Figure-ground separation: evidence for asynchronous processing in visual perception?

Burgi, Pierre-Yves

universität bern
institut für informatik
und angewandte mathematik

Proceedings
Third Annual Meeting
of the
Swiss Group for Artificial Intelligence
and
Cognitive Science

SGAICO'91

Editors

Hansruedi Kaiser, Rene Bach, Horst Bunke

IAM-91-004
March 1992

Adresse:
Institut für Informatik und angewandte Mathematik
Universität Bern
Länggass Str. 51
CH - 3012 Bern

e-mail: office@iam.unibe.ch
Telefon: (0041) 31 65 86 81
Fax: (0041) 31 65 39 65

Proceedings

Third Annual Meeting

of the

Swiss Group for Artificial Intelligence and Cognitive Science

SGAICO'91

Editors

Hansruedi Kaiser, Rene Bach, Horst Bunke

Abstract: October 10-11, 1991 the third annual meeting of the Swiss Group for Artificial Intelligence and Cognitive Science was held in Biel/Bienne (Switzerland). This report contains the full version of all papers presented at this meeting (except two).

CR Categories and Subject Description: I.2 [Artificial Intelligence]

FIGURE-GROUND SEPARATION: EVIDENCE FOR ASYNCHRONOUS PROCESSING IN VISUAL PERCEPTION?

P.-Y. BURGI

Artificial Intelligence and Vision Group
Computing Science Center, University of Geneva
12 rue du Lac, CH-1207 Geneva, Switzerland

ABSTRACT

The performance of the human visual system in extracting noisy figures from a noisy background is astonishingly good. Even in situations of very poor contrast, boundaries emerge clearly. Conversely, typical edge detectors fail to give good results for such images. In an attempt to explain the discrepancies in these performances we have developed a neural network model relying on two assumptions, both of which are based on neurophysiological findings. Firstly, the processing of visual information is considered to be asynchronous: stimuli are delayed accordingly to their intensity. Secondly, emergent boundaries have the property of producing coherent responses corresponding to the (near)-simultaneous responses of cells in the cortical orientation columns. Results show that such neural network can indeed benefit from the asynchrony when treating images with ratio signal-to-noise particularly low.

1. Introduction

Resolution of the figure-ground separation problem is essential in a visual system. Generally this problem implies that contours separating objects from background are extracted. For synthetic images where background and objects have distinct light intensities this problem is easily solvable, especially if there is no noise. However, this problem becomes thorny when noise is added or when natural images are used. Techniques of segmentation robust to noise have been developed[1,2,3] but are not discussed here. Until now, no evidence has been found that such techniques are biologically plausible. Conversely, lateral inhibitory neuronal networks performing high pass filtering are known to exist and are commonly found in biological systems. In particular, on-center off-surround receptive fields as well as Gabor-like oriented filters are found in the human visual system. Furthermore, orientation columns found in the primary cortex are quite suited for cooperative mechanisms relaxing the constraint of building contours by using local information only.

Another aspect of perception known for a long time is the relation between latency of responses of retinal ganglion cells and light intensity. For images with low signal-to-noise ratio it is hypothesized that this relation can somehow alleviate the resolution of the figure-ground separation problem. In addition, a generalization of the relation between latency and response amplitude is done. Owing to the introduction of latencies this approach is termed as asynchronous by opposition to conventional approaches where no such temporal constraint is incorporated, thus called synchronous.

On the basis of this neurophysiological evidence an architecture has been developed and a comparison study made between synchronous and asynchronous approaches. It will be shown in particular that for this specific figure-ground separation problem the asynchronous approach performs better than the synchronous one.

2. Model

The various elements involved in the resolution of the figure-ground separation problem are now discussed. There are two main cornerstones for this study. First, a relation between the brightness of a picture element and the latency of the response it produces is supposed to exist; a direct consequence of this brightness-latency relationship for images containing regions varying in intensity is that processing must be considered asynchronous as visual information is not treated in one step. Second, near-simultaneous spatially-distant featural responses are supposed to reflect global structures found in the image; contours being of particular interest in the figure-ground separation task, oriented image columns were chosen to support mechanisms detecting simultaneity in responses. Oriented image columns are sets of connected points crossing the image from one end to the other along a specific orientation.

Models yielding the asynchrony as well as representing image columns are thus required. Both are described separately in the next two sections while their integration in a general architecture comprising the different processing levels constitutes a third section.

2.1 Origin of the Asynchrony

For decades it has been well known to psychologists that there exists a dependence of visual latency on light intensity[4]. Illustrations of this phenomenon are the Hess effect and the Pulfrich effect. The Hess effect, first reported by Carl von Hess in 1904, is a monocular illusion in which a difference in target luminance causes a change in the relative apparent location of two laterally moving targets[5]. The Pulfrich stereo-effect, described in 1922 by Carl Pulfrich, is easily experienced when wearing two different lenses, one being dark gray and the other clear: the image coming in from the dark lens is perceived a little later than the image coming in from the clear lens; this latency induces a stereoscopic visual illusion when an observer is looking at a swinging pendulum bob. Visual latencies appear also in reaction times for stimuli with varying physical intensity[6,7] and for different contrasts[8,9].

