A connectionist model of attentional enhancement and signal buffering.

Judith M. Shedden and Walter Schneider
University of Pittsburgh

Abstract. The connectionist/control simulation of attentional enhancement, signal maintenance, and buffering of information is described. The system implements a hybrid connectionist architecture incorporating auto-association in the hidden layer and gain control on the hidden and output layer. The structure of the model parallels major features of modular cortical structure. The attentional selection simulations show that as one channel is attenuated, the system exhibits attentional capture in which only the more intense stimulus is transmitted to higher levels. The signal maintenance simulations show that small levels of auto-associative feedback can faithfully maintain short bursts of input for extended periods of time. With high auto-associative feedback, one module can buffer information from a previous transmission while the module blocks the interference resulting from concurrent transmissions. The combination of auto-associative feedback and gain control allow extensive control of information flow in a modular connectionist architecture.

This paper examines signal control issues in a connectionist processing system. It examines three basic cognitive operations of **attentional enhancement**, **signal maintenance**, and **buffering** of information. In human processing these operations represent control functions that are often associated with mechanisms of attention. Attention is the selection for processing of some subset of available information in the environment. For example, when humans switch attention from one item to another there is an attentional capture effect where they perceive one signal to the almost total exclusion of another signal (e.g., Treisman & Riley, 1969). Information that is briefly presented is maintained in a short-term sensory buffer that is available for a substantial period after the sensory stimuli are removed (e.g., Sperling 1960). Humans can also buffer information in short-term memory while encoding and acting on new information (Klapp, Marshburn, & Lester, 1983).

Traditional Models of Attention. The mechanism of attentional focus has been a topic of interest to psychologists for years (see Shiffrin, 1988). Broadbent's (1957, 1958) theory of selective attention suggests a filter mechanism, in which information from all channels is initially processed in parallel, but at some point in the system information converges on a limited capacity channel. The model originally stated that selection is all-or-none, with no information being passed from other unattended channels. This view was modified when experiments showed that unattended information is available under some circumstances. Treisman's (1960) attenuation model is a modification of the filter model. It states that both the attended and unattended channels receive processing, but the processing in the attended channel is complete, and the processing in the unattended channel is 'attenuated' to some degree. Later models have combined selective processing with limited parallel processing. The Shiffrin and Schneider (1977) model assumes there is a parallel processing of multiple channels of consistent well learned information, and serial controlled processing of novel or inconsistent information. All of these models assume some mechanism of attentional enhancement. From this perspective, the mechanism for enhancement must still be determined.

Connectionist Models of Attention. Recently there have been several connectionist models of attention. The present models simulate either the selection issue or enhancement issue. The Koch and Ullman (1985) model involves a winner-take-all hierarchical model to indicate which pathway is to be selected. The Mozer (1988, and Mozer & Behrmann, 1989) model involves a topographic multidimensional map, with attention represented as a set of units that gate the flow of activity from lower levels in proportion to the strength of the attentional map. The Cohen, Dunbar and McClelland (in press) model examines the effect of the enhancement of the attended channel via control of the resting levels of the attended and unattended pathways. The model is applied to performance in the Stroop task. Our current simulation examines the enhancement effect under the assumption that attention involves changing control parameters of a connectionist module. The selection issue is not dealt

Shedden, J. M., & Schneider, W. (1990). In Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, Cambridge (pp. 566-573). Hillsdale, NJ: Lawrence Erlbaum.

with in this paper and will be addressed in future papers.

Physiology of Attention. The idea of attenuation, or inhibition of an unattended, competing signal is supported by physiological findings. Moran and Desimone (1985) measured the effects of attention on single cells in primate extrastriate visual cortex (area V4). Attention influences the output rate of V4 neurons as measured with microelectrode recordings. The post-stimulus time histograms show a strong attentional effect (Moran & Desimone,1985 and personal communications Desimone). The effect of selective attention is the attenuation of the signal from the unattended stimuli, and not an enhancement of the signal from the attended stimuli. This attenuation is a lateral inhibition within the receptive field of the responding cell, and the reduction in response of the unattended cells is to 1/3 of the attended state. Thus, there is roughly a 3:1 ratio of attended/unattended signal. This ratio is an important one from a modeling standpoint, because it provides a physiological bench mark against which to test the performance of the model.

