Vision Res. Vol. 18, pp. 101 to 105 Pergamon Press 1978. Printed in Great Britain.

BINOCULAR CROSS-CORRELATION IN TIME AND SPACE

Christopher W. Tyler
Smith-Kettlewell Institute, San Francisco, California 94115, U.S.A.

and

Bela Julesz
Bell Laboratories, Murray Hill, New Jersey 07974, U.S.A.

(*Received 6 July* 1976; *in revised form* 1 *April* 1977)

Abstract—Studies of the ability to perceive the degree of binocular correlation between random inputs (where no explicit binocular disparity occurs) show that detection can occur for presentation times of the order of the duration of one neural impulse. This supports a parallel model of neural processing for binocular correlation. In addition we describe the spatio-temporal integration properties of the binocular correlation system.

A fundamental property of the nervous systems of organisms is the ability to compare one pattern of input stimulation with another pattern presented at a different time or place.

Binocular stereoscopic vision can serve as a model system for studying the ability to make correlations between neural inputs. Correlation processes have frequently been proposed as a possible mechanism underlying human pattern recognition (Licklider, 1948; Reichardt, 1961). The time is now ripe to explore the characteristics of binocular correlation processes, and recent advances in display technology allow this to be accomplished on-line to the limits of visual resolution (Breitmeyer, Julesz and Kropfl, 1975).

For random-dot matrices, correlation may be defined by assigning the mean luminance level a value of 0, a bright dot a value of +1 and a dark dot (no bright dot) a value of −1. Now multiplication of two matrices (the mathematical operation for correlation) gives the result of 1 whenever a matrix position is filled with two bright dots or with two dark dots ($-1 \times -1 = 1$), but a value of −1 when the elements differ ($-1 \times 1 = -1$). If all the positions are identically filled in the two matrices (either bright or dark) the correlation has a value of $(1 \times n)/n = 1$ (for n matrix elements). If the two matrices are randomly filled on average 50% will be identical and 50% opposite in contrast. The correlation will then have a value of

$$\frac{1 \times n/2}{n} + \frac{-1 \times n/2}{n} = 0.$$

If half the elements are identically filled but half are randomly filled the correlation will be

$$\frac{1 \times n/2}{n} + \frac{1 \times n/4}{n} + \frac{-1 \times n/4}{n} = 0.5.$$

In general the correlation equals the proportion of matrix positions that are identically filled when the remainder are randomly filled.

An interesting feature of spatial corss-correlation is that it is essentially a parallel process in which all elements of one input channel are compared with all elements of a second input channel. This is true even though the parallel process may be computable in a serial fashion. The serial/parallel dichotomy is relevant when considering the organization of neural processing using elements that take a finite time to pass information. In this case a parallel process will be much faster than the equivalent serial operations. The parallel nature of the process is particularly evident when the cross-correlation occurs between two random matrices which are organized only in terms of a correlation between them. This is a case where there is no organized information prior to the cross-correlational process, and hence it dramatizes the cross-correlation operation. An example of this type of process in human perception was developed by Julesz (1960) in his random-dot stereograms for binocular depth perception. In the original static stereograms a correlation change in the form of a binocular disparity could be detected in a 50 msec presentation (Julesz, 1964). This may seem brief enough in terms of everyday perception, but in comparison to the 1 msec required for a neuron to fire, 50 msec provides time for a good deal of neural processing of a serial nature.

In some recent experiments, observers were fusing two identical arrays of dynamic noise (zero disparity) and were instructed to report the collapse of fusion of a small target (24 × 24 arc min) whose disparity was changed suddenly from zero (Breitmeyer *et al.*, 1975; Julesz *et al.*, 1976). In these experiments observers were instructed not to report the target in depth, but only to detect the transition from binocular correlation to uncorrelation in the zero disparity plane at minimum target duration. While this duration varied greatly within 1° eccentricity from a fixation marker, several observers had duration thresholds under 18 msec, that is about a third of the duration for perceiving depth. Since in one of these studies (Julesz *et al.*, 1976) similar results were obtained with static noise (provided the static noise array was changed to a new one at the onset and offset of the target in order to avoid monocular movement cues), the reduced duration threshold did not occur because of dynamic noise, but presumably of the simpler criterion of detecting correlation change (instead of depth change). However, the target size in these studies was relatively small, and it seems of considerable theoretical and practical interest to determine the minimal duration for detecting correlation changes with increased target size.

