The Phase Tracker of Attention

Erik D. Lumer
Xerox Palo Alto Research Center
Palo Alto, CA 94304
lumer@parc.xerox.com

Abstract

We introduce a new mechanism of selective attention among perceptual groups as part of a computational model of early vision. In this model, selection of objects is a two-stage process: perceptual grouping is first performed in parallel in connectionist networks which dynamically bind together the neural activities triggered in response to related features in the image; secondly, by locking its output on the quasi-peridic bursts of activity associated with a single perceptual group, a dynamic network called the *phase-tracker of attention* produces a temporal filter which retains the selected group for further processing, while rejecting the unattended ones. Simulations show that the network's behavior matches known psychological data that fit in the descriptive framework of object-based theories of visual attention.

Introduction

In most elaborate perceptual systems with limited processing resources, mechanisms that focus the attention on small parts of the sensory inputs are often necessary in order to cope with the complexity of the sensed world. The importance of attention in everyday activity has been a major impetus for its extensive study by psychologists and neurophysiologists (Eriksen & St-James, 1986; Crick, 1984; Duncan, 1984; Treisman & Gelade, 1980). As a result of their work, a number of theories have been developed, that fall in general under either one of two broad classes, known as the location-based and object-based theories of visual attention (Duncan, 1984). The first class stipulates that, at any given moment, attention is entirely allocated to a single convex region of space. In this model, the spatial dimension is the basic cue used to direct attention, which is therefore often compared to a mental spotlight. The psychological evidence that supports location-based theories along with the spotlight metaphor is reminiscent of a variety of experimental paradigms, including response competition (Eriksen & St-James, 1986), spatial precueing, and visual search (Treisman & Gelade, 1980). Location-based theories may be contrasted with object-based theories, which assume that attention can be allocated to one or more perceptual groups, regardless of their spatial locations. Object-based theories describe early perception as a two-stage process: the segmentation of images into distinct perceptual groups is done according to low-level, data-driven mechanisms of perceptual organization that exploit detected properties of proximity, continuity, similarity, or common motion, among others. In contrast with the parallel preattentive stage, a second stage of visual processing, called *focal attention*, is serial and consists in the selection and analysis of a particular perceptual group (Neisser & Becklen, 1975).

The experimental evidence in support of an object-based form of attention is multiple. It indicates in particular that subjects are better able to report two properties of the same object than one from each of two objects that are at the same spatial location (Duncan, 1984). Rock and Gutman (1981) also observed that subjects who were directed to attend to only one of two overlapping and novel figures (say the red figure among a red and green one) showed no recognition of form for the unattended one. Such result is not predicted by standard location-based theories of attention. More recently, Driver and Baylis (1989) showed that in response competition experiments, the grouping of target and distracting elements by common motion can have more influence than their proximity. These results are consistent with the hypothesis that attention can be directed to perceptual groups whose components are not spatially contiguous. Despite these and many other experimental facts, very little has been done to address the computational and neurological issues raised by the existence of an object-based form of attention. This situation contrasts with the flurry of recent work devoted to the modeling of location-based mechanisms of attention (Ahmad, 1991; Mozer, 1988).

The goal of this paper is to map the conceptual framework of object-based theories into a crisp computational model which has better predictive value. Note

This work was partially supported by the Air Force Office of Scientific Research contract No. F49620-90-C-0086 given to B.A. Huberman.

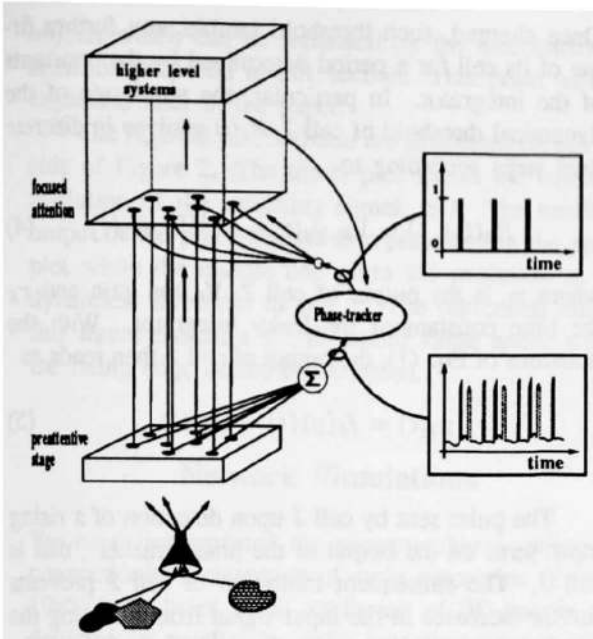

Figure 1 : Functional model of selective attention.