Cortical neuroanatomy provides connection patterns and unit types that can be a basis for models of attentional functioning. Most of cortex post the first sensory areas (e.g., post V1 in vision) shows a similar layering of cells and cell types. Studies of V2 cortex identify cortical connections and components (Lund, Hendrickson, Ogren, & Tobi, 1981). Cortical processing appears to occur in a modular structure of columns (Mountcastle 1979). The forward information flow passes through two layers of pyramidal cells and inputs into layer 4, then to layer 2-3 pyramidal cells, and then out to the next module (see Lund et. al. 1981). The layer 4 cells connect recurrently to themselves providing a feedback path. In addition there are inhibitory interneurons that are primarily within a layer. There is a special class of axon-axon inhibitory cells called chandelier cells that attenuate large numbers of output cells (see Peters, 1984). This cell can inhibit the output of sets of pyramidal cells. The fast large scale inhibitory effects of chandelier cells make them a good candidate for the attenuation effects seen by Moran and Desimone (1985).

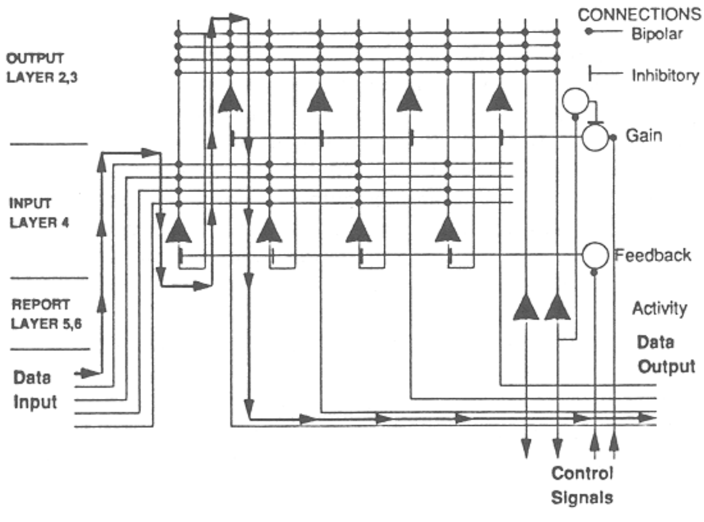

FIGURE 1: A connectionist model of cortical processing. It parallels the cortical processing in that it is a two layer structure (input layer 4 and output layer 2,3). Input feedback and output are controlled by two inhibitory modulation units. Control signals are provided by report units (layer 5,6) transmitting information regarding the activity and priority of the input to more central control structures and influencing the local gating of layer 2,3 activity. These control structures are assumed to influence the feedback and gain signals within a module.

CAP2 Architecture. The CAP2 architecture is a computer simulation environment designed to implement a modular connectionist architecture incorporating the major features of cortical processing to predict human attentional effects. The model consists of a variable number of modules, units, layers, and control elements, which can be combined to create several different connectionist architectural configurations. A module consists of an input layer of units, an associative matrix, and an output layer of units (see Figure 1). Modules can be added to the system in breadth, so that several input modules connect to one output module. In addition, modules can be added to the system in depth, creating several hierarchical layers.

Control elements are an important part of the CAP2 environment, and include internal feedback control, output gain control, and module activity and priority reports (see Schneider & Detweiler, 1987, for a discussion of the activity and priority report). These control elements are scaler values derived from the activity of the module itself, or assigned from outside the module, and act to modulate attentional processing. Feedback control increases or decreases the strength of the signal already in the system. Gain control increases or decreases the strength of the output signal.

$$\text{net input} = (\text{feedback} * \text{internal input}) + (\text{gain} * \text{external input})$$

The model includes standard connectionist layers with additional association and control effects. The simulation described in this paper consists of two modules connected hierarchically. Each module is a connectionist network consisting of three layers, including the **data input**, the module **input layer** (standard hidden layer in back propagation), and the **output layer**. Figure 1 shows the layers and connections. Each layer has 50 units, and each unit in one layer is connected by a set of weights to every unit in the adjacent layers. The modules can be cascaded so the output layer of one module is the data input of the next. Learning of input output patterns is accomplished via back propagation (see Rumelhart, Hinton, & Williams, 1986).

