SPEEDED PHASE DISCRIMINATION :
EVIDENCE FOR GLOBAL TO LOCAL PROCESSING

J.M. Barr
Experimental Psychology Department
University of Oxford. U.K.

Abstract

In recent years much attention has been given to the advantages of multiple resolution pre-processing methods in computer vision. There is strong evidence that the parallel extraction of luminance changes over different spatial scales also occurs in human visual perception. The first experiment confirms the strong evidence that responses to low resolution signals can be elicited as much as 100 msec faster than to high spatial frequency stimuli of the same contrast. A further experiment measured the reaction time to discriminate the relative phase of the higher frequency component of a luminance grating comprising a fundamental and its second harmonic. It was found that the decision can be made more rapidly when the fundamental is low than high frequency. On the assumption that the gross structure of spatial forms is conveyed by low spatial frequencies, this supports the idea that the substrate for the subjective impression that rough descriptions of visual forms precede detailed perception, is the progressive increase in response time with frequency, of the visual mechanisms implementing spatial filtering.

Current Address :
THORN EMI Central Research Laboratories.
Hayes. Middlesex. U.K.

Key Words :

Visual Perception, Phase Discrimination,
Multi-resolution Image Processing

Introduction

Within most natural scenes, intensity changes occur over a range of spatial scales, so it has been suggested that a general purpose vision system may require some form of early representation that captures and makes this explicit. Over the past decade a number of multi-resolution schemes have been described in the image processing literature, [1,2,3,4,5,6].

Over the same period and independently, considerable empirical evidence has been amassed from both psychological and neurophysiological research, supporting the idea that a form of multiple resolution representation is employed in the early processing stages of biological visual systems. It is asserted that at each point in the retina there exist several contrast sensitive mechanisms each detecting luminance changes over a different spatial scale. As spatial scale is equivalent to spatial frequency (for signals containing a single frequency component) the neural mechanisms can be described as filters in the spatial frequency domain. The simplest model of the filter impulse response is the difference of two circularly symmetric Gaussian functions. The output of a set of filters tuned to the same frequency, covering the entire retina is termed a channel, and is equivalent to the parallel convolution of the image with a single operator. Each channel is assumed to act independently of any other, hence this processing scheme is known as the multiple independent channels model [7].

The independent channels model has been remarkably successful in predicting contrast thresholds in a variety of laboratory experiments, and in generating a great deal of further research attempting to specify the spatial filter characteristics. However little attention has been given to their organization, function or utility. A description of the properties of individual mechanisms is insufficient to explain the fundamental problems of shape description and object recognition.

Early papers proposed that the hypothesised neural filters implement spectral analysis. If phase information is discarded it would then be possible to perform pattern recognition with translational invariance, and if reference templates are stored and matched in terms of the ratio of frequencies, with size invariance [8,9]. However there are several objections to this notion. Spatial phase is of cardinal importance to the visual system. It is still possible to recognise objects if amplitude information is discarded by equalizing all components, provided phase information is preserved [10]. The one octave bandwidth of the filters is too great to permit high resolution spectral analysis, and furthermore, while Fourier Transform based methods may permit the recognition of simple shapes presented in isolation, it would be much more difficult to acheive more complex tasks such as scene segmentation in the frequency domain.

Marr and Hildreth [11] propose a space domain function for spatial frequency filters, in which at each image location, mechanisms tuned to different frequencies provide independent evidence of luminance changes over different scales. This information is then integrated in local oriented edge detector units. Critical problems with this scheme are the alignment of the output from different operators, its performance on blurred edges and its performance in the presence of noise.

In this paper the bank of spatial filters will be considered as a multi-resolution pre-processing front end, which provides a rich description of luminance changes over a range of spatial scales, upon which higher level interpretive processes may operate.