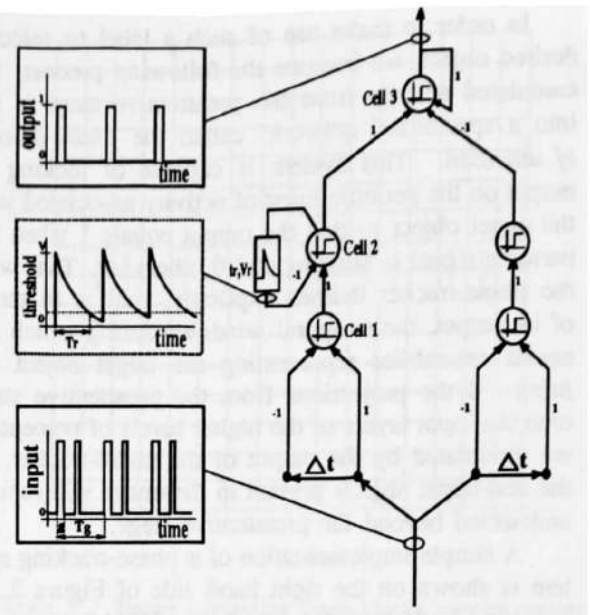

Figure 2 : Connectionist phase-tracker.

that in this framework, the objects that the focal attention can select or discard are defined at a preattentive stage. Thus, one expects mechanisms of attention to be intimately tied to the ones underlying perceptual organization. In what follows, this relationship is unraveled in terms of compatible mechanisms of interactions among neurons. In our model, the organization of visual scenes is based on a biologically inspired mechanism of labeling of perceptual groups (Gray et al., 1989; von der Malsburg, 1981), in which neural assemblies express their membership to a perceptual group by firing simultaneously and in a pseudo-periodic fashion, while being out of synchrony with neurons stimulated by other groups or a background. A number of authors have recently demonstrated the feasibility of perceptual grouping via synchronization of neural activity in large heterogeneous networks (Lumer & Huberman, 1992; Sporns, Tononi, & Edelman, 1991; Baldi & Meir, 1990). We refer to (Lumer 1992) for a description of how this mechanism is implemented in our model. In this paper, we focus more specifically on the issue of internal access to perceptual groups constructed in this way: given *implicit* temporal labels, i.e. the relative phases of neural oscillations distributed across a population of detectors, we still face the problem of how to use them *explicitly* in a mechanism of visual attention that selects among groups for further processing. A solution to this problem is proposed in the next section in the form of a dynamic network called the *phase-tracker of attention*. This system is then tested on examples which mimic the conditions and observed behaviors in a number of psychological experiments. The paper ends with a short discussion about the implications of our work.

The Phase-Tracker

We develop and study below a computational model of a two-staged visual system of the kind described by Neisser (1975) and others.

A coarse schematic representation of the model is given in Figure 1. Let us assume that the segmentation of perceived images is achieved at a preattentive stage via the synchronization of neurons that fire periodically in response to the local properties of a same object. The discussion of how this is actually done is reported elsewhere (Lumer, 1992). In the present context, suffices it to notice that the cumulated activity emerging from the preattentive stage evolves in time as one or several intertwined and periodically bursting signals superimposed over a low level stochastic noise. The noisy activity results from the asynchronous firing of cells stimulated by an incoherent percept or background. Each periodic burst of activity, on the other hand, is associated with a single perceptual group so that its phase can serve as a unique label referencing that group.