The system differs from a standard three layer network in that the hidden layer is connected through an auto-associative matrix to itself (see connections for Layer 4 cells to themselves as well as layer 2-3 cells). The auto-associative matrix is taught to reproduce the hidden layer using delta-rule learning (see McClelland & Rumelhart, 1988).

The system includes control elements that modulate the output from one layer to the next. These control elements are connected in a manner similar to the connections of cortical chandelier cells (e.g., one unit providing a scaler reduction of the population of units to which it is connected). The two control units determine **feedback** and **gain** (right of Figure 1). Manipulation of the feedback control element affects the strength of the signal through the auto-association matrix. This allows the system to hold a signal after the external input is turned off. The signal can be perpetuated by cycling through the auto-associative matrix. Control of the gain cell limits the output of the information from one module to the input of the next. The model also includes report cells used for determining where to switch attention and automatic processing within a module (see Schneider 1985, Schneider & Detweiler 1987). These however are not used in the current model.

Learning. The vector space for the simulation consists of ten input vectors and ten target vectors. Activation levels have maximum values of 1.0, minimum values of -1.0, and a resting activation of 0.0. Input vectors are of length 50, with initial activations set randomly to 1.0 or -1.0, and then clipped to 0.9 or -0.9 respectively for each unit. Target vectors are also of length 50, with initial activations set randomly to 1.0 or -1.0 for each unit. The set of all input and target vectors are forced to have correlations below 0.15 with all other vectors in the set.

Input vectors and target vectors are paired, and training of the network consists of back propagation of error (Rumelhart, et. al. 1986) after presentation of each input pattern. The activation function on the output is a logistic:

$$\text{activation} = \frac{1}{1 + e^{-\text{netinput}}}$$

The weights of the autoassociation matrix are changed using delta rule learning (see McClelland & Rumelhart, 1988), with the hidden layer vector as the input and the target. One epoch is defined as the presentation of each input pattern once, although the presentation order is randomized for each epoch. The network is trained for 15 epochs, at which point the system has learned to a criterion of 100% accuracy and correlations between output and target vectors are above 0.98 for each input pattern.

The network learned 10, 50, and 100 vector pairs of 50 units each in 2, 7 and 13 epochs, perfectly choosing the best matching response for the input. The correlation between output and target vectors climbs more slowly reaching correlations of .98, .73, and .55 respectively. This performance illustrates the large potential storage capacity of the model, which is important when relating model performance to brain performance. One cortical hypercolumn has been estimated to contain tens of thousands of cells (Mountcastle, 1979), suggesting that information is coded not as a single unit being on, but rather vectors in which a subset of units are turned on. We have worked with 10, 25, 50 and 200 unit vectors and find the performance of the model works well with large vectors. We use 50 element vectors that allow coding of large numbers of vectors with most vectors showing little correlation. The 50 element vectors do not require the long simulation times of very large vectors. All further testing discussed below is done with input and target vector sets of ten vectors each, after learning is completed.

Attentional Enhancement and Attentional Capture. A basic characteristic of attentional switching in humans is the attentional capture effect. As one moves attention from one stimulus to another, there is a sudden change from perception of the first stimulus to perception of the second. One does not see a mixture of the two stimuli. This is illustrated by observing a Necker cube. One perceives it in one orientation or the other, and this perception may switch between orientations, but one does not see a gradual shift or a combination of the two orientations. This effect is one of the characteristics investigated in this model.

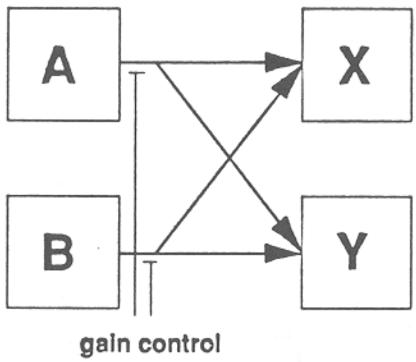

FIGURE 2: Diagram of attentional control between modules. The connections between modules represent vectors. The gain control is a scaler multiplication of all the elements of the vector. Each module connects to and receives from multiple modules. Attentional enhancement of A involves having a higher gain on the A module relative to the B module. Buffering involves loading a message from A to X and storing the message in X while a second message is sent from B to Y. Since B is connected to both X and Y the X module must hold the A message during the transmission from B (see text).