Such latencies have also been measured physiologically but the literature is scarce about that topic, particularly for primates. Nevertheless, physiological recordings have been performed on the Pulfrich effect for the cat[10]. Furthermore, the latency-luminance relation under dark and light adaptation has been determined on retinal units of the bullfrog retina[11], of the marine toad Bufo marinus[12] and of the cat[13]. For the dark-adapted condition, Chapman reported a simple linear function relating reciprocal of latency of the first impulse to the logarithm of light luminance; other functions have also been proposed[14].

By merely considering cells to be leaky integrators and information transmitted between cells coded in spike frequency it is possible to explain the light intensity-latency relation. Frequency modulation is largely used in neuronal structures of the brain to code response amplitudes by train of spikes. When receiving a spike on excitatory synapses, cells increase their membrane potential; conversely, between spikes, their membrane potential diminishes due to leakages. Clearly, a train of spikes makes the membrane potential increase and decrease; after a while there is a convergence towards an average value. In such conditions, one is naturally led to consider the problem of determining for a given frequency when the cell will first reach a fixed threshold. The time needed for that is the latency of the cell. It can easily be shown that the relation between frequency and latency is nonlinear and mimics neurophysiological data[15]. It must be noted that the relation frequency-latency holds also when spikes are replaced by constant potentials whose amplitude has for value the frequency of the spikes. However, to have a good correspondence, the period separating spikes must be negligible with respect to the cell time constant[16]. For the retina the dependence of latencies on light intensity seems to be induced by specific chemical mechanisms in the photoreceptors[12], but this fact will not change the logic presented in this paper. In summary, low frequencies or low potentials have correspondingly much larger latencies than those brought out by high frequencies or high potential; this observation is generalized to all neuronal structures. For this generalization it interesting to note that Cattell (1885) demonstrates psychologically that the intensity of electric shock or of light yield similar effects on reaction time[4].

For a system composed of various levels of processing, the relation amplitude-latency has an obvious effect: asynchrony. Indeed, signals with differing amplitudes are converted into trains of spikes of corresponding frequencies before being fed into the first level of processing. Cells pertaining to this level will yield output signals with differing latencies. Subsequent levels will thus be fed with signals arriving at different times, those corresponding to high amplitudes coming earlier than those with lower amplitudes. This distribution of information over time forms asynchronous data flows between the different levels of processing. The aim of this paper is to show that for a given architecture such data flows produce results differing from those that would be obtained with conventional synchronous approach where no time differences would be taken into account. Before presenting the architecture allowing a comparison between the two approaches, namely synchronous versus asynchronous, the question of cooperation must be addressed.

2.2 *Towards a Cooperative Model*

From a point of view of signal processing, it is rather obvious that local information is seldom sufficient. A very simple example is shown in Figure 1 where gaussian noise has been added to an edge. While the vertical edge is clearly perceived when the whole image is displayed, a reduced view as seen through a window is hardly enough to accurately determine the position of the edge.

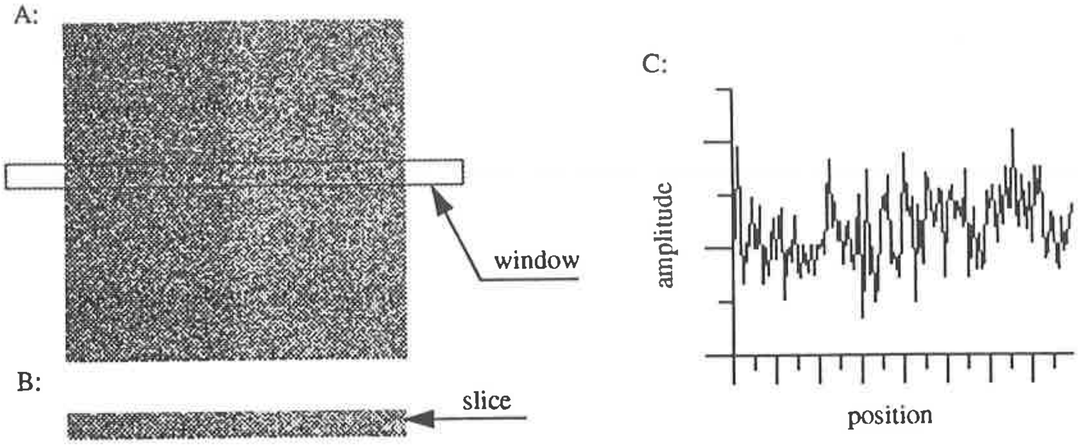

Figure 1 : A: Gaussian noise of standard deviation 35 has been added on a step of height 25. The resulting signal-to-noise ratio (SNR) is 0.5. In spite of the low SNR, position of the edge between the dark and bright part is clearly perceived. Image size is 128 x 128; B: Image shown in A has been windowed and perception of the edge is very dim and inaccurate. Image size is 11 x 128; C: Slice of B showing the brightness in function of the position. In this graph the position of the edge does not appear anymore.

