued in favour of an absolute or event-driven coding of position.

### 5.2 Evidence against absolute position

Because Ng's groups were of equal size however, her data are equally well compatible with a representation of relative position. To contrast the predictions of absolute and relative position, Henson [7] examined the transpositions between groups of different size, such as a group of three followed by a group of four (Figure 3B). Transpositions were more likely between the ends of groups (Positions 3 and 7) than between the same absolute position within groups (Positions 3 and 6). In a second experiment, Henson [7] also found that protrusions between trials of different length were more likely between the ends of trials than the same absolute position within trials. These data favour the representation of relative position over the representation of absolute position.

One caveat with the experiments of Ng [35] and Henson [7] is that neither controlled for covert rehearsal. In Ng's experiment for example, subjects may have rehearsed the groups at the same rate, in spite of the differences in their presentation rates. If so, the predictions of temporal and absolute position would be confounded. However, this possibility seems less likely for Henson's experiments, because there is no apparent reason why subjects should rehearse longer sequences faster than shorter ones, such that their total rehearsal time is equated. Thus the conclusion that positional representations in STM are relative rather than absolute will be maintained.

## 6 A Connectionist Model of Relative Position

The model of relative position described by Henson [6, 9] is a psychological-level model that is useful in allowing quantitative fits to data and making predictions for future experiments. However, it is not a connectionist model and makes no attempt to offer a neural-level implementation of relative position. Below we sketch such an implementation in terms of temporal oscillators, given the considerable evidence for such oscillators in the brain [36]. This implementation offers a novel positional representation that combines aspects of both temporal and relative position.

The basic idea behind the model is that many oscillators of different frequencies compete to best represent the input.[4] The winning oscillators are those with a half-period closest to the temporal duration of a sequence. Thus groups of different temporal duration (by virtue of different presentation rates or different numbers of items) are represented by different oscillators. During recall however, oscillators of a single frequency are selected (depending on the recall rate), and items are cued to the extent that the phases of these oscillators match the phases of the oscillators that won the competition to represent the sequence. The model is therefore able to adapt to different presentation and recall rates. Moreover, by having all oscillators compete in parallel, but only selecting those with a half-period close to the sequence duration (apparent as soon as no more items occur; see below), the model finesses the problem of predicting the end of a sequence that plagues models with an end marker.

More specifically, an array of oscillator pairs $p = 1..N$ with half-periods $H_p = 1..N$ time units is assumed, such that the phases $\phi(t)$ and $\phi'(t)$ of the oscillators in pair $p$ are:

$$\phi_p(t) = sin(\frac{\pi t}{H_p}) \qquad\qquad \phi'_p(t) = sin(\frac{\pi t}{H_p} + \frac{\pi}{2})$$

In other words, the two oscillators are phase-lagged by 90 degrees [37].

During presentation, the occurrence of the first item in a sequence causes all oscillators to be reset such that $t = 0$. The oscillators then run at their different speeds, each item being associated with the phases of all oscillators at its time of occurrence, until no further items occur. The winning oscillator pair then selected to represent the sequence is the pair $p = w$, where $H_w$ is the total duration of the sequence. The positional code for the item presented at time $t_1$ after the start of the sequence is the two-dimensional vector $< \phi_w(t_1)\ \phi'_w(t_1) >$. Assuming that oscillator pair $p = r$ is chosen for recall (and reset at the start of recall), then the strength with which the item presented at time $t_1$ is cued at time $t_2$ relative to the start of recall is given by the inner product of the positional codes for presentation and recall, equal to:

$$sin(\frac{\pi t_1}{H_w})sin(\frac{\pi t_2}{H_r}) + sin(\frac{\pi t_1}{H_w} + \frac{\pi}{2})sin(\frac{\pi t_2}{H_r} + \frac{\pi}{2}) = cos(\frac{\pi t_1}{H_w} - \frac{\pi t_2}{H_r})$$

In other words, the similarity gradient for the item occurring at time $t_1$ during presentation is a cosine function of the angle between the phase of the first oscillator in the pair selected at presentation and the phase of the first oscillator in the pair chosen for recall at time $t_2$ (without needing to average over oscillators with different initial phases).