An advantage of pyramidal processing schemes in computer vision is the possibility of information flow in several directions within the data structure [12]. Projection operations permit information acquired at low resolution to guide processing at higher levels. For example in matching applications it seems an efficient strategy to rapidly discard the maximum number of incorrect alternatives by first matching at a coarse level of resolution, saving the more computationally intensive high resolution processs for a reduced set of alternatives. Similarly in shape analysis it is more efficient to first locate the approximate boundaries of the form with a coarse analysis before focussing local operations on optimal regions. Data flow in the opposite direction involves the integration of high resolution information and its reduction to lower levels, eg. block quantization. Lateral processing is restricted to a single level. It is not known whether there are such interactions between the levels of the multiple resolution stucture of the human visual system but it is well established that the temporal properties of low spatial frequencies channels differ from those tuned to higher spatial frequencies in that the former mechanisms exhibit a greater sensitivity to temporal transients [14]. Low spatial frequency mechanisms behave as derivative operators to temporal changes in contrast while high spatial frequency mechanisms operate as temporal integrators. Hence the possibility is raised that the output of each spatial frequency channel is produced asynchronously.

Intuitively it appears that on first glance of a scene, we obtain an immediate rough impression of the approximate forms, locations and extents of the principle objects. Perception rapidly becomes more detailed over time [14]. This form of global precedence can be interpreted in terms of the independent channels model as the progressive acquisition of representations of increasing spatial frequency. Immediately after stimulus presentation only the blurred output of low spatial frequency filters is available, and over a fraction of a second the image representation is sharpened by the addition of higher frequency components. It is not known whether this phenomenon has any utility in terms of neural implementations of projective multi-resolution algorithms, or whether it is merely a processing bottleneck, a dysfunctional epiphenomenon.

The idea of asynchronous parallel channels has several interesting consequences, two of which will be addressed empirically in this paper. The first is that the visual detection of stimuli containing lower spatial frequencies should be faster than of stimuli containing only higher frequencies. Secondly, visual discrimination should be faster if the stimuli differ in their low spatial frequency content than if they only differ in their high frequency content.

Experiment 1

The first experiment measures simple reaction time (sRT) to the presentation of sine wave gratings, replicating a result obtained by Breitmeyer [15]. The stimulus is the abrupt appearence of a luminance grating displayed on a CRT screen. The observer's task is to indicate his detection of the change from a blank to a luminance modulated screen by pressing a microswitch as quickly as possible.

A very simple model of the subject's performance in the reaction time task comprises two components, a sensory, and a motor stage. On the abrupt presentation of a visual stimulus it is assumed that only those spatial filters tuned to the stimulus produce a sensory response, and that their response magnitude follows some growth function, increasing over a short duration following stimulus onset. The decision to press the microswitch is taken when the perturbation of the sensory output function exceeds a criterion value. The duration of motor response is assumed to be constant (approx. 100 msec), independent of sensory factors including the spatial frequency of the stimulus.

The total sRT is the sum of the durations of these two stages. Any variation in the empirically obtained latencies with stimulus spatial frequency can therefore be entirely attributed to differences in the time taken for the outputs of the spatial filters to exceed some criterion level. The stimulus is a one dimensional luminace function sinusoidally modulated about a mean level. This is spread vertically on the screen by a high frequency oscillator to give the appearence of a vertical grating.

Results for one subject at two contrast levels are shown in Figure 1. It should be possible to obtain estimates of sensory latency from the sRT values by subtracting from them the constant motor time. It can be seen that there is indeed an increase in the time course of response with increasing stimulus frequency. A response to the detection of a low contrast 9 cpd grating is not made until up to 100 msec after that to a 0.5 cpd stimulus of the same contrast. It can also be seen that reaction time decreases with increasing stimulus contrast and the extent of the spatial frequency based time difference diminishes. On the basis of the assumptions made above, this can be taken to imply that spatial filters tuned to higher frequencies have longer latencies than those tuned to lower frequencies.