With respect to this example it is easy to comprehend that a system can benefit from using spatially-distant information along an orientation. It is interesting to note that long range connections between cells corresponding to different locations but sensitive to same orientations are known to exist in the primary visual cortex[17,18]. Cooperative models based on the columnar architecture found by Hubel and Wiesel[19] and using long-range receptive fields have been developed[20]. A recent approach makes use of the synchronization of oscillatory responses among columns[21]. Such cooperative interactions have been physiologically measured on the cat[22].

In our model we make use of oriented image columns within which a cooperative mechanism operates. Along an image column this cooperative mechanism detects the coherency possibly existing among cell responses. Following the idea that there is a correspondence between cell responses and spike frequencies, which usually follows a sigmoidal function, there is also a dependence between cell responses and the latency for the appearance of the first spike. Thus they are two effects influencing the measure of similarity between responses, namely the frequency and the latency. To express these two factors, cell responses are represented by vectors whose length is related to the frequency and whose angle is related to the latency. This conversion, called vector conversion, is illustrated in Figure 2. Latency being deduced from frequency there is obviously a redundancy in this transformation. The function chosen for converting frequency into vector magnitude is linear with a range defined by the maximum and minimum values fixed by the sigmoidal function. Vector angles are chosen to vary linearly with latencies in the range $[0, \pi]$ to avoid cancellation of vectors pointing in opposite directions. This situation would correspond to inhibitory connections which are not wanted.

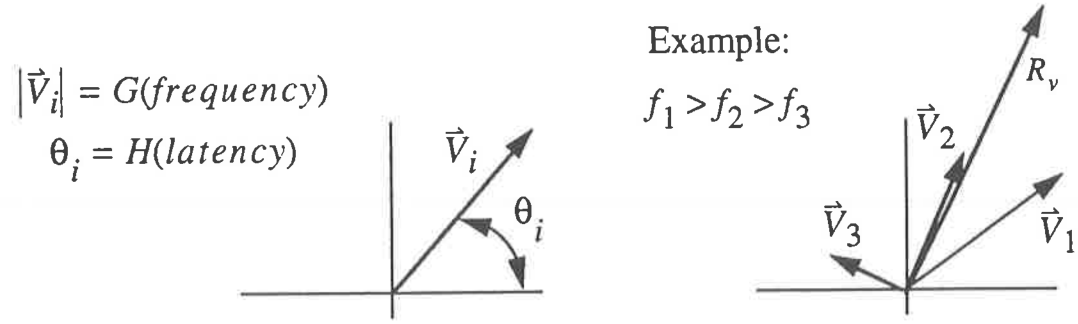

Figure 2 : Illustration of the vector conversion. There is a linear relation between spike frequency f and vector magnitude $|\vec{V}|$ as well as between latency and vector angle θ. The example shows how the vectors corresponding to three different frequencies are disposed.

Measuring the coherency among responses for a specific image column is made in three stages: (i) convert responses into vectors according to the previous rules; (ii) calculate the vectorial sum of the vectors determined in the previous step. The resulting vector gives a measure of strength of the responses along an image column with a weak measure of the extent of spreading of the vectors; (iii) make a measurement of the dispersion of the all the vectors with respect to the resulting vector calculated in the previous step. The dispersion of n vectors of length $|\vec{V}_i|$ and angle θ_i with respect to a resulting vector of length R_v and angle θ_{R_v} is measure by applying the following definition*:

$$c = R_v \cdot \sum_{i=1}^{n} |\vec{V}_i| \cos(\theta_{R_v} - \theta_i) \tag{1}$$

As the angle of the vectors is related to the appearance of the first spikes, the measure of dispersion c corresponds to determining the temporal coherency between responses (along a specific image column). It is clear that cell responses with particularly strong values are favored as their corresponding vectors have small angles and great magnitudes. This goes with the spirit that signals coming first, thus the strongest ones, could exert a stronger action than those coming next.

2.3 Architecture

There are two questions relative to the present system architecture whose task is to extract edges separating a figure from a background. Firstly, what are the different levels of processing? Secondly, how is the asynchrony disposed? To allow an evaluation of the results of the synchronous versus asynchronous approaches two steps are necessary: (i) the core of the synchronous architecture, based on knowledge in neurophysiology and signal processing fields, is designed; (ii) latencies and possibly temporal integration are consid-

*. A similar definition is used by Rao and Schunck (1991) to measure the flow orientation coherence of vectors corresponding to gradients[23].

ered. Thus, introducing the asynchrony will just be a matter of adding new elements on a structure devised for the synchronous processing. Note that the type of image to be processed has the constraint that the foreground brightness is distinct from the background brightness. Images relying only on spatial information are not considered.