In order to make use of such a label to select a desired object, we imagine the following process: the cumulated activity from the preattentive stage is fed into a specialized network, called the *phase-tracker of attention*. This system is capable of locking its output on the periodic burst of activity associated with the target object so that the output equals 1 when the periodic signal is bursting and 0 otherwise. That way, the phase-tracker defines explicitly, that is in terms of its output, the temporal windows during which the neural assemblies representing the target object are firing. If the projections from the preattentive stage onto the input layers of the higher levels of perception are modulated by the output of the phase-tracker, all the non-target objects present in the image will remain undetected beyond the preattentive stage.

A simple implementation of a phase-tracking system is shown on the right hand side of Figure 2. It consists of a hybrid dynamic network exhibiting transient states, delayed propagation and feedforward as well as feedback connections. Each unit in the network connects its total input (i.e. presynaptic) activity, x, with its (postsynaptic) output, y, via a sharp thresholding function that is defined as

$$y = f_\theta(x) \tag{1}$$

where

$$f_\theta(x) = \begin{cases} 1 & if\ x \geq \theta \\ 0 & otherwise. \end{cases} \tag{2}$$

The input to the phase-tracker, $s(t)$, is propagated along a left and a right branch. The two branches act as rising and falling edge detectors, respectively. Let us first take a closer look at how the rising edge detector works. The presynaptic connection to cell 1 (see Figure 2) produces the first order difference of the input signal. When larger than the threshold θ, this difference causes cell 1 to fire. Stated more formally, the output of cell 1, $y_1(t)$, is related to the input signal s(t) by the relation

$$y_1(t) = f_\theta(s(t) - s(t - \Delta t)) \tag{3}$$

where Δt is a positive time delay. With a proper value assigned to Δt, cell 1 will turn on as a result of any sharp increase of its input. It therefore plays the role of a rising edge detector.

The output of cell 1 is fed into cell 2, which possesses a dynamical threshold. The properties of networks of cells with dynamical thresholds have recently been studied by a number of people (Abbot, 1990; Horn & Usher, 1989). In essence, a dynamic threshold is a transient feedback link from the thresholding cell onto itself. It is usually modeled as a leaky integrator which gets charged by the output activity of its cell. Once charged, such threshold inhibits any further firing of its cell for a period determined by the constants of the integrator. In particular, the amplitude of the dynamical threshold of cell 2, $R_2(t)$ evolves in discrete time steps according to

$$R_2(t+1) = V_R.y_2(t) + e^{-1/\tau_R}.R_2(t) \tag{4}$$

where y_2 is the output of cell 2, V_R the gain and τ_R the time constant of the leaky integrator. With the notations of Eq. (1), the output of cell 2 then reads as

$$y_2(t) = f_\theta(y1(t) - R_2(t)). \tag{5}$$

The pulse sent by cell 2 upon detection of a rising input turns on the output of the phase-tracker , that is cell 3. The subsequent inhibition of cell 2 prevents further increases in the input signal from affecting the output of the phase-tracker for a period of time T_R. This refractory period is related to the parameters of cell 2 via

$$T_R = Ceil\left(\tau_R ln\left(\frac{V_R}{1-\theta}\right) + 1\right) \tag{6}$$

where the function *Ceil(.)* rounds its argument up to the nearest integer value and accounts for the discrete nature of the dynamics. Because of the static feedback connection from cell 3 onto itself, its output remains at a high level until a pulse from the falling edge detector resets it to zero. The falling edge detector is very similar to the rising edge detector. By changing the sign of the first order difference computed at the input of the right branch with respect to that used in the left branch, its output will fire in response to any sharp decrease of the input to the phase-tracker.

Consider thus a periodic signal placed at the input of the phase-tracker, which consists of bursts lasting for an interval of time g and separated from each other by regular intervals T_g. The rising edge of the first burst causes the phase-tracker to turn on, a state which is kept by the system until the burst dies off, at which point the falling edge detector emits a pulse and the output of the phase-tracker switches to a low value. The phase-tracker is then inhibited during an interval of time T_R following the rise of the detected burst. It will therefore accurately track the phase of the incoming signal provided that the refractory period is shorter than T_g. Furthermore, other signals added to the input in between two successive bursts of the tracked activity will be ignored by the system as long as they precede or follow a detected burst by an interval of time larger than $T_g - T_R$. This difference defines the *resolution* of the phase-tracker and constrains in part the number of

objects which can be separated by the mechanism of attention proposed in this section. This point will be expanded later in the paper.