This simulation examines what is perceived when the relative signal strength of the attended and unattended message is altered. Does the receiving module get a clear signal of one of the messages? At what relative strength are the messages clear? Figure 2 illustrates the basic structure of the model. The A and B signals are input to the X module. As the relative strength of the two inputs changes, the X module settles on either the A or B signal. The two input vectors are presented to the network simultaneously for eleven testing trials. On each trial the gain (strength) of each vector is increased or decreased by 0.1, so that the total strength of both vectors is always 1.0. This is done for all combinations of two input vectors. Note the attentional capture tests were done with learning turned off, using the auto-associative and normal associative matrices developed in learning the 10 vector patterns.

The relative gains were:

Trial:	1	2	3	4	5	6	7	8	9	10	11
Vec A Gain:	1.	.9	.8	.7	.6	.5	.4	.3	.2	.1	.0
Vec B Gain:	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.

The question being asked is what vector or combination of vectors will be received by the output module? Accuracy is determined as the best match between output and target. Thus, if testing vector A, the trial is correct if the output has a higher correlation with target A than with any of the other possible target vectors. The left panel of Figure 3 shows the accuracy measure and correlations for the averaged data over all input pairs, and shows that at a .7:.3 ratio, all vectors are accurately identified without contamination from the interfering vector. Of more interest is an analysis of individual vector pairs, an example of which is shown in the right panel of Figure 3. Note that with a 0.1 change in relative strength there is a complete shift from perfect recognition of the A vector to perfect recognition of the B vector. In every case a difference of 0.1 in vector activation is enough to cause a sharp transition between the perception of vector A or vector B. There is variability between vector pairs for the position of this transition point, and that accounts for the wider crossover in the averaged data (Figure 3, left panel).

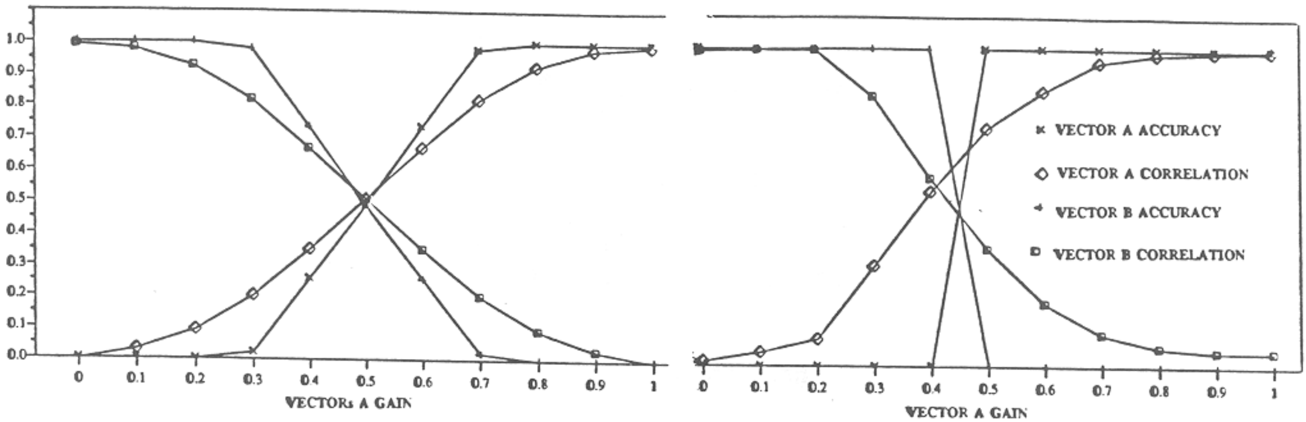

FIGURE 3: Attentional Capture Effects. The left panel shows the averaged data for 45 vector pairs, each pair presented simultaneously to the network. Total gain is always equal to 1.0. For example, if the gain of vectorA = 0.7, then the gain of vectorB = 0.3. Shown are the best match accuracy data and the correlation data. The right panel shows the data for a single vector pair. VectorA and vectorB are presented simultaneously. Best match accuracy and correlation with matching target vector are shown for each vector at each level of gain.