In this paper we report detection of binocular correlation changes for presentation times as low as 2.5 msec when a decorrelation is presented in dynamic visual noise (DVN). Since this brevity in detection is obtainable only with large fields containing several hundred dots, it seems that only an explicitly parallel process would be fast enough to make all the necessary correlations. Furthermore, the possibility of storage of the monocular stimuli in the form of an afterimage or iconic memory is made very unlikely because the DVN continues after the stimulus, effectively masking all preceding events in that region of the visual field after very few successive frames.

METHOD

The apparatus consisted of a Hewlett-Packard cathode ray tube display controlled by a PDP 11/20 computer. This is capable of producing dynamic binocular stimulation with matrices of 30 × 30 adjacent dots at repetition rates up to 666 frames/sec. The matrices subtended 5° of visual angle under our viewing conditions. They could be filled randomly with a specified density of dots which had an average density of 6%. The degree of correlation between the dot positions in the two eyes could be varied from +1 through zero to −1. A correlation of +1 (correlated DVN) appears as a flat plane of randomly moving dots. A correlation of zero (uncorrelated DVN) appears as a confused impression of a three-dimensional time-varying snowstorm. The display had an average luminance of 3.0 cd/m² measured on a Spectra-spot meter.

Experiments consisted of presenting an initial correlation state of 600 msec duration which briefly changed to a different test correlation state. The duration of the test correlation state plus the return to the initial state added up to a total of 1500 msec. A change from correlation to uncorrelation and back will be called a decorrelation task, whereas a change in the reverse direction will be called a correlation task. The square area of change in correlation was in the center of the matrix and could be varied in size by the experimenter while the remainder of the matrix did not change. For the smallest area only 2 × 2 dots were changed, but at 1.5 msec per frame many dot combinations could occur within the 50 msec required for best detection. Threshold duration of the correlation state change for detection of stimulus change was measured by a forced-choice randomized-staircase method (Julesz and Tyler, 1976) in which a stimulus or no stimulus were presented randomly each with a probability of 0.5. The observer's task was to detect the occurrence of the event based on any stimulus cue.

No monocular event occurs specific to the binocular correlation change test period, and trials with one eye occluded demonstrated that no threshold could be obtained for durations up to 1000 msec. The observers reported that the appearance of the decorrelation task at threshold was a flash of some kind of brightness or salience of the stimulus, rather than any impression of depth. On the other hand, the appearance of the correlation task at threshold was of a change in the way the dots appeared to move, becoming slower and more cohesive when the correlated plane was detected, rather than perceiving the depth plane as such.

The two observers were both fully corrected myopes with visual acuity of 20/20 and good stereopsis on the basis of random dot stereogram tests. Both were well-practiced in the task and aware of the purpose of the experiment, although neither knew what form the data would take prior to his threshold judgements.

AREAL INTEGRATION

The first question we ask is what is the retinal area over which binocular cross-correlation can be accomplished? This may be estimated by measuring detection as a function of stimulated retinal area. The point beyond which increasing area does not improve detection should indicate the area over which the cross-correlation is integrated. Areal integration data are shown for two observers in Fig. 1. Each point represents the mean of approximately 100 observations, which were obtained in full randomized order over several days of observation. The filled circles