The regimes just outlined are illustrated on the left side of Figure 2. The lower plot shows the temporal evolution of the incoming signal, $s(t)$. The resulting output of the phase-tracker is represented in the upper plot while the middle one gives the evolution of the dynamical threshold of cell 2. The horizontal line in this figure indicates the threshold value under which the rising edge detector is enabled.

Network Simulations

We have implemented the phase-tracker as part of a connectionist architecture of early perception (Lumer, 1992). In brief, local attributes of 2D-images are detected in parallel by cells organized in a number of feature maps. The segmentation of images via synchronization of activity is done in grouping maps whose outputs are projected in a one to one fashion on the maps which define the first stage of the higher levels of perception. The inputs to the grouping maps can be restricted by a coarse location-based mechanism of attention which works cooperatively with the phase-tracker. More will be said about spatial attention in the final part of the paper. The cumulated output from the grouping maps is fed into the phase-tracker. A simple control mechanism allows the use of top-down information in the selection of objects with specified features: a global detector is associated with each map at the entrance of the higher levels of perception and signals whether the corresponding feature is present in the group currently selected. If this is the case and the feature does not match the description of a target object, the state of the phase-tracker is automatically reset so as to track the next label which is available at its input. To get a rough estimate of the time scales involved in the studied mechanism of selection, we equate one time step in our simulation with 1 msec of real time. The time constants of the grouping cells are set so that the cells fire once every 25 cycles, that is at a frequency of 40Hz. This number is consistent with the observed frequencies of neural oscillations in the primary visual cortex of cats (Gray et al., 1989). The resolution of the phase-tracker, as defined above, is equal to 1 msec. Finally, the visual field in the simulations is an array with 16 by 16 pixels.

The system was tested on a number of examples in which the phase-tracker takes advantage of the temporal separation of perceptual groups that cannot be easily discriminated spatially. Thus, a simple spatial spotlight of attention will fail in these cases. In particular, we have demonstrated the ability of our system to selectively focus on either one of two overlapping figures distinguished from each other by their respective colors. This behavior is in agreement with psychological observations (Rock & Gutman, 1981). Similarly, attention can be restricted to a non-contiguous set of objects animated by a common motion in a setting that mimics, albeit in a caricatural fashion, recent experiments performed by Driver and Baylis (1989). To save space, we will only detail the second example. The image in Figure 3 is composed of five 2x2 objects. The center and two far end elements are animated by a common upward motion while the intermediate objects move downward. The control system is instructed to focus only on the objects moving upward during the first 100 iterations before shifting attention to the other group of objects. Figure 4 displays the input (lower plot) and output (upper plot) of the phase-tracker as a function of time. After a transient period of about 25 iterations during which the temporal labels are formed, the phase-tracker locks on the index to the objects moving upward (their shared label is represented in grey for illustrative purpose only). Attention is released from this group at t=100 msec and redirected towards the objects moving downward after an equivalent time of about 20 msec (the corresponding label for these objects is shown in black). Notice that the locking of the phase-tracker on a label translates into the selection for further processing of the entire group indexed by that particular label.

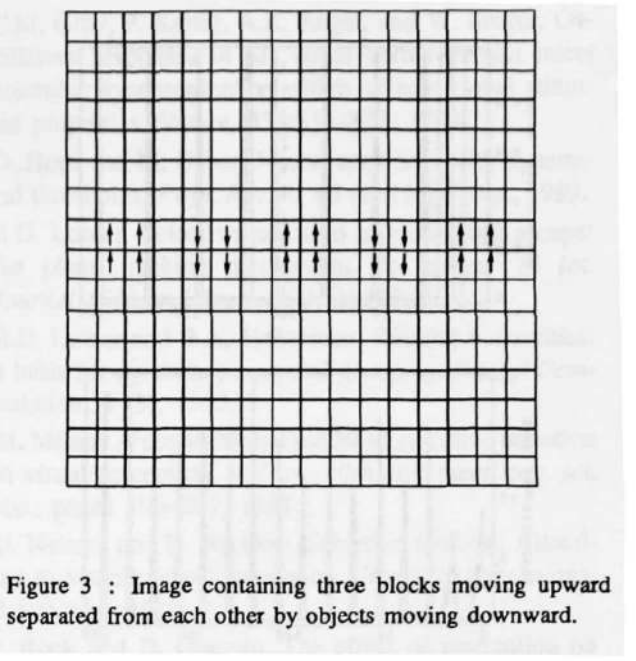

Figure 3 : Image containing three blocks moving upward separated from each other by objects moving downward.