The simulation shows a clear signal capture effect with only the stronger vector being perceived at a strength ratio of 7 to 3. Recall that the attentional data from Moran and Desimone (1985) shows a required difference of 3:1 for the attended/unattended ratio, a point at which the current network provides clear capture. Note in this model that although the attention effect involves gradual attenuation, the network interactions produce an all or none switch to the enhanced signal. All of these tests were with vectors with low correlations; we expect to see some mixture effects with more highly correlated vectors.

Signal Maintenance and Normalization. Much of perception involves the input of brief bursts of information that can then be read out over a period of time. A brief visual stimulus can be read out of iconic storage for over a half a second (e.g., Sperling, 1960). One can fixate a stimulus briefly and then move one's eyes and still recall the initial stimulus. The auto-associative feedback common in cortex provides a mechanism for maintaining short duration signals. In this simulation this is implemented in the layer 4 auto-associative connections. Feedback in this system provides a way to latch and hold a signal for some extended period of time after the external input has been turned off. If the stimulus burst is very short, feedback can act to boost the signal and thereby normalize the input.

Input vectors are presented to the network for 5 iterations, and then the stimulus is turned off for another 5 iterations. An iteration consists of one pass through the system, from presentation of the input vector to production of the output vector. Five different levels of feedback are tested. The left panel of Figure 4 shows the activity of the hidden layer of units, and illustrates that the signal drops out quickly with a feedback value of 0.0 (maximum possible activity equals 1.0). As feedback increases, activity also increases. At a feedback value of 0.1 the module holds activity stable, accurately maintaining the vector activity pattern in the module. At feedback levels higher than 0.1 the activity continues to increase, but the signal evoked begins to stray from the desired target, and the system begins to 'hallucinate,' or identify stimuli which are not there.

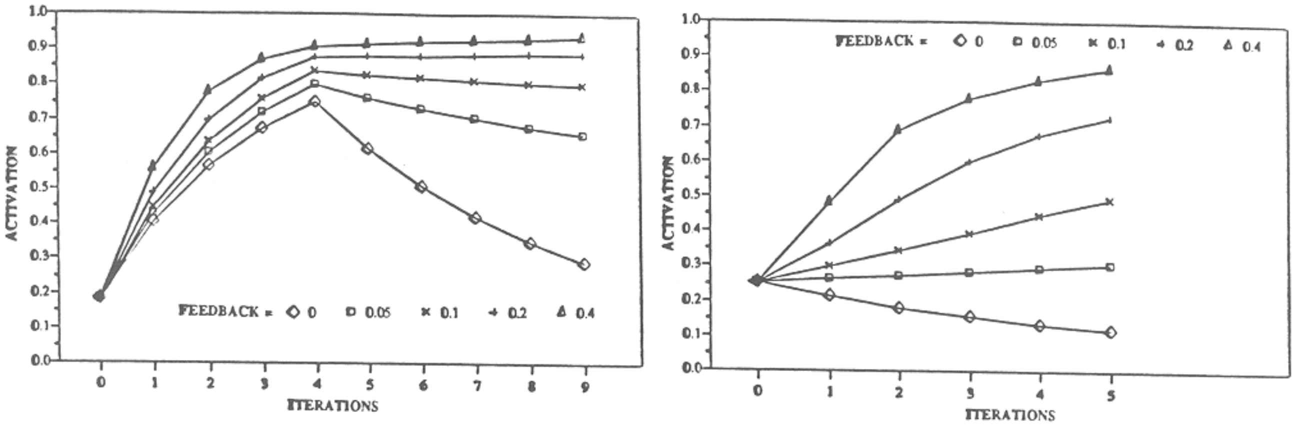

FIGURE 4: Latching and holding a signal. Left panel data are averaged over all stimulus vectors. A stimulus vector is presented for 5 iterations and then removed. Activation levels of the hidden layer are shown at different levels of feedback. At feedback of 0.0 the signal drops out quickly. At feedback of 0.1 the signal holds. At higher levels of feedback activity increases, but the accuracy of the signal begins to deteriorate. Right panel illustrates stimuli presented for only one iteration. Activation levels of the hidden layer are shown for different levels of feedback. At feedback of 0.0 the signal does not hold, but at higher levels of feedback activity increases, approaching normal levels by 4 or 5 iterations.

