The Divergent-Reconvergent Model of Serial Order Encoding and Retrieval

Danke X. Shieh (dxie@crl.ucsd.edu)
Jeffrey L. Elman (elman@crl.ucsd.edu)
Department of Cognitive Science, University of California, San Diego
La Jolla, CA 92093

Abstract

This paper presents a new connectionist model for serial order encoding and retrieval. It is based on a divergent-reconvergent structure, which encodes a list of items by superimposed distributed representations of each of the items, where early items are coded by more active units than late ones due to lateral inhibition. Thus, the first item is the strongest and the last the weakest. Retrieval is based on the activation gradient: selecting the most active item, and then inhibiting it to allow the next item to be retrieved. Side effects of this mechanism gives rise to the primacy effect, the recency effect and basic similarity effects, similar to those found in human immediate serial recall. The divergent-reconvergent structure strikingly resembles the anatomical structure of the basal ganglia. This model provides a plausible computational and functional account for the sequencing function of the basal ganglia, and other brain areas that may implement it.

Introduction

In 1951, Karl Lashley identified the *serial-order* problem as fundamental to understanding the brain and cognition. The serial-order problem can be loosely defined as how behavioral sequences are produced. It underlies much human and animal behavior, ranging from locomotion, reaching, grasping to language and memory. In the field of memory, many theories have been proposed to account for the nature of serial order encoding and retrieval (see Henson, 1998 for a review).

One influential theory, the *ordinal theory*, assumes that order is represented by relative values of a continuous property (e.g., the activation level of the items) with the first item 'strongest' and the last item 'weakest'. The order of items is retrieved by iteratively selecting the most active item, with each retrieved item temporary suppressed. This mechanism is also known as *Competitive Queuing* (CQ, Houghton, 1990).

Models based on CQ can reproduce dominant patterns of order errors (omission, intrusion, transposition). A CQ system extended with rehearsal and multiple presentation modalities (visual and auditory) can account for an even larger amount of data (Page & Norris, 1998). The simplicity and explanatory power of CQ is appealing. However, CQ itself cannot explain phenomena where positional coding appears to be essential. For example, in the recall of a temporally grouped sequence: 4 5 2 – 3 9 7, if the middle item "5" of the 1st group is transposed with an item in the 2nd group, it is most likely to be the middle item "9" than others. To account for these psychological findings, positional information is considered to be required, though there are disputes concerning its nature (Page & Norris, 1998).

From current behavioral and modeling studies of working memory for serial order, it may best be assumed that CQ interacts with some positional or contextual coding mechanism. This view has been explicitly or implicitly adopted in most major models. Burgess and Hitch (1999) used fast synaptic changes to associate contextual code to items, where the item-level representations by themselves work on a CQ basis. Henson (1998)'s start-end model can be seen as integrating CQ in one of its positional signals (the start signal), because this signal is equivalent to an activation gradient, and it is used, among other information, to cue stored items.

Now that CQ has directly or indirectly played important roles in working memory models for serial order, how can the *activation gradient* that CQ works upon be generated in the brain? Many mechanisms have been proposed, but there has not been a consistent and neuropsychologically-based account.

The primacy model (Page & Norris, 1998) assumes that the primacy gradient "results from some association of each item with some representation of the start-of-list context". SARDNET (James & Miikkulainen, 1995) assumes that all items are activated at presentation to the same level and begin to decay, which makes the last item the 'strongest', and the first the 'weakest'. Burgess and Hitch's (1999) model assumes that each item is inhibited to a low level after presentation, but the inhibition is gradually removed, leaving the first item 'strongest' and the last item 'weakest', the opposite of SARDNET. This mechanism meets the requirement of CQ (starting from the most active item), but the indicated inhibition implies two conditions – occurring immediately after presentation, and lasting throughout recall, which are not yet found in neuropsychological studies. To support the CQ framework, other mechanisms may be needed.

How can the activation gradient be generated in a neuropsychologically plausible way? This paper presents a simple neural mechanism that has not been explored before. It is notable that the connectivity pattern in this model resembles that of the basal ganglia, and this pattern may also be available in the neocortex. It is hoped that this model will shed light on the computation and function of the basal ganglia, and other brain areas that may implement it.