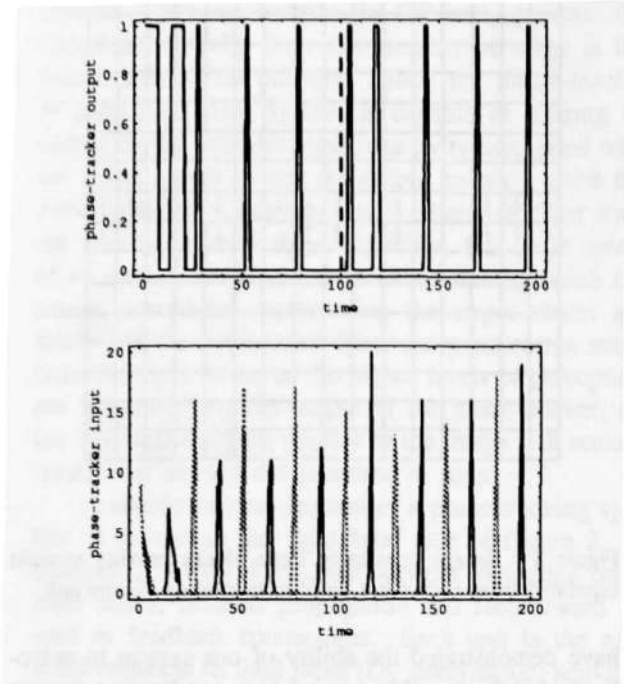

Figure 4 : Tracking of objects with common motion.

Discussion

In this paper, we have presented a non-spatial process of selection among perceptual groups, which overcomes the shortcomings of location–based models of visual attention. The proposed mechanism of selective attention presupposes the segmentation and labelling of perceptual groups via synchronization of the neurons responding to the local properties within a group. Our work is therefore complementary to the flurry of recent reports showing that this type of grouping can be achieved using simple dynamical networks. Indeed, any improvement of image segmentation via dynamic grouping will augment the potential of selection by the phase-tracker. Furthermore, the use of oscillatory dynamics in our model leaves the door open for a better modelling of the cortical tissues in which these regimes have been observed.

The embedding of the phase-tracker in a connectionist architecture of early visual processing reveals its capabilities and limitations. With the parameters used for our simulations, we observed that the bursts of activity, or labels, associated with a single perceptual group have a temporal duration on the order of 3 time steps. Furthermore, the resolution of the phase-tracker, as defined in the second section, is equal to one time steps. We therefore know that each unambiguous label produced by the segmentation system occupies on the time axis about 4 time steps. Since the consecutive firings of grouping cells are spaced by 25 time steps, we conclude that selection cannot operate on more than approximately 6 perceptual groups whose labels are placed at the input of the phase-tracker. This observation places a strong constraint on the interaction between a coarse mechanism of spatial orientation and selection: the former must be tuned so as to limit the number of objects that the dual mechanisms of segmentation and selection operate on at any given moment. To our knowledge, this constitutes the first embodied prediction of a possible interdependence between two modes of attention, i.e. location-based and object-based, that have traditionally been considered as orthogonal. We expect that future work, both experimental and computational, will further elucidate this relation.