Experiment 2

Having confirmed the evidence for temporal asynchronies between spatial frequency channels in a detection task, it was decided to investigate whether the effect transfers to discrimination. Most natural visual stimuli do not comprise a single frequency component but complex spectra. Furthermore it is the phase rather than the amplitude spectrum that determines shape recognition [10]. Therefore a choice Reaction Time (cRT) task was chosen requiring the discrimination of the relative phase of a two component compound grating. If channel asynchrony imposes deterministic delays on the transmission of information then decisions based on higher frequency information should be delayed relative to those based on low spatial frequency information.

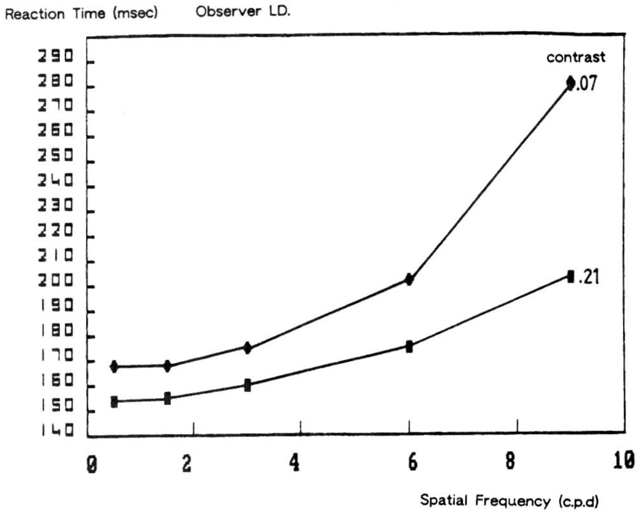

Figure 1. Reaction Time to sinusoidal gratings for one subject at 2 contrast levels

The stimuli used in the experiment were the first two sinusoidal components of a square wave, added in either square wave or triangle wave phase. The luminance at each point, L(x) is given by Equation 2:

```
L(x) = Lm + c sin(2Pif + Phi1)   (Equat-
        + c/3 sin(2Pi3f + Phi2)    ion 2)
```

where Lm is mean luminance, and c,f and Phi are contrast frequency and phase respectively. Their luminance profiles are shown in Figure 2. Observers had to identify each stimulus as rapidly as possible. The model for the performance of the cRT task is similar to that for sRT except that a discrimination process must intervene between sensory detection and the initiation of the motor response. A motor response can only be initiated after the detection of the third harmonic and the discrimination of its phase relative to the fundamental. Assuming the speed of phase discrimination is constant with spatial frequency, it should not be possible to make the discrimination until the higher frequency information has reached the

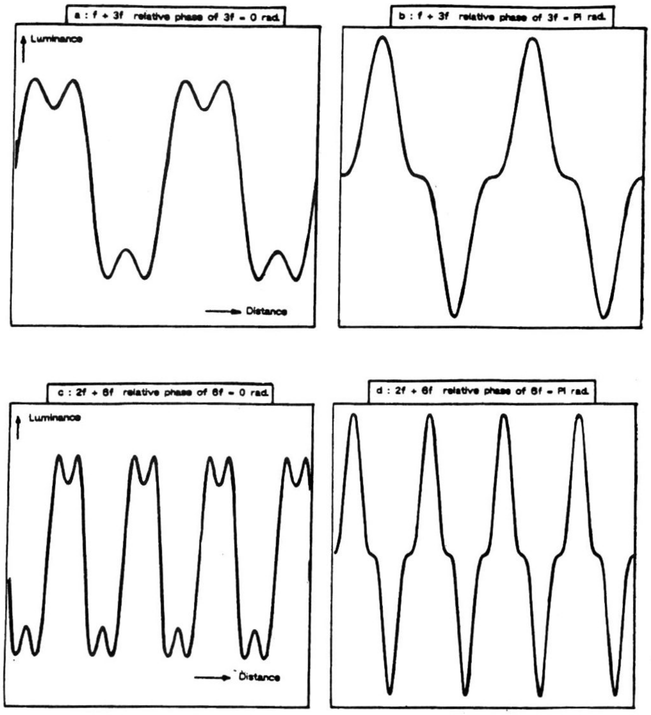

Figure 2 Examples of the luminance profiles of the stimuli used in Experiment 2.

decision mechanism. Therefore there should be an increase in choice reaction time with frequency that should depend solely on sensory delay and should increase at the same rate as reaction time to detect the high frequency component presented alone.