2.3.1 Synchronous Architecture

While optimized filters have been designed to extract edges[24], the architecture proposed herein (Figure 3) is inspired from the human visual system and is restricted to three levels.

The first level of processing is a high pass isotropic filtering consisting of a small on-center and a large off-surround structure. This level corresponds physiologically to the ganglion cells in the retina[25,26]. By using appropriate functions this level of processing can demonstrate adaptive properties like the Weber-Fechner law as shown by Grossberg[27]. In our model we make use of linear filtering which does not take into account this law.

The second level implements anisotropic filtering by means of even Gabor filters. This level has its physiological correspondence in the visual cortex[28,29,30]. Results of this stage are half-rectified by imposing the rule that positive signals are unaltered in their transmission while negative responses are completely attenuated.

Finally, the third level applies a measure of coherency among responses taken along an image column whose orientation corresponds to the orientation of the anisotropic filter which was used to obtain them. Physiological evidence for this level has already been discussed in the previous section 2.2.

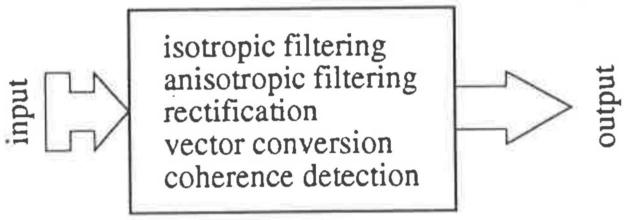

Figure 3 : Synchronous processing to detect contours separating a foreground from a background. Outputs give, for a specific orientation, values of coherence within image columns.

2.3.2 Asynchronous Architecture

Two elements are now introduced and complete the synchronous architecture described in the previous section.

First of all, latencies are added, requiring the addition of a stage which converts brightness information into latencies before filtering is applied (Figure 4). Signals entering the system are thus delayed according to their intensity and once they have appeared they keep their value, replacing the original value set to zero; this corresponds to sustained responses. The appearance of such signals is thus temporally distributed and produces a dynamic data flow; this flow makes outputs vary along time.

Secondly, all responses of the coherence detection stage are temporally integrated. The reason for that stage relies on the direct relation existing between the dynamic data flow and the image structure. The foreground being supposedly brighter than the background, the data flow will first reflect the foreground structure even if noise is added to such images.Thus, there exists a specific period during which there is a trade-off between the appearance of the image structure and the degree of noise. Concretely, this means that an optimum filtering occurs. No assumption being made on the image it is not quite obvious to decide when this optimum is reached. One way of avoiding this decision is to assume that this optimum in filtering yields responses which subsist even when they are integrated. This solution is illustrated in Figure 4. Note that for the synchronous case this specific period does not exist as information from both background and foreground are simultaneously considered.

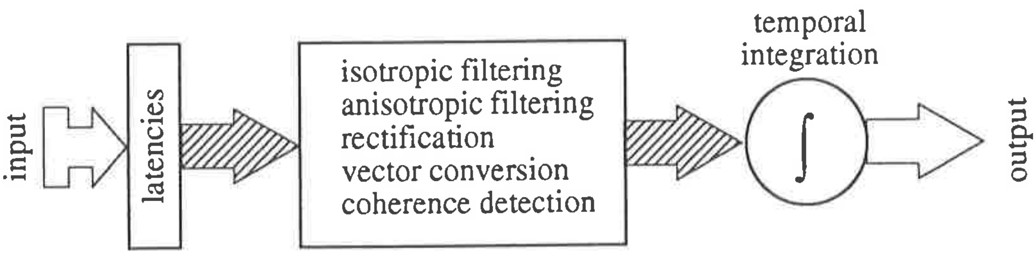

Figure 4 : Asynchronous processing based on the architecture given in Figure 3 but with two supplementary stages: latencies computation and temporal integration. Hatched arrows stand for the dynamic data flow.

3. Results

Comparison between synchronous and asynchronous approaches is made for two noisy images. Parameters are defined for the isotropic and anisotropic stages. There is one supplementary parameter needed for the asynchronous approach. Indeed, the asynchronous processing would require that each time new information appears its action on the output would have to be determined. However, for reasons of computational cost, a sampling time value must be fixed. While there is no inferior limit to this value except for the computational cost, a large value would lead to the synchronous case (Figure 7).