The Divergent-Reconvergent Model

The Divergent-Reconvergent Model is a novel neural network model for serial encoding and retrieval. The goal of the current model is to show 1) a special connectivity pattern in the basal ganglia, the divergent-reconvergent structure (Graybiel, Aosaki, Flaherty, & Kimura, 1994), can give rise to serial order encoding and retrieval in a way that has not been explored before; 2) side effects of this structure induce essential characters of human working memory: the *serial order effects* and *similarity effects.*

The basic idea

The main input station of the basal ganglia, the striatum, collects inputs from the entire neocortex and sends processed information, through other parts of the basal ganglia, to the motor and frontal areas of the forebrain. It has puzzled investigators for years that there are millions of projection neurons in the primate striatum, but they project to a much smaller set of neurons in the basal ganglia output nuclei, the global pallidus (GP). Graybiel et al. used anatomical tracers on the basal ganglia pathway, and found a *divergent-reconvergent* connection structure. They proposed that this network works as "local experts" performing distinct computational tasks, and their results are selected by a gating network.

We propose a different function that the divergent-reconvergent structure may also perform – the sequencing of motor and cognitive elements. They may include hand and foot movements, percepts as perceived letters and words, and other cognitive entities as mental arithmetic operations and chess moves. Basal ganglia involvement in high-level cognition has also been argued for by other researchers (Graybiel et al., 1994; Lieberman, 2000).

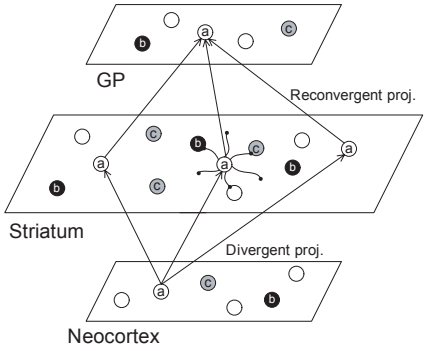

Figure 1. The divergent-reconvergent connectivity scheme

Figure 1 shows the divergent-reconvergent connectivity structure that can support sequence encoding and retrieval. For comparison purposes, the layers are labeled by their putative brain counterparts, but the diagram is not intended to show the fine anatomy of the brain.

Units in the neocortex represent input items. When an item is presented, it activates a set of units that are spatially distributed in the striatum. This set of units convergently activates an output unit in the GP. With this connectivity, the network remaps an input unit to a corresponding output unit, via a distributed internal representation. But, if recurrent loops and lateral inhibition are added to the striatal units, an interesting phenomenon appears: presented items are actively maintained in the striatum (discussed later), and early items inhibit a proportion of striatal units representing late items. This will make the distributed representation of late items "weaker" than that of early ones. In this way, an activation gradient can be produced in the striatum. The GP can then perform a competitive selection process to iteratively retrieve items sequenced in the striatum.

It is interesting that the structural and functional requirements in the above mechanism (lateral inhibition, competitive selection) are indeed available in the basal ganglia. The primary neurons in the striatum (the spiny neurons) are characterized by a dense and extensive local axon collateral field, and they have inhibitory effects on neighboring spiny neurons. The GP (more specifically, its internal segment, GPi) can function as an action selector, performing a kind of winner-take-all function, or as some researchers prefer, a "loser-take-all" function, given its inhibitory output which in turn disinhibits a selected action (Berns & Sejnowski, 1998).

Though the striatal spiny neurons do not have local recurrent loops, their activity may be maintained in indirect ways: 1) there are loops between the basal ganglia and the prefrontal cortex (PFC) (Beiser & Houk, 1998); 2) the striatum may constantly receive input from the PFC, which maintains activation by its loops with other parts of the brain.

It can be noted that the divergent-reconvergent connection pattern may also be available in the neocortex, where sequential information needs to be processed (e.g., auditory perception). In this model, we focus on the basal ganglia because of its privileged role in sequencing. This serial order mechanism, either implemented in the basal ganglia or the neocortex, can be a plausible mechanism for the brain.