Each stimulus was the sum of two sinusoids separated in frequency by a factor of three. In each block of trials the same fundamental frequency was always presented, but the relative phase of the harmonic randomly varied between either 0 or Pi on each trial with equal probability. The subject was given two response keys, one assigned to each alternative, and instructed to indicate the perceived phase relationship as rapidly as possible by pressing the appropriate switch. The subjects were also presented with the third harmonic stimuli in isolation in a sRT task.

Figure 3 shows the results for 4 subjects. The increase in sRT with the frequency of the third harmonic stimuli follows the pattern of Experiment 1. The results for the cRT task are plotted in terms of the frequency of the third harmonic. Phase discrimination for the low frequency compound stimuli appears to take approximately 80 msec longer than detection of its slower component. If this reflected the additional complexity of the phase discrimination decision then reaction time to the high frequency compound would be expected to show a similar additional delay. However the surprising result is that the the relative phase of the third harmonic can be discriminated as rapidly as the detection of the same stimulus presented in isolation.

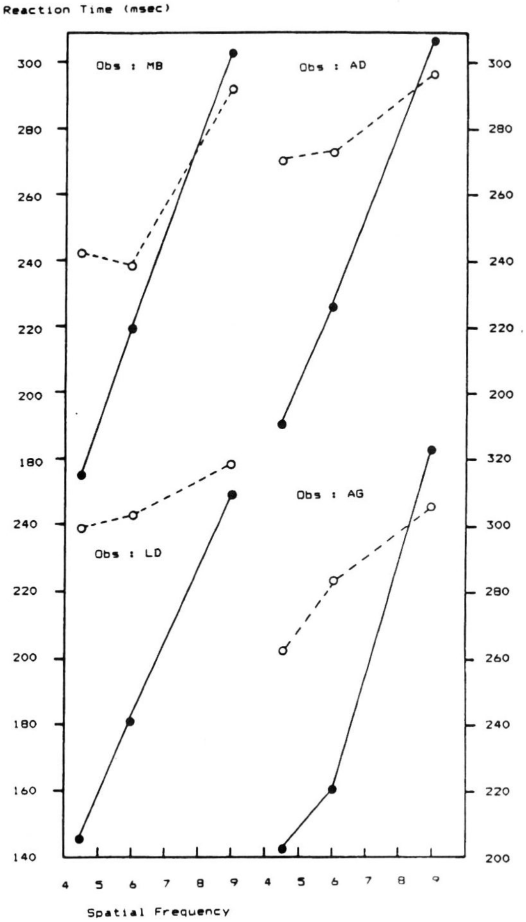

Figure 3. Results of Experiment 2 illustrating Reaction Time as a function of Spatial Frequency for 4 subjects.

Closed circles represent mean simple Reaction Times to single component gratings.

Open circles represent mean choice Reaction Times for correct responses in the speeded phase discrimination task. The symbols are plotted against the frequency of. the third harmonic component of the compound stimulus.

It appears that the slope of the cRT function is more nearly parallel to that of the sRT function if plotted in terms of the frequency of the fundamental. rather than the third harmonic. The shallower slope follows the shallower increase in sRT with frequency at lower spatial frequencies and higher contrasts. One possible explanation for the anomalous result of Experiment 2, therefore is that the performance of the task is based not upon discrimination of relative phase but upon discrimination of the peak to peak amplitudes of the