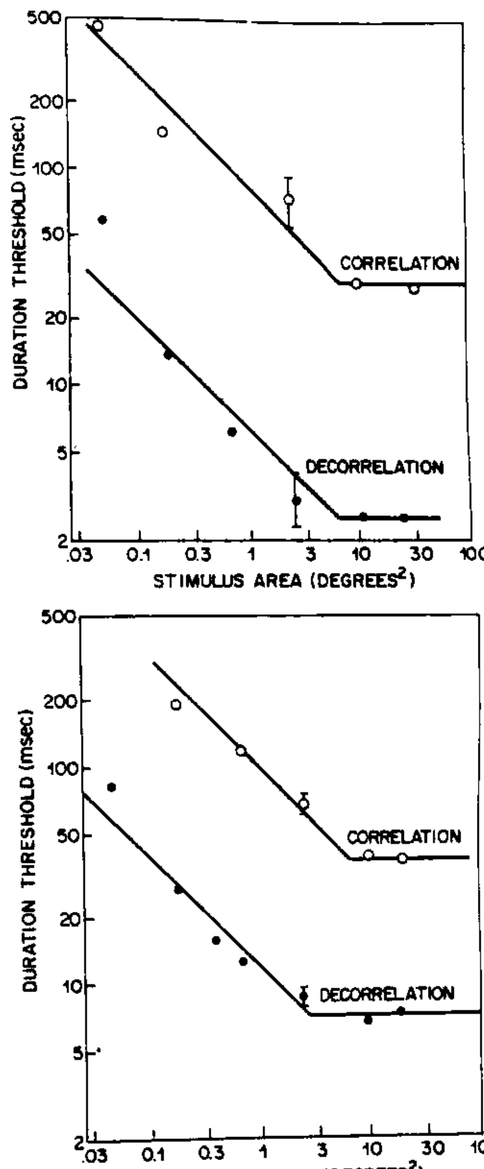

Fig. 1. Duration thresholds for detection of binocular correlation (open circles) and decorrelation (filled circles) of visual noise as a function of stimulus area for two observers (upper and lower figures).

show duration thresholds for detection of the occurrence of uncorrelation from a pre- and post-test field of correlated DVN. This areal integration appears to continue up to some value after which threshold becomes independent of area. The data are therefore fitted by inspection with a linear segment with a slope of -0.5 and a second segment with a slope of 0. (The data are not sufficiently extensive to justify use of a more elaborate, second order curve). The fit of the data to such two-segment curves has a mean standard deviation of 0.13 log units (averaged over both observers in both tasks), whereas if the points had been fitted by a single line with a slope of -0.5 the mean standard deviation would be nearly doubled (0.25 log units). Thus the construct of a maximum integration area is supported by the improvement in fitting the data. (The slope of -0.5 is considered below.) It is clear that areal integration continues up to about 3–5 square degrees for both observers. This corresponds within the accuracy of the data to the extent of the human fovea (2° dia or π square degrees), so it is probable that the limit of integration is set by the acuity limitations of the retinal projection system.

Areal integration was also measured for the opposite task, detection of a correlated test area in a pre- and post-test field of uncorrelated DVN (Fig. 1, open circles). It is notable that this task requires an increase in presentation time of a factor of 10 or more. This increase is required even though the discrimination during the test period is identical for both tasks: namely whether the test area was correlated or uncorrelated. Despite the radical change in detection duration, integration area for the detection of correlation in uncorrelation remained almost unchanged, with a value of about 5 square degrees for the two observers.

ANALYSIS OF THE BINOCULAR CROSS-CORRELATION PROCESS

These data indicate that there is a spatiotemporal trade-off in the cross-correlational process up to a spatial limit of 2°–3° dia. However, it is not a true reciprocity since the data are well fitted by lines with a slope of -0.5 in logarithmic coordinates. This means that the duration increases with the square root of the reciprocal of stimulus area ($\propto$ the number of dots/sec). Such a relationship suggests that detection is operating as a signal/noise discriminator for events between each corresponding pair of dot positions, since the theoretical limit for signal discrimination increases with the square root of the number of events over which discrimination may be integrated.