Architecture

Figure 2 shows the structure of the proposed model involving the PFC, basal ganglia and the motor cortex. The Neocortex-Striatum-Motor pathway embodies the divergent-reconvergent structure illustrated above. The winner-take-all function of the GP and its subsequent triggering of an action in the motor cortex by way of the thalamus are simplified as a single striatomotor *reconvergent* projection. The Neocortex-PFC pathway actively maintains the items (without order information in the current model). The PFC provides additional item representations, which are inhibited when the items are retrieved. This kind of item-level inhibition is also often assumed in other working memory models of serial order (e.g., Burgess & Hitch, 1999).

Layers in this model are described below. In addition, there is a single *recall inhibitor* unit that inhibits the *motor* layer during item presentation. The signs + / − on a projection designate excitatory and inhibitory, respectively.

Neocortex layer. This is the input layer of the system. It has n_f units, each corresponding to an input feature. An item can have multiple features. When it is presented, its feature units are activated for 1 time step. This layer projects to the *striatum* layer (via the *divergent* projection) and to the PFC via an all-to-all projection. In the *divergent* projection, each

neocortex unit connects to the same number of units that are randomly selected from the *striatum* layer. For simplicity, *striatum* units receiving input from different *neocortex* units do not overlap. If overlapping is allowed, the system will work under a bit more interference.

PFC layer. This layer actively maintains item-level representations. Its units should be at least as many as items. This layer is trained with unsupervised learning to recognize items presented to the *neocortex* layer. After learning, each item should activate a unique unit in this layer. Units that do not respond to any item will remain inactive. The activation function is sigmoid,

$$f(x) = \frac{1}{1 + e^{-(ax+b)}}, \qquad (1)$$

where x is the net input, and parameters a = 7, b = -6. Items are actively maintained by self-recurrent connections (weight = 2). The dynamics is *bi-stable*, in the sense that a strong input (>0.8) can stimulate a unit to a stable activation level close to 1, and it will stay active until a strong inhibition (<-1.3) resets it to 0. Spontaneous decay has not been modeled but can be easily done with a leaky-integrator neural model.

Striatum layer. This layer is the substrate for order encoding. It has n_s units. The main input source is the *neocortex* via the *divergent* projection. Units in this layer are randomly divided into subsets of the same size (n_s / n_f), each receiving input from a *neocortex* unit (weight = 1). This layer uses the same neural model as (1). The layer size can be chosen fairly freely, as long as there are enough units to produce fine-grained activation gradients. Another input source is the PFC, from where actively maintained items continuously activate their corresponding striatal units (weight = 1).

Each unit in this layer exerts strong lateral inhibition (weight = -2) on 4 neighboring units in the cardinal directions, except for the edges, where there are fewer neighbors. When a unit is active, it dominates and completely inhibits those neighboring units from firing. If a neighbor unit receives input from the same *neocortex* feature as this unit, the inhibition will eventually be removed by Hebbian learning since they will frequently fire together. This reduces undesired variations in the activation gradient. Other numbers of inhibited neighbors can also be used.

Motor output layer. This is the output layer of the system. As the PFC layer, it should have more available units than items, and is to be trained by unsupervised learning to respond to the items. The actual neural pathway from the striatum to the motor cortex is a complicated chain of inhibitory and gating processes through the GP and thalamus (Berns & Sejnowski, 1998). It is summarized here by a single winner-take-all projection. After training, item-specific units will summate striatal activation for the items. Units whose net input is above the threshold (0.2) can compete for output. The strongest unit will be reported, and it will inhibit an item representation in the PFC, which, in turn, deactivates the item in the striatum, allowing the next item to be retrieved.