The ability of this representation to capture the appropriate similarity gradients for grouped lists is shown in Figure 4. For these simulations, a number of further assumptions were made about grouping. Firstly, a group was assumed to end when no item occurred in a given interval after the most recent item, where that interval was equal to the time between the presentation of the previous two items. In other words, prior rhythmic parsing of the input was assumed, such that a group boundary was located whenever an item failed to occur before or in time with the beat defined by the last two items. Secondly, two sets of oscillator pairs were assumed: one in which all oscillators were reset at the start of a new group, and one in which oscillators were only reset at the start of a list. The winning oscillator pair(s) in the first set represented the relative position of items within groups, whereas the winning oscillator pair in the second set (those with a half-period equal to the length of the whole list) represented the position of groups within the list. Positional codes therefore consisted of two vectors, one comprising the phases of the winning group oscillator pair at the time each item was presented, and one comprising the phases of the winning list oscillator pair at the time the first item in each group was presented. During recall, the inner products between these vectors and those derived from the group oscillator pair and list oscillator pair chosen for output were combined by simple addition (and halved to normalise similarity to the range -1 to 1). Finally, it was assumed that reset of the group oscillator pair during recall was triggered when the phases of the list oscillator pair chosen for output matched those possessed at the start of presentation of each group by the list oscillator pair chosen for input (these phases perhaps being stored in temporary associations between the two sets of oscillators).

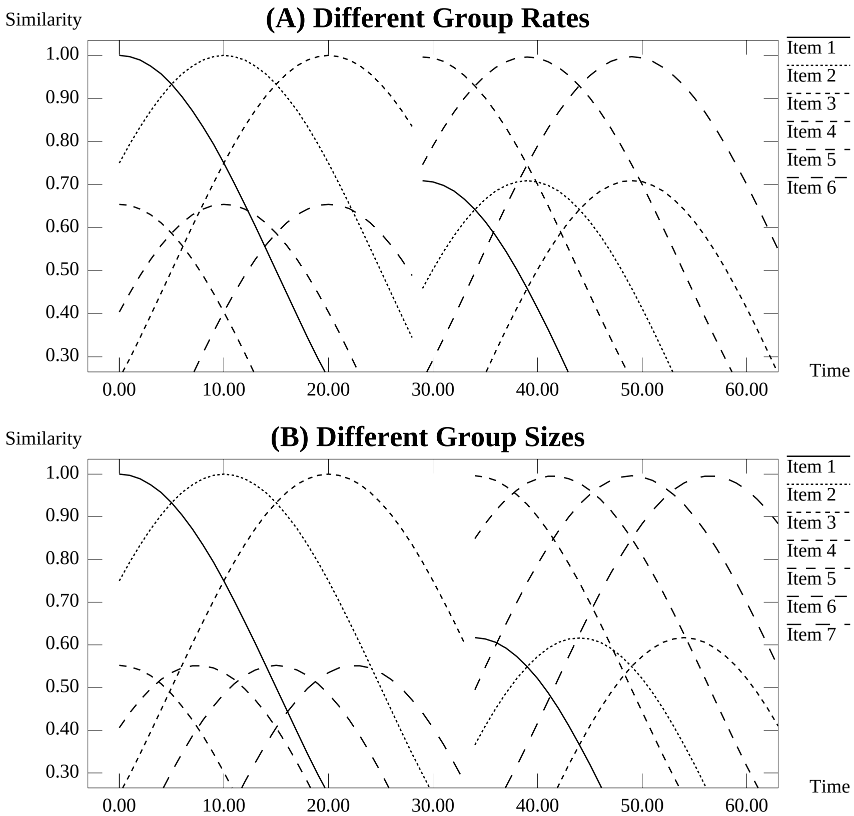


Figure 4: Similarity gradients for two groups of three items, the first presented every 10 time units and the second presented every 20 time units (A), and for a group of three items followed by a group of four items, presented every 10 time units (B), with an extra 10 time units between each group (half-periods of group and list oscillator pairs used for recall were 30 and 80 time units respectively).