Both images have 128 x 128 pixels with the background filling the left half and the foreground filling the right half. Thus there is a edge in the middle to be located (Figure 5A and Figure 6A). Background and foreground are deteriorated with additive noise. In the first image the noise has a gaussian distribution while in the second it has an uniform distribution. Both images have the same signal-to-noise ratio (SNR) but the one with the uniform noise has a smaller contrast. Results in Figure 5B and Figure 6B show values of coherence along the vertical image columns. The higher the coherence is, the whiter it appears. When the edge is correctly detected the peak-to-peak ratio indicated on the right in Figure 5C and Figure 6C is greater than 1. Interestingly, synchronous processing yields peaks in the values of coherency which do not correspond to the edge between background and foreground while the asynchronous processing yields the highest peak rightfully in the middle. These results demonstrate the superiority of the asynchronous processing on the synchronous version. To illustrate the effect of varying the sampling time different values have been chosen to process the image asynchronously with gaussian noise. Results shown in Figure 7 make clear that an increase in the sampling time value yields a decrease in the performance, the limit case of such variations corresponding to the synchronous processing.

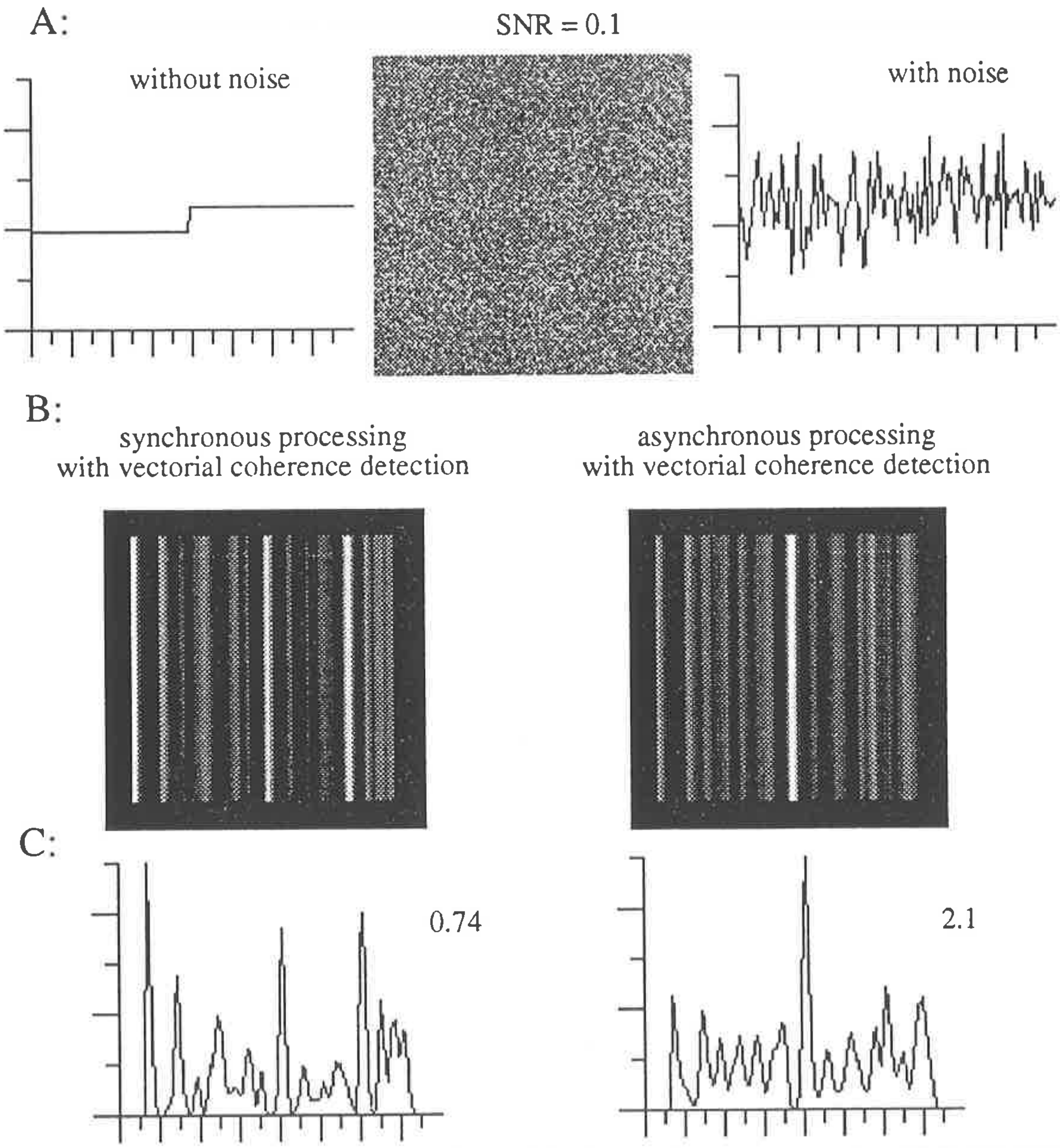

Figure 5 : A: Gaussian noise of standard deviation 79 has been added on a step of height 25; B: Coherency detection in vertical image columns. For the asynchronous processing, sampling along time of 2 ms; C: Slices of B, with numbers on the right indicating the worst peak-to-peak ratio. Images 128 x 128; isotropic filter 11 x 11; anisotropic filters 21 x 21, aspect ratio 0.5. Brightness range is [0 .. 255] corresponding to the latency range [22 ms .. 2 ms]. Note that due to the reproduction on paper maybe the edge in A is not visible.