The marked difference between detection of correlation and of decorrelation raises the question of the underlying basis of such performance. The first possibility is that the difference is connected with depth information perceived in the stimulus when uncorrelated noise is observed for long durations. Under the concept the uncorrelated DVN can be considered as consisting of a set of random disparities, each point in one eye being at some horizontal disparity from any point in the same horizontal line in the other eye. A set of cortical disparity detectors would then process the random disparities to produce a multiple depth image. Fully correlated noise appears as only a single depth plane.[1] If disparity detection provides the basis for detection of changes in correlation, it could be argued that it should be easy to detect stimulation of multiple depths following stimulation at only one depth (the decorrelation task) because most depths were previously unstimulated. On the other hand the correlation task requires detection of one depth when all depths were previously stimulated.

There are at least two arguments against this position. Firstly, if the visual system is capable of making random disparity correlations there should be nearly as many of these in the correlated stimulus as the uncorrelated stimulus. The main difference is that the correlated stimulus also contains full correlation in the zero depth plane. This description is diagrammed in Fig. 2a, where the distribution of elements in the left and right-eye matrices are shown as the square blocks. The distribution of possible disparity correlations in space (assuming that disparity is only detectable between elements of the same sign of contrast (Julesz, 1964)) is shown by the circles, dark or light according to the sign of contrast. Figure 2b shows the equivalent diagram for binocularly uncorrelated visual noise. In this case only 50% of the elements are correlated in the plane of fixation. There are also proportionately more disparate elements, but this increase in number is unlikely to be significant. The number of disparate elements in correlated noise is, r^2-r, where r is the number of elements in a horizontal row, whereas in uncorrelated noise this changes to r^2-$r/2$. When r is at its minimum of 2 this change is only from 2 to 3, or a 50% increase, and it becomes a minute 1.7% as r increases to 30. Such a small change is unlikely to be important in comparison to the 100% increase in correlated elements in the plane of fixation with change from uncorrelated to correlated noise. Furthermore, the data show that the correlated/uncorrelated difference remains approximately constant across stimulus area, with the possible exception of the very smallest area of 0.4°.

The second argument against the range of depth planes as an important factor in the correlation/decorrelation difference is that the correlation depth plane can be well cued without improving detection. The correlation data for observer Z were obtained with a fixation point and a correlated spatial surround, both in the depth plane in which the uncorrelated noise became correlated (which was the zero disparity plane). The task is repeated many hundreds of times during the session, so the observer is very familiar with the expected appearance of this plane. Nevertheless the duration threshold remains close to a log unit above that for the decorrelation task.

These two arguments make it difficult to maintain that the correlation/decorrelation difference is related to the possible stimulation of non-zero depth planes.

[1] The term "depth plane" is used to connote a plane of some constant disparity relative to the zero plane defined by the fully correlated binocular display matrix. This will not correspond exactly to zero retinal disparity for a variety of physiological reasons too complex to enumerate in a footnote [see Helmholtz H. von (1910) *Handbuch der Physiologische Optik. III* (Translated by Southall J. P. C. Dover, New York. 1925) Chap. 31].

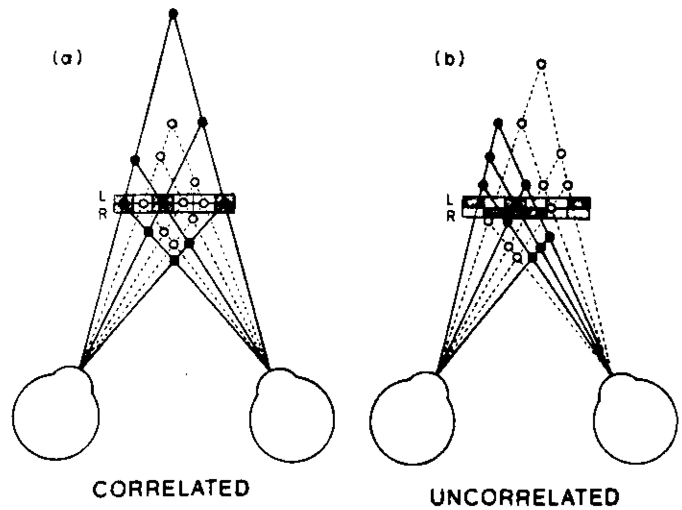

Fig. 2. Depiction of the possible disparity relationships present in correlated and uncorrelated noise. Block elements represent examples of a line of the left and right stimulus matrix. Open and closed circles represent the geometric depth of like-contrast binocular stimulation.