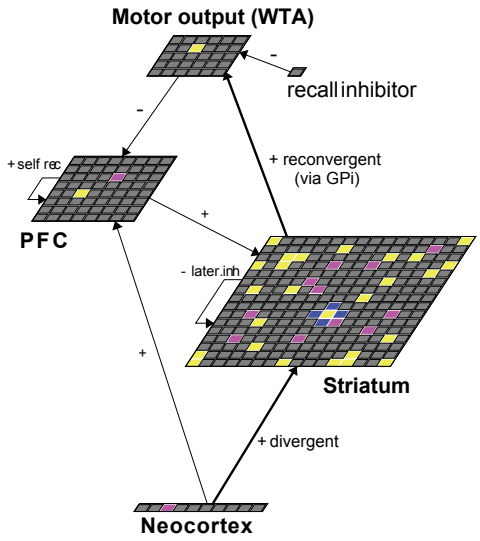

Figure 2. The Divergent-Reconvergent Model

Training

Two projections in the network are fixed. They are the divergent projection from the *neocortex* to the *striatum*, and the self-recurrent connections in the PFC. Learning will apply to other projections. At initialization, each striatum unit is given strong inhibitory weights on 4 neighboring units, and small random weights are assigned to other modifiable projections.

During training, the *recall inhibitor* is set *inactive*, so that the network can produce output at any time. Each item is presented to the *neocortex* layer in turn. Before the next item is presented, network activity is reset to 0. The network only receives individual items but not sequences in training.

After an item is presented, competitive learning (Rumelhart & Zipser, 1985) is applied to two layers: the *motor output* layer and the PFC layer. Units in these layers will learn to recognize patterns in their input layers. For example, after an item is presented, it produces an activation pattern in the striatum layer. *Motor* units will learn this pattern. The most strongly activated unit "wins" and its weight vector w is adapted to the pattern vector x,

$$w = w + \eta(x / \Sigma_i x_i - w), \qquad (2)$$

where η = 0.1 is the learning rate. The pattern x is normalized to sum to 1 by the factor $\Sigma_i x_i$. This is helpful because otherwise patterns with more features will simply be more active and output early. Although it may be a valid phenomenon that a salient item (supposedly highly active) is recalled first by humans, this paper does not intend to model this phenomenon. Units in the PFC are similarly trained to learn the item patterns in the *neocortex* layer.

Hebbian learning is applied to adapt the *lateral inhibition* between *striatum* units. As mentioned above, initial inhibition between co-active units will be eventually removed. Hebbian learning is also applied to the projection from the PFC to the *striatum* layer (max weight = 1). This will associate items in the PFC with their distributed representations

in the *striatum.* This association is important for the maintenance of the striatal activity. Anti-hebbian learning is applied to the projection from the *motor output* layer to the PFC layer (min weight = -2), so that a motor response will inhibit its corresponding item unit in the PFC.

The model is trained with 40 passes of all items. At the end of training, every item should reliably activate one PFC unit and one motor unit. It is notable that although the network is not trained to encode and retrieve sequences explicitly, this ability has emerged from its structure and what it has learned about the items.

Simulations

Simulation 1 - Serial recall

In this simulation[1], we test the network's basic ability to encode and retrieve sequences. Recall accuracy and serial position effects will be investigated to assess its validity as a model for human working memory. During the testing phase, no further learning occurs. As noted above, the serial recall effects are a consequence of the network's architecture, once it has learned to recognize items in isolation.

The materials are lists of 3 to 8 items, randomly drawn from a family of 20 items without repetition (the total number of items is important for this model, discussed later). For each length, 800 lists are used. In this simulation, each item is represented by a single feature. Items containing multiple features will be studied in Simulation 2.

A list is presented and retrieved as follows. First, set the *recall inhibitor* unit active, so that the network will store the sequence without generating any output. Then, the items are presented to the *neocortex* layer, each for 1 time step followed by an interval of 5 time steps. After a list has been fully presented, the *recall inhibitor* is deactivated to start a recall. Every time a response is generated, the *recall inhibitor* is temporarily set *active* for 5 time steps. No output is produced in this interval, allowing the network to stabilize after an item is retrieved. The recall process ends when the network doesn't generate any output in 15 time steps.