Figure 4A shows the similarity gradients for two groups, the second of which was presented twice as slowly as the first (as in Figure 3A); Figure 4B shows the gradients for a group of three items followed by a group of four (as in Figure 3B). In both cases, the similarity between items in different groups is greatest for items at the same relative position within a group. Thus, in the case of groups presented at different rates, Item 3 is more likely to transpose with Item 6 than with Item 5, and in the case of groups of different size, Item 3 is more likely to transpose with Item 7 than with Item 6.

As well as being compatible with the error patterns in STM, it is interesting to note that such a representation also shows some influence of temporal factors. With the increasing or decreasing presentation schedules of Neath and Crowder [33, 34] for example, the parser will not identify any rhythm or grouping during presentation, and will therefore code the whole list with a single oscillator pair. Because the increasing and decreasing schedules will result in successive positions being represented with increasing and decreasing phase differences respectively, the model predictions coincide with those of temporal distinctiveness theory. In other words, the representation coded by a single oscillator pair is of relative, temporal position. Where the model might differ to temporal distinctiveness theory is with rhythmic presentation schedules that allow hierarchical representations of position. In these cases, the present model predicts that errors between sequences will respect relative rather than temporal position within sequences (because different-duration sequences are coded by different-frequency oscillators), consistent with the above data, whereas the predictions of temporal distinctiveness theory are unclear. Future work will hopefully clarify the similarities and differences between the two approaches.

## 6.1 Future Work

Several questions remain of the model outlined above. Foremost is the problem of synchronising the phases of the group and list oscillators used in recall. If such synchronisation is not achieved, the relative phases of the group and list oscillators selected after presentation will not match the relative phases of the group and list oscillators chosen for recall, and appropriate similarity gradients are not guaranteed. This problem was overcome above by coding items for the position of their group in the list (rather than directly for their position in the list), and by resetting the group oscillator pair during recall whenever the appropriate phases of the list oscillator pair were reached. There is still a problem in choosing the frequency of the group oscillator pair relative to that of the list oscillator pair at recall: If the group oscillator pair is too slow relative to the list oscillator pair, the group oscillators may be reset before they have completed one half-period. Alternative solutions to the problem of synchronisation might be worth considering. Finally, the assumptions behind the rhythmic parser need to be examined. For example, the current parser predicts that an input of *1 2  3* would be parsed as a group of two followed by a group of one, whereas an input of *1  2 3* would be parsed as a single group of three. This would be simple to test empirically, yet surprisingly, we know of no data on the effects of such temporal presentation schedules on grouping.

## 7 Summary

The problem of serial order illustrates both the importance of representations in psychological theorising and connectionist modelling, and the valuable interplay between these two fields. Since their inception as psychological theories, all three approaches to the problem, interitem associations, ordinal representations and positional representations, have subsequently been instantiated in various connectionist models. Though the data from STM experiments favour positional representations, it is only through the precision required of connectionist instantiations that the nature of positional representations, whether they are temporal, absolute or relative, has been questioned. These connectionist models are therefore now feeding back into psychological theorising to make predictions that can be tested empirically. As such, the problem of serial order exemplifies both the theme of this year's Workshop, and the rationale behind the Neural Computation and Psychology Series as a whole.


### Acknowledgements

Rik Henson was supported by BBSRC Grant Number SO5000. Neil Burgess was supported by a Royal Society University Research Fellowship. The authors can be emailed via r.henson@ucl.ac.uk and n.burgess@ucl.ac.uk.