which presumably leaves the conclusion that the critical events are those occurring in the zero disparity plane. These events are defined in terms of the number of correlated points or the degree of correlation rather than the range of disparities, and justify the emphasis placed in this paper on the binocular cross-correlation process in the zero disparity plane. A more detailed analysis of the mathematical properties of this process has been presented elsewhere (Julesz and Tyler, 1976; Tyler and Julesz, 1976).

One point that can be considered here is the robustness of the correlation/decorrelation difference when changes occur between different degrees of partial correlation (i.e. decorrelation of only a small percentage of the positions in the matrix). Does the difference arise from a nonlinearity in the neural response to different degrees of correlation, or does it occur for any degree of correlation change? To answer this question we modified the apparatus so that it could present any degree of correlation between the dots presented to each eye in either the test or the pre- and post-test periods. We then measured duration thresholds for either changes from full correlation to some intermediate degree of correlation ($C\rightarrow$), which was varied as an experimental parameter, or the reverse change from an intermediate correlation to full correlation ($C\leftarrow$). The stimulus area was set at 2.5° × 2.5° for maximum sensitivity, and a new density of 5%/frame.

The results are shown in Fig. 3, for one observer with the percent change in correlation either from ($C\rightarrow$) or to ($C\leftarrow$) full correlation along the abscissa. The logarithmic difference between detection of correlation and decorrelation is maintained for all degrees of change down to 1%. This difference therefore seems to be inherent in the detection process, since it is unaffected either by area or degree of correlation. It must represent some kind of *adaptation* to the current state of the DVN, since the extent of the change for detection should be the same regardless of the direction of change if no adaptation were occurring. This adaptation may be analogous to the Weber law adaptation to stimulus intensity, where the increment threshold is proportionately larger when the adaptation level is high than when it is low. Increments and decrements from the same adaptation level have similar threshold values.

These observations—particularly the large difference between correlation and decorrelation tasks—may be explained qualitatively in terms of the dipole model of stereopsis (Julesz, 1971). For DVN the magnetic dipoles are regarded as electromagnets, that change their polarity at a high rate. When an observer converges at a fixation marker, say at zero disparity,

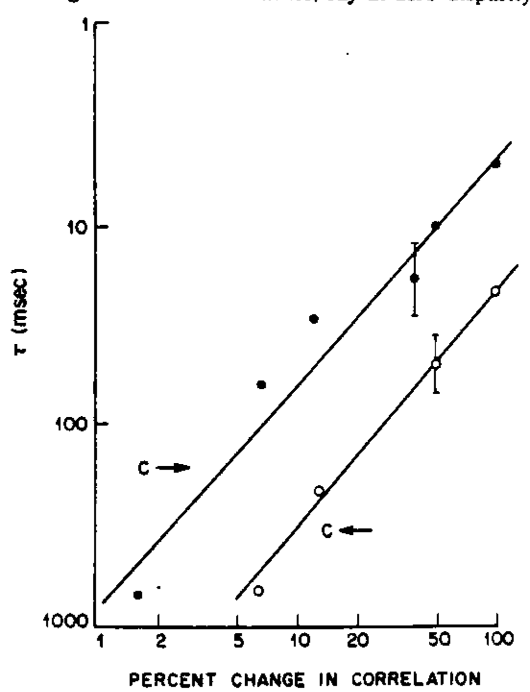

Fig. 3. Duration thresholds for degrees of binocular correlation, changing from full correlation to the value on the abscissa $C\rightarrow$, filled circles) and for the reverse change ($C\leftarrow$, open circles) for one observer (CWT).

it takes some time for the correlated left and right dipole-arrays to interlock (by finding the correct amount of shift that brings the corresponding dipoles of the two arrays within their near-fields). Then decorrelation from an already correlated state simply means that the interlocked dipoles of the left and right dipole-arrays suddenly repel each other (at least 50% of them). This sudden repelling of half the dipoles and the spring-coupling between neighboring dipoles that forces the remaining interlocked dipoles to unlock, results in a rapid collapse of the interlocked state.