While items are presented, an activation gradient is formed in the striatum (Figure 3, Plots 1-3). It is interesting that after an item is inhibited after retrieval, some neighboring units that it has inhibited are now liberated. Items that make use of these liberated units will see an increased activation due to the input from PFC (Plots 4-5). This is a special property of the system: after an early item is retrieved, late items can fill up the gap it leaves, allowing the network to add new items to the end of the list. Thus, this model can elegantly handle continuous input and output. It is immune to some criticisms against early buffer theories of human memory, which need to be emptied before new items can be stored.

[1] The network sizes are neocortex layer $n_f = 20$, striatum layer $n_s = 3600$ (60×60), PFC layer 400 (20×20) and motor output layer 400 (20×20). Gaussian noise ($\sigma = .05$) is injected into the motor output layer.

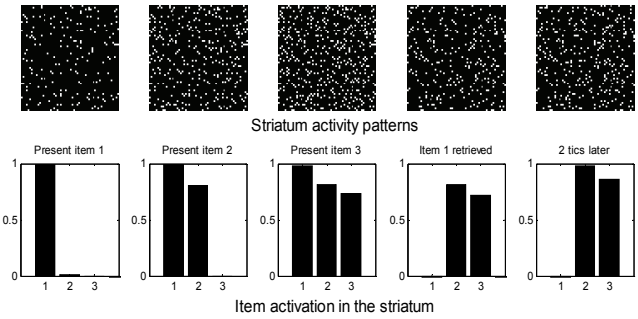

Figure 3. Network activity in sequence encoding and retrieval

In human immediate serial recall, three types of errors are commonly found: omission, transposition and intrusion. Omissions are very rare in this simulation since decay has not been modeled. Intrusions are also rare, because the only input to the model is the current list, and there is no interference from long-term memory or other input sources. Most errors are transposition errors, caused by the noise injected to the recall process.

Recall accuracy is plotted as a function of list lengths, and serial positions (Figure 4). The result shows clear list length, primacy and recency effects. Long lists are recalled worse than short ones, because 1) there are more items competing for output at each point, 2) exponentially fewer units represent each additional item, because of the added-up lateral inhibition from all previous items. The memory span is approximately a log function of the memory resource, and is considerably stable in face of variances of the latter.

The *primacy* effect occurs because activation levels are more distinct among early items than late ones. This is a result of the exponential-like activation gradient. The *recency* effect occurs in this model for two reasons. First, no confusion can happen at the last position, since all other items have been retrieved. This explanation is also given in some other models (Page & Norris, 1998). Second, the activation of late items grows up after the removal of early items. This also contributes to the recency effect, since the increased activation makes late items more distinct. The second factor rests on the assumption that retrieved items are inhibited from the memory, or have decayed for they are no longer needed, as is found in cell recordings in monkey's delayed match-to-sample tasks (Fuster & Jervey, 1982). It is valid only to the extent that this assumption is true. Other factors that may contribute to the recency effect include stronger activation of recent items due to less decay (Davelaar & Usher, 2003), or additional memory components, such as the precategorical acoustic store (PAS, Crowder & Morton, 1969).

Discussion

The total number of items needs to be chosen with some care, because it indirectly determines how many *striatum* units are allocated for each item. If there are too few items, too many striatum units may be allocated for each, and the lateral inhibition may be so strong that the network has too little a memory span. If there are too many items, the stria-

tum units allocated for each may be too sparse, and the weak lateral inhibition may not generate good activation gradients. In this paper, the total number of items is always chosen to balance this trade-off. The number or density of striatum units for an item or feature may be defined as an independent parameter in the future.

Since the network parameters have not been optimized to fit human data, quantitative comparison is not yet available. However, qualitative results have provided initial support for the proposed model.

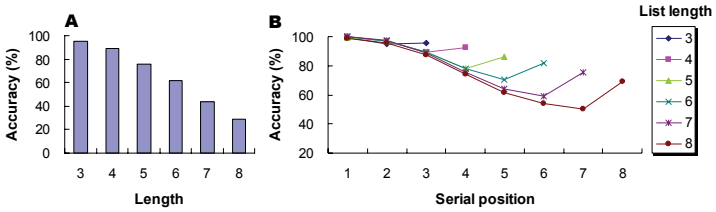

Figure 4. List length and serial-position effects

Simulation 2 - Similarity effects

Typical working memory models based on CQ adopt localist representation for the items. Since localist representations do not represent item similarity by themselves, many models have resorted to a two-stage process to account for similarity effects: first, selecting a correct item, and then, passing it to a confusion process which probabilistically produces erroneous output items, according to their similarity to the item selected in the first stage (Henson, 1998; Page & Norris, 1998). The model proposed in this paper does not require two separate stages to account for similarity effects. Rather, it naturally allows interaction between similar items based on the neural network's ability to handle distributed representations.