Feedback also allows the network to latch and increase the activity of a short burst stimulus. Input vectors are presented to the system for 1, 2, 3, or 4 iterations, and then the stimulus is turned off for 5 iterations. As can be seen in the right panel of Figure 4 which illustrates a single iteration stimulus burst, without feedback the signal dies quickly away. At higher levels of feedback the signal is latched and the activity level increases, approaching the strength of a normal signal after 4 or 5 iterations.

Signal Buffering during concurrent loading. When attention is directed to one location, the information can be buffered such that it is not destroyed by attention moving to another location. In reading for example, fixating one word does not clear memory of all previous words. The problem of buffering with concurrent loading is illustrated in Figure 2. Assume the A and B modules both project to the X and Y modules. If the A signal is transmitted it goes to both the X and Y modules. Physiologically, once a signal exits a cortical module it outputs to all sites it is connected to. To buffer information one would like to transmit the A signal to X and the B signal to Y. The problem is that transmitting the B signal to Y produces interference in the X module causing possible loss of information. The same problem occurs when serially loading a set of inputs that must be examined in parallel. In this case the A module must output to the X module for the first stimulus and the Y module for the second. Since the A module is connected to both modules, it is critical that the second transmission to Y does not delete the previous signal to X.

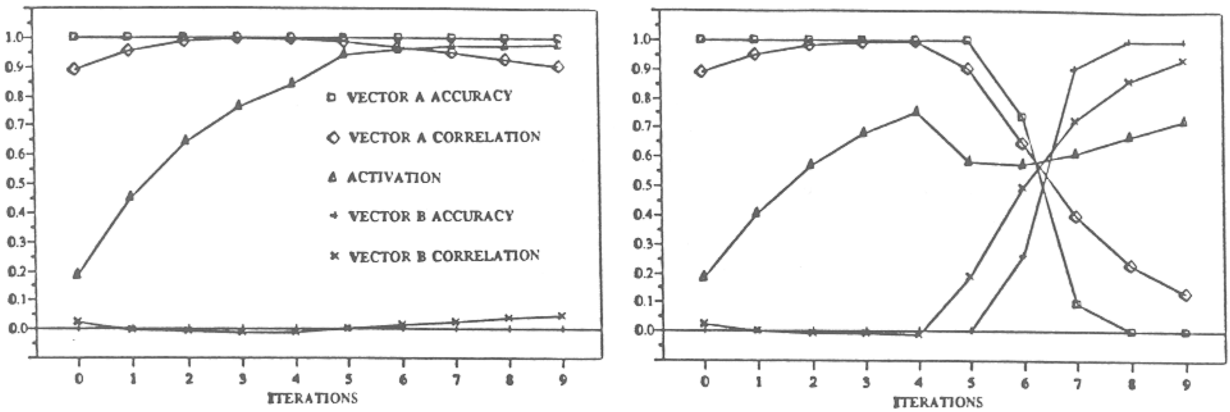

FIGURE 5: Signal maintenance with feedback. In left panel, on iterations 0 to 4 vectorA only is presented. On iteration 5 feedback is increased from 0.1 to 1.5. On iterations 5 to 9 vectorB only is presented. Shown are best match accuracy for vectorA and vectorB, correlation with targets for vectorA and vectorB, and activity of the hidden layer. The signal for vectorA holds throughout. In right panel there is no feedback. The signal for vectorA dies out as soon as the incoming stimulus is replaced with vectorB, and the signal for vectorB is increased.