Another very interesting observation can be drawn from the fact that the combined mechanisms of segmentation and selection have a maximum "capacity" of about 6 elements. Indeed, in trying to determine how the time required to quantify a collection of n items presented to view was function of n, it was found (Chi & Klahr, 1975) that a striking discontinuity occurs in the region of $n=6\pm1$. This phenomenon, known as subitizing, is characterized by a very rapid apprehension of the number of items below the discontinuity, while the reaction time increases linearly with the number of elements by a much larger increment above the critical point. It is tempting to speculate that a transition from an object-based form of attention to a serial scanning of spatial locations in the display might be related to the observed phenomenon. We also notice that sensory segmentation and selective attention are not the unique attributes of vision. For example, the auditory modality parses complex sound fields into independent streams, each one being associated with a specific external source. This capability is best illustrated in cocktail parties where one is able to distinguish several voices, and selectively attend to one, among a noisy crowd. Since the processes of segmentation of the auditory fields have been modelled as neural oscillators which either synchronize or desynchronize their phases (von der Malsburd & Schneider, 1986), a phase-tracker could likewise be used in this context to implement selective attention.

Last but not least, this paper illustrates the richness of computational mechanisms which can be derived from the use of dynamical networks having transient states. As connectionist models of cognition become larger and develop modularity, the issues of communication and coordination between the heterogenous modules become central. In this context, the connectionist equivalents of communication devices, such as signal multiplexers, clock synchronizers, and phase-locked loops of the kind studied here, are expected to play a fundamental role.

Acknowledgments

I thank Bernardo Huberman, Eric Saund, Roger Shepard and David Rumelhart for helpful discussions as well as Jeff Shrager for mentioning the phenomenon of subitizing.

References

L.F. Abbot. A network of oscillators. *Phys. A: Math. Gen.*, 23:3835–3859, 1990.

S. Ahmad and S. Omohundro. Efficient visual search: a connectionist solution. *In Proc. 13th ann. conf. cog. sci. soc.*, 1991.

P. Baldi and R. Meir. Computing with arrays of coupled oscillators: an application to preattentive texture discrimination. *Neural Comp.*, 2 (4):459–471, 1990.

M.T.C. Chi and D. Klahr. Span and rate of apprehension in children and adults. *Journal of Experimental Child Psych.*, 19:434–439, 1975.

F. Crick. Function of the thalamic retcular complex: The searchlight hypothesis. *Proc. Natl. Acad. Sci. USA*, 81:4586–4590, 1984.

Jon Driver and G.C. Baylis. Movement and visual attention: the spotlight metaphor breaks down. *J. Exp. Psych.: Human Perc. and Perf.*, 15 (3):448–456, 1989.

J. Duncan. Selective attention and the organization of visual information. *J. Exp. Psych.: General*, 113 (4):501–517, 1984.

C.W. Eriksen and J.D. St.James. Visual attention within and around the field of focal attention: a zoom lens model. *Percept. and PsychoPhys.*, 40 (4):225–240, 1986.

C.M. Gray, P. Konig, A.K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337, 1989.

D. Horn and M. Usher. Neural networks with dynamical thresholds. *Phys. Rev. A*, 40 (2):1036–1044, 1989.

E.D. Lumer. Selective attention to perceptual groups: the phase tracking mechanism. *To appear in Int. Journal of Neural Systems*, 3 (1), 1992.

E.D. Lumer and B.A. Huberman. Binding hierarchies: a basis for dynamic perceptual grouping. *Neural Computation*, 4 (3), 1992.

M. Mozer. A connectionist model of selective attention in visual perception. *In Proc. 10th ann. meet. cog. sci. soc.*, pages 195–201, 1988.

U. Neisser and R. Becklen. Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7:480–494, 1975.

I. Rock and D. Gutman. The effect of inattention on form perception. *J. Exp. Psych.: Human Perc. and Perf.*, 7 (2):275–285, 1981.

O. Sporns, G. Tononi, and G.M.Edelman. Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proc. Natl. Acad. Sci. USA*, 88:129–133, 1991.

A.M. Teisman and G. Gelade. A feature-integration theory of attention. *Cog. Sci.*, 12:97–126, 1980.

A. Treisman. Focused attention in the perception and retrieval of multidimensional stimuli. *Percep. and Psychophys.*, 22:1–11, 1977.

C. von der Malsburg. The correlation theory of brain function. *Intern. report 81-2*, Dept. of Neurobiology, Max Planck Institute., 1981.

C. von der Malsburg and W. Schneider. A neural cocktail-party processor. *Biol. Cybern.*, 54:29–40, 1986.