Specifically, two hypotheses are to be tested. First, are lists with confusable items more difficult to retrieve? Second, is the current architecture adequate to account for the following phenomenon? In lists containing a pair of phonologically confusable items (e.g., BHPYX), this pair of confusable items are likely to be transposed with each other, while other items are kept in their correct positions.

In this simulation[2], a new set of 20 items are used, each containing 5 features. Each of the first 16 items contains 5 completely distinct features (Items 1-16). In addition, there are two pairs of similar items: a pair sharing 1 feature (labeled A and B), and a pair sharing 2 features (labeled X and Y). This item set requires 97 different features.

The test material is 3 types of experimental lists. The Dissim type contains 5 distinct items. The Sim1 and Sim2 types contain 3 distinct items plus either of the above two pair of similar items, respectively. For each type, the total number of lists is 800.

[2] The network sizes are, neocortex layer $n_f = 97$, striatum layer $n_s = 6400$ (80×80), PFC layer 20 and motor output layer 20. Gaussian noise ($\sigma = .05$) is injected into the motor output layer.

Some modification has been made to the network for this simulation. In Simulation 1, competitive learning is used to train the motor and PFC layers. That was to show that the system can work in an unsupervised fashion. However, competitive learning requires many more units than the number of patterns. To simplify the system and reduce computational costs, in this simulation, only 20 motor and 20 PFC units are used, and each of them is designated to be the "winner" of one item. Although the designation of output units is similar to supervised learning, this is only to simulate the effect of having a large number of units in unsupervised learning.

The result (Figure 5A) shows the serial position curves of the 3 types of lists. A clear similarity effect can be observed: lists with higher confusability induce more errors. The effect is produced because the activation levels of items with shared features are less distinctive than those without shared features (shared features cannot contribute a difference to the items' activation levels). This basic similarity effect is consistent with the fact that people have greater difficulties recalling phonologically confusable lists than non-confusable ones (Henson, 1998). Now, the first hypothesis to be tested has been shown to be true.

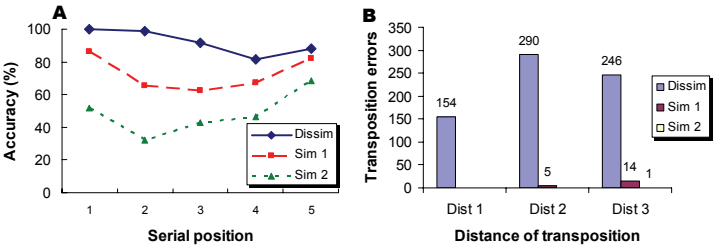

Figure 5. Similarity effects on accuracy and transposition error distributions

Next, transposition errors are studied in more details. In human immediate serial recall, transposition errors tend to be local, occurring between items that are close to each other in the list. They are explained by some models as having similar position coding (Burgess & Hitch, 1999; Henson, 1998). This characteristic has also been found in this model (Figure 5B), especially in the Sim2 condition, where a clear transposition gradient can be seen: more transposition errors occur between close items than distant ones. The locality of transposition errors occurs because when an item is not produced at its right position, it will tend to be produced in the next position, because it has higher activation than later items. The same property is true of other CQ models (Page & Norris, 1998).

An important question is, are the pair of confusable items more likely to transpose with each other than with other items? In human experiments, this phenomenon has been confirmed. Phonologically confusable items tend to transpose with each other, while keeping non-confusables items in the right positions. Unfortunately, this property has not been found in this model. When a pair of confusable items is included, their transposition did occur. For example, sometimes "Y 6 1 X 5" is recalled as "X 6 1 Y 5". But it is accompanied by increased transposition errors between con-

fusable item and non-confusable items. The sawtooth-like accuracy curve for alternating phonologically confusable and non-confusable items have not been found either, in a separate simulation.