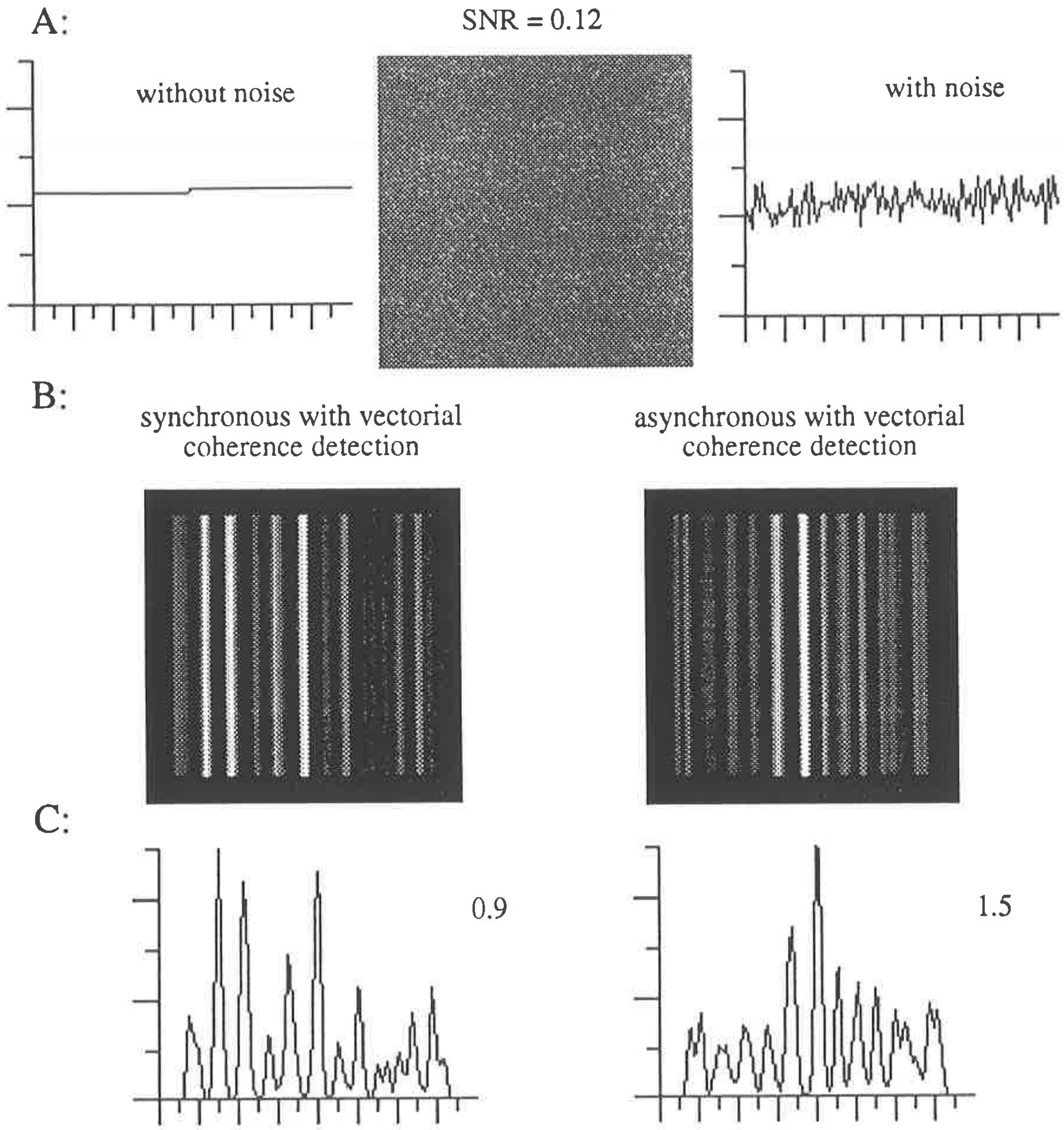

Figure 6 : A: Uniform noise of range 50 has been added on a step of height 5; B: Coherency detection in vertical image columns. For the asynchronous processing, sampling along time of 1 ms; C: Slices of B, with numbers on the right indicating the worst peak-to-peak ratio. Image 128 x 128; isotropic filter 11 x 11; anisotropic filters 21 x 21, aspect ratio 0.5. Brightness range is [0 .. 255] corresponding to the latency range [22 ms .. 2 ms]. Note that due to the reproduction on paper maybe the edge in A is not visible.