### Notes

1. There are other possible solutions of course. With respect to repeated items, one can appeal to the type/token distinction, so that two occurrences of the same type have nonidentical token representations [38]. With respect to errors in recall, open-loop chaining models like TODAM [39] can cue with previous items whether or not they are recalled correctly (as opposed to closed-loop chaining models in which responses feedback to cue subsequent items [9]).

2. Neath and Crowder [34] found that slower presentation rates did aid recall when much faster rates were used (around five items per second, in comparison with the maximum of two items per second in the study of Baddeley & Lewis [32]). However, under such rapid presentation of items, there is a danger that faster presentation rates do not allow as effective encoding.

3. Numerous experiments have shown that measures of recency are positively correlated with the ratio of interitem interval to retention interval [40, 41], in support of temporal distinctiveness theories [42, 43]. However, these experiments have tested mainly recognition, or free recall, rather than memory for serial order (apart from those discussed in Section 5.1). Moreover, recency is not necessarily the best index of temporal distinctiveness because it is a relative measure, comparing performance on the last position with that on penultimate positions, and is therefore likely to depend on other factors, including overall performance level. Indeed, when overall performance levels are compared, greater temporal spacing does not always aid free recall [40] (though see [42]).

4. We thank Tom Hartley for this idea.

## References


[1] Jordan M I, Serial order: a parallel distributed approach (ICS Report 8604), San Diego: University of California, Institute for Cognitive Science, 1986.

[2] Grossberg S, Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? Journal of Mathematical Psychology, 17, 1978, pp 199-219.

[3] Houghton G, The problem of serial order: A neural network model of sequence learning and recall, in Dale R, Mellish, C & Zock M (Eds.), Current Research in Natural Language Generation, London: Academic Press, 1990, pp 287-319.

[4] Brown G D A, Preece T, & Hulme C, Oscillator-based memory for serial order, manuscript submitted for publication.

[5] Burgess N, & Hitch G, Toward a network model of the articulatory loop, Journal of Memory and Language, 31, 1992, pp 429-460.

[6] Henson R N A, Short-term memory for serial order: The Start-End Model, manuscript submitted for publication.

[7] Henson R N A, Positional information in short-term memory: relative or absolute?, manuscript submitted for publication.

[8] Lashley K S, The problem of serial order in behavior, in Jeffress L A (Ed.), Cerebral Mechanisms in Behavior: The Hixon Symposium, NY: John Wiley & Sons, Inc., 1951, pp 112-136.

[9] Henson R N A, Short-term memory for serial order, unpublished doctoral dissertation, University of Cambridge, England, 1996.

[10] Ebbinghaus H, Memory: a contribution to experimental psychology, New York: Dover, 1885/1964.

[11] Wickelgren W A, Short-term memory for phonemically similar lists, American Journal of Psychology, 78, 1965, pp 567-574.

[12] Slamecka N, Ebbinghaus: some associations, Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 1985, pp 414-435.

[13] Elman J L, Finding structure in time, Cognitive Science, 14, 1990, pp 179-211.

[14] Taylor J G, Neural network capacity for temporal sequence storage, International Journal of Neural Systems, 2, 1991, pp 47-54.

[15] Houghton G & Hartley T, Parallel models of serial behaviour: Lashley revisited, PSYCHE, 95-2-25, 1996.

[16] Estes W K, An associative basis for coding and organisation in memory, in Melton A W & Martin E (Eds.), Coding Processes in Human Memory, Washington DC: V. H. Winston & Sons, 1972, pp 161-190.

[17] Page M P A & Norris D G, The primacy model: a new model of immediate serial recall, manuscript submitted for publication.

[18] Conrad R, Order error in immediate recall of sequences, Journal of Verbal Learning and Verbal Behavior, 4, 1965, pp 161-169.