The auto-associative feedback shown in this model provides a possible mechanism for signal buffering during concurrent loading. If the feedback is high enough, the module will maintain its signal and block out competing signals. In the simulation, a stimulus vector (vectorA) is presented to the network for 5 iterations, allowing the strength of the vectorA signal to build within the module. On the sixth iteration feedback is turned up from 0.1 to 1.5, and a different stimulus vector (vectorB) is presented for 5 iterations. The increase in feedback is sufficient to maintain the signal of vectorA without allowing any contamination from vectorB. The left panel of Figure 5 shows the accuracy and correlations for both stimuli throughout the ten iterations. VectorA remains the strong and accurate signal throughout, and the incoming vectorB does not have an effect. Compare these results to the case (Figure 5, right panel) where feedback was not utilized (feedback = 0.0). In the case where there is no feedback the vectorA signal begins to die as soon as presentation of vectorA is replaced with presentation of vectorB. The competing vectorB signal builds quickly and takes over the system.

Conclusion. The present connectionist control architecture implements some major features of cortical structure. The control of the gain of the output of a module allows selective enhancement of one message by attenuating competing messages. The attenuation of the unattended signal at levels representative of physiological attenuation produces a strong attentional capture effect. The use of auto-associative feedback provides a way to control input at the receiving module. The auto-associative feedback allows the system to maintain briefly presented information for later output after the stimulus has ceased. The feedback control can also latch a signal within a module and reduce the interference of concurrent transmissions that are directed at other modules but still input to that module. This architecture will be explored further to directly simulate human attentional effects, and determine the computational performance of incorporating modern cortical connectivity in a modular connectionist architecture.

References

Broadbent, D.E. (1957). A mechanical model for human attention and immediate memory. Psychological Review, 64, 205-215.

Broadbent, D.E. (1958). Perception and communication. London: Pergamon Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (in press). On the control of automatic processing: A parallel distributed processing account of the Stroop effect. Psychological Review.

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. Journal of Experimental Psychology: General, 112, 240-264.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology, 4, 219-227.

Lund, J. S., Hendrickson, A. E., Ogren, M. P., & Tobin, A. E. (1981). Anatomical organization of primate visual cortex area VII. Journal of Comparative Neurology, 202, 19-45.

McClelland, J. L., & Rumelhart, D. E. (1988). (Eds.), Explorations in parallel distributed processing: A handbook of models, programs, and exercises. Cambridge, MA: MIT Press.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science, 229, 782-784.

Mountcastle, V.B. (1979). An organizing principle for cerebral cortical function: The unit module and the distributed system. In F.O. Schmitt & F.G. Worden (Eds.), The neurosciences: Fourth Study Program (pp.21-42). Cambridge, MA: MIT Press.

Mozer, M. C. (1988). A connectionist model of selection attention in visual perception. In Proceedings of the Tenth Annual Conference of the Cognitive Science Society (pp. 195-201). Hillsdale, NJ: Erlbaum.

Mozer, M. C. & Behrmann, M. (1989). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia, Tech. Rep. CU-CS-441-89. University of Colorado at Department of Computer Science, Boulder, CO.

Peters, A. (1984). Chandelier cells. In A. Peters & E. G. Jones (Eds.), Cerebral cortex: Vol. 1: Cellular components of the cerebral cortex (pp. 361-381). New York: Plenum.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by backpropagating errors. Nature, 323, 533-536.

Schneider, W. (1985). Toward a model of attention and the development of automaticity. In M. Posner & O. S. Marin (Eds.), Attention and performance XI (pp. 475-492). Hillsdale, NJ: Erlbaum.

Schneider, W. & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), The psychology of learning and motivation, Volume 21 (pp. 54-119). New York: Academic Press.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey & R. D. Luce (Eds.), Steven's handbook of experimental psychology, Volume 2: Learning and cognition (739-809). New York: John Wiley & Sons.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II: Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Sperling, G. (1960). The information available in brief visual presentations. Psychological Monographs, 74 (Whole No. 498).

Treisman, A. (1960). Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 12, 242-248.

Treisman, A. M., & Riley, J. G. A. (1969). Is selective attention selective perception or selective response? A further test. Journal of Experimental Psychology, 79, 27-34.

Acknowledgements: This research was supported in part by ONR contract N00014-87K-0397 and N00014-86-K-0678 and ARI contract MDA903-89-K-0174.