Discussion

What does the result suggest? Does it mean that the model is wrong? Here we provide two analyses. First, the serial order model is partially motivated by the neural anatomy of the basal ganglia. As discussed before, this mechanism may apply to the sequencing of many types of elements: motor, perceptual, cognitive and linguistic. Though it is true that phonologically encoded lists exhibit an obvious sawtooth-like serial position curve if the lists contain alternate confusable and non-confusable items, this curve is generally not obvious in other modalities like vision. Thus, the model, which is based on the basal ganglia, may correctly not exhibit this effect. Second, further analysis of the difficulty from similarity reveals the same issue as repeated items – repeated activation of a feature does not induce a new representation. This is an intrinsic difficulty for CQ and arguably the brain. The brain's solution to the repetition problem appears to be non-trivial (consider the binding error of repeated items, such as 233 recalled as 223, or 858 as 585), and this model may provide a starting point for studying the brain's solution.

In a more advanced simulation, we have confirmed that if additional phonological information is used to cue items in this model, the effects found with alternate phonologically confusable and non-confusable items will emerge.

General discussion

This paper presents a novel computational mechanism (the divergent-reconvergent scheme) for encoding and retrieving sequences. It has been demonstrated with computer simulations that the Divergent-Reconvergent Model indeed possesses the capability. The side effects of the mechanism give rise to the list length effects, serial position effects and basic similarity effects. More importantly, it provides a viable account for the function of the basal ganglia, which has puzzled researchers for decades.

It is notable that the model has not been explicitly trained to encode and retrieve sequences. Instead, it is only exposed to individual items in isolation. Its ability to recall sequences emerges from the architecture of the model. This model has obvious psychological advantages: humans have little difficulty recalling sequences after being familiarized with the items. The amount of training appears to be right.

It might appear that this model requires too rich lateral inhibition among all items in the brain, which is problematic, given that mental objects of different modalities are located in different areas of the brain, and there isn't comprehensive lateral inhibition between these areas. However, this is exactly the reason why the brain's sequencing function may resort to a centralized mechanism. In our opinion, this model is best matched with the computation going on with one module of the striatum (Graybiel et al., 1994). The lack of lateral inhibition across modules (e.g., foot and hand) can explain very well why mixed foot and hand movements are harder to sequence than foot or hand movements alone.

The Divergent-Reconvergent Model provides a neurologically viable account for the brain's mechanism for serial order. However, it has not addressed many important issues (e.g., recall of non-words, temporal grouping and item repetitions). They will be directions for future research.

Acknowledgement

We would like to thank Hal Pashler, Marty Sereno, Gary Cottrell, Jochen Triesch, Marta Kutas, and Ginny de Sa for insightful and supportive comments on this research, and three anonymous reviewers for their expert feedback that greatly improved this paper. Danke owes inspiration to the Geisel Library and PSY201B. This research was supported in part by grant NIH MH60517 to the second author.

References

Beiser, D. G., & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology, 79*, 3168-3188.

Berns, G. S., & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience, 10*, 108-121.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106 (3)*, 551-581.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (pas). *Perception & Psychophysics, 5*(6), 365-373.

Davelaar, E. J., & Usher, M. (2003). *A neurocomputational account of short- and long-term recency effects.* Paper presented at the CogSci 2003.

Fuster, J. M., & Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience, 2*, 361-375.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science, 265*(5180), 1826-1831.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73-137.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock (Eds.), *Current research in natural language generation*. London: Academic Press.

James, D. L., & Miikkulainen, R. (1995). Sardnet: A self-organizing feature map for sequences. In G. Tesauro, D. S. Touretzky & T. K. Leen (Eds.), *Advances in neural processing systems* (Vol. 7).

Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Cambridge: Mass: Harvard University Press.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105 (4)*, 761-781.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9*, 75-112.