[19] Lee C L & Estes W K, Order and position in primary memory for letter strings, Journal of Verbal Learning and Verbal Memory, 16, 1977, pp 395-418.

[20] Lee C L & Estes W K, Item and order information in short-term memory: evidence for multilevel perturbation processes, Journal of Experimental Psychology: Human Learning and Memory, 7, 1981, pp 149-169.

[21] Burgess N & Hitch G, Memory for serial order: a network model of the phonological loop and its timing, manuscript submitted for publication.

[22] Burgess N & Hitch G, A connectionist model of STM for serial order, in Gathercole S (Ed.), Models of short-term memory, UK: Psychology Press, 1996, pp 51-71.

[23] Hitch G H, Burgess N, Towse J N & Culpin V, Temporal grouping effects in immediate recall: a working memory analysis, Quarterly Journal of Experimental Psychology, 49A, 1996, pp 116-139.

[24] Conrad R & Hull A J, Information, acoustic confusion and memory span, British Journal of Psychology, 55, 1964, pp 429-432.

[25] Henson R N A, Norris D G, Page M P A & Baddeley A D, Unchained memory: error patterns rule out chaining models of immediate serial recall, Quarterly Journal of Experimental Psychology, 49A, 1996, pp 80-115.

[26] Wickelgren W A, Associative intrusions in short-term recall, Journal of Experimental Psychology, 72, 1966, pp 853-858.

[27] Henson R N A, Item repetition in short-term memory: Ranschburg repeated, manuscript submitted for publication.

[28] Henson R N A, Mathematical analysis of a linear compound chaining model, unpublished manuscript.

[29] Ryan J, Grouping and short-term memory: different means and patterns of grouping, Quarterly Journal of Experimental Psychology, 21, 1969, pp 137-147.

[30] Wickelgren W A, Rehearsal grouping and hierarchical organization of serial position cues in short-term memory, Quarterly Journal of Experimental Psychology, 19, 1967, pp 97-102.

[31] Conrad R, Serial order intrusions in immediate memory, British Journal of Psychology, 51, 1960, pp 45-48.

[32] Baddeley A D & Lewis V J, When does rapid presentation enhance digit span?, Bulletin of the Psychonomic Society, 22, 1984, pp 403-405.

[33] Neath I & Crowder R G, Schedules of presentation and temporal distinctiveness in human memory, Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 1990, pp 316-327.

[34] Neath I & Crowder R G, Distinctiveness and very short-term serial position effects, Memory, 4, 1996, pp 225-242.

[35] Ng L H H, Are time-dependent oscillators responsible for temporal grouping effects in short-term memory?, unpublished Honours Thesis, University of Western Australia, 1996.

[36] Treisman M, Cook N, Naish P L N & McCrone J K, The internal clock: electroencephalographic evidence for oscillatory processes underlying time perception, Quarterly Journal of Experimental Psychology, 47A, 1994, pp 241-289.

[37] Hartley T, Syllabic phase: bottom-up representations of the structural properties of speech, manuscript in preparation.

[38] Wickelgren W A, Context-sensitive coding, associative memory, and serial order in (speech) behavior, Psychological Review, 76, 1969, pp 1-15.

[39] Lewandowsky S & Murdock B B Jr, Memory for serial order, Psychological Review, 96, 1989, pp 25-57.

[40] Bjork R A & Whitten W B, Recency-sensitive retrieval processes in long-term free recall, Cognitive Psychology, 6, 1974, pp 173-189.

[41] Neath I, Distinctiveness and serial position effects in recognition, Memory & Cognition, 21, 1993, pp 689-698.

[42] Glenberg A M & Swanson N G, A temporal distinctiveness theory of recency and modality effects, Journal of Experimental Psychology: Learning, Memory and Cognition, 12, 1986, pp 3-15.

[43] Marks A R & Crowder R G, Temporal distinctiveness and modality, Journal of Experimental Psychology: Learning, Memory & Cognition, 23, 1997, pp 164-180.