On the other hand, when the arrays are initially uncorrelated, it takes considerable time for the dipole-arrays to interlock. In the uncorrelated state, just when correlation starts, half of the dipoles are interlocked, half repel each other, and by doing so might often face some neighboring dipole of the same kind (i.e. opposite polarity) and by chance interlock. These turned interlocked dipoles correspond to depth organizations outside the zero disparity plane. Then, if the arrays become correlated (zero disparity plane) it takes some time to drag away the interlocked dipoles (corresponding to non-zero disparities) through the spring coupling exerted by the pool of dipoles that by chance were correctly facing their partners at the moment when correlation started.

CONCLUSION

The data described show a remarkable sensitivity in the binocular correlation process. Detection is possible down to correlation changes of only about 1%, and, under different conditions, of changes lasting only 2–3 msec in fields of continually changing random noise. These results attest to a rapid, parallel operation of a neural cross-correlation process of unexpected precision and sensitivity. Possible reasons why this study showed more rapid cyclopean processing than earlier studies (e.g. Julesz, 1964) are (a) that decorrelation is an unusually strong stimulus and (b) that because it only requires detection of the collapse of fusion from a single depth plane the area over which correlation change can be integrated is an order of magnitude greater than has been previously used in DVN tasks (Breitmeyer *et al.*, 1975).

Note added in proof. A recent study of the minimum duration for stereopsis in DVN (Uttal, Fitzgerald and Eskin, 1975) contains a report of consistent stereoscopic discrimination for presentation times of 5 msec. However, this was found without the use of masking DVN before and after the stimulus epoch. Addition of masking DVN brought performance close to chance levels for 20 msec presentation times. This is therefore approaching an order of magnitude greater than the duration required for the pure cross-correlation task reported here.

Acknowledgement—Our thanks to W. J. Kropfl for extensive technical and experimental assistance.

REFERENCES

Licklider J. C. R. (1948) The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Am.* **20,** 150–159.

Reichardt W. (1961) Autocorrelation: A principle for the evaluation of sensory information by the central nervous system. In: *Sensory Communication* (Edited by Rosenblith W. A.), Ch. 17. Wiley, New York.

Breitmeyer B., Julesz B. and Kropfl W. J. (1975) Dynamic random-dot stereograms reveal an up-down anisotropy and left-right isotropy between cortical hemifields. *Science* **187,** 269–270.

Julesz B., Breitmeyer B. and Kropfl W. J. (1976) Binocular-disparity-dependent upper-lower hemifield anistropy and left-right hemifield isotropy as revealed by dynamic random-dot stereograms. *Perception* **5,** 129–141.

Julesz B. (1960) Binocular depth perception of computer-generated patterns. *Bell Syst. Tech. J.* **39,** 1125–1162.

Julesz B. (1964) Binocular depth perception without familiarity cues. *Science* **145,** 356–362.

Julesz B. and Tyler C. W. (1976) Neurontropy, an entropy-like measure of neural correlation, in binocular fusion and rivalry. *Bio. Cybernetics* **23,** 25–32.

Tyler C. W. and Julesz B. (1976) The neural transfer characteristic (neurontropy) for binocular stochastic stimulation. *Bio. Cybernetics* **23,** 33–37.

Uttal W. R., Fitzgerald J. and Eskin T. E. (1975) Parameters of tachistoscopic stereopsis. *Vision Res.* **15,** 705–712.