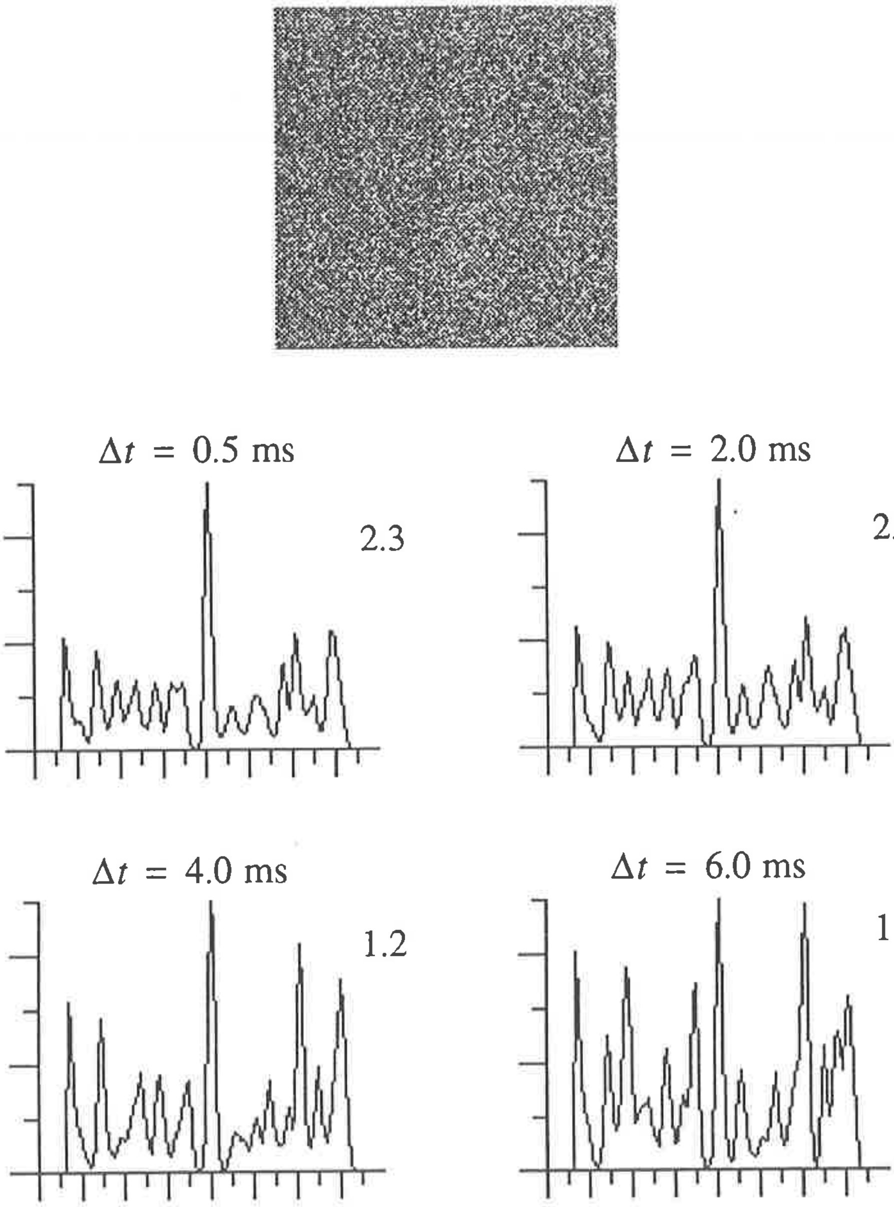

Figure 7 : Noisy step as in Figure 5 (Gaussian noise and SNR=0.1). Effect of varying the sampling along time on the performance; values are shown above graphs. Performance is measured as the peak-to-peak ratio with the values shown on the right of the graphs. Increasing sampling time diminishes the asynchrony, the limit being the synchronous case. Filter parameters as in Figure 5.

4. Discussion

While the model has been physiologically oriented it only crudely mimics reality. Stress has been more heavily laid on concepts rather than on details. A more faithful model would not only estimate latencies in the first and last levels but also at the different stages in-between. This would require more sophisticated processing as receptive fields would receive asynchronous signals. Spikes would have to be considered individually imposing a higher computational cost. Yet small differences in phases between signals impinging on receptive fields would not produce great differences in responses.

About the concept of time two points must be touched on which are of particular interest for those who work in image processing. They involve respectively two questions:

Are thresholds equivalent to sampling time values?

Sampling time values are just an artifact of modeling reality. The ideal case would be to consider a continuous flow, corresponding to continuously varying a threshold. By definition thresholds are discrete values and thus the correspondence with the temporal model would not hold. Another point of view is that thresholding needs to fix values above which signals are kept and below which signals are suppressed. With the temporal flow the system does not need to decide anything about signals; it just receives them as they arrive accordingly to latencies determined by a general law.

Would preprocessing be equivalent to temporal precedence?

Temporal precedence happens every time the data flow is considered asynchronous and some data arrive before other. In the present case points of high brightness appear before those of low brightness. One possible interpretation would be that strong signals are favored as they come first and could exert some specific action first. It terms of image processing this favoring would be equivalent to multiplying signal amplitudes with coefficients whose value would be greater for strong amplitudes and smaller for weaker ones. Nevertheless the comparison would stop short here. Using the temporal approach means that filtering is applied according to a structure implicit to the image. This structure is reflected through temporal precedence of some data over others. Conversely, image processing techniques first preprocess signals and then apply filtering on the image without taking into account any structure.

In spite of the seemingly elusive aspect of temporal precedence we believe it could play a main role also in competitive mechanisms. While our current results point out the index of coherency for every image column, the positions of contours have still to be delimited. Competitive mechanisms are thus required and latencies are thought to be of importance for that task. Other evidence that the information contained in the neuronal latency is very efficient for the resolution of some competitive problems, like the winner-take-all one, has indeed been reported[31].

5. Conclusions

The evidence tends to establishing that images containing points of differing brightness produce asynchronous data flows. When such flows are processed through a classical architecture with the supplementary constraint that responses are integrated temporally, we can show an increase in the robustness of processing images of low SNR. This result gives strong evidence that the figure-ground separation problem can benefit from asynchronous processing.

6. Acknowledgments

The author would like to thank Prof. Les Kitchen, Prof. T. Pun and D. Brunet for their various comments on the manuscripts and stimulating discussions.

7. References

1. P.B. Chou and C.M. Brown, in *Proceedings of Darpa Image Understanding Workshop* (1987).
2. P.A. Devijver and M.M. Dekesel, in *Reconnaissance des formes et intelligence artificielle*, ed. AFCET, (Dunod, 1987).
3. A. Rosenfeld and A.C. Sher, *Pattern Recognition* **21** (1988) 147.
4. J.M. Cattell, *Brain* **8** (1885) 512.
5. J.M. Williams and A. Lit, *Vision Research* **23** (1983) 171.
6. H.G. Vaughan Jr., L.D. Costa, and L. Gilden, *Vision Research* **6** (1966) 645.
7. M.E. Rudd, *Vision Research* **28** (1988) 179.
8. M.W. Greenlee and B.G. Breitmeyer, *Vision Research* **29** (1989) 1575.
9. D. Thompson and N. Drasdo, *Vision Research* **29** (1989) 309.
10. T. Carney, M.A. Paradiso, and R.D. Freeman, *Vision Research* **29** (1989) 155.
11. R.M. Chapman, *Journal of the Optical Society of America* **51** (1961) 1102.
12. R.J.W. Mansfield and J.G. Daugman, *Vision Research* **18** (1978) 1247.
13. P. Lennie, *Vision Research* **21** (1981) 815.
14. K. Donner K., *Visual Neuroscience* **3** (1989) 39.
15. P.-Y. Burgi and T. Pun, in *Channels in the Visual Nervous System: Neurophysiology, Psychophysics and Models*, eds. B. Blum, S. Hochstein, and W. Cavonius (Freund Publishing House, London, 1991).
16. P.Y. Burgi, submitted for publication.
17. C.D. Gilbert and T.N. Wiesel, *Journal of Neuroscience* **3** (1983) 1116.
18. K.S. Rockland and J.S. Lund, *Science* **215** (1982) 1532.
19. D.H. Hubel and T.N. Wiesel, *Journal of Physiology* **195** (1968) 215.
20. S. Grossberg and E. Mingolla, *Psychological Review* **92** (1985) 173.
21. O. Sporns, J.A. Gally, G.N. Reeke, and G.M. Edelman, *Proceedings of the National*

Academy of Science USA **86** (1989) 7265.
22. C.M. Gray, P. König, A.K. Engel, and W. Singer, *Nature* **338** (1989) 334.
23. A.R. Rao and B.G. Schunck, *Computer Vision, Graphics, and Image Processing* **53** (1991) 157.
24. J. Canny, in *Readings in Computer Vision*, eds. M.A. Fischler and O. Firschein (Morgan Kaufmann, 1987).
25. J. Richter and S. Ullman, *Biological Cybernetics* **43** (1982) 127.
26. U. Wehmeier, D. Dong, C. Koch, and D. Van Essen, in *Methods in Neuronal Modeling*, eds. C. Koch and I. Segev (The MIT Press, 1989).
27. S. Grossberg, *The Behavioral and Brain Sciences* **6** (1983) 625.
28. J.G. Daugman, in *Models of the Visual Cortex*, eds. D. Rose and V.G. Dobson (John Wiley and Sons, 1985).
29. V.D. Glezer, in *Models of the Visual Cortex*, eds. D. Rose and V.G. Dobson (John Wiley and Sons, 1985).
30. L.A. Palmer, J.P. Jones, and W.H. Mullikin, in *Models of the Visual Cortex*, eds. D. Rose and V.G. Dobson (John Wiley and Sons, 1985).
31. S.J. Thorpe, in *Parallel Processing in Neural Systems and Computers*, eds. R. Eckmiller, G. Hartmann, and G. Hauske (Elsevier Science Publishers B.V., North-Holland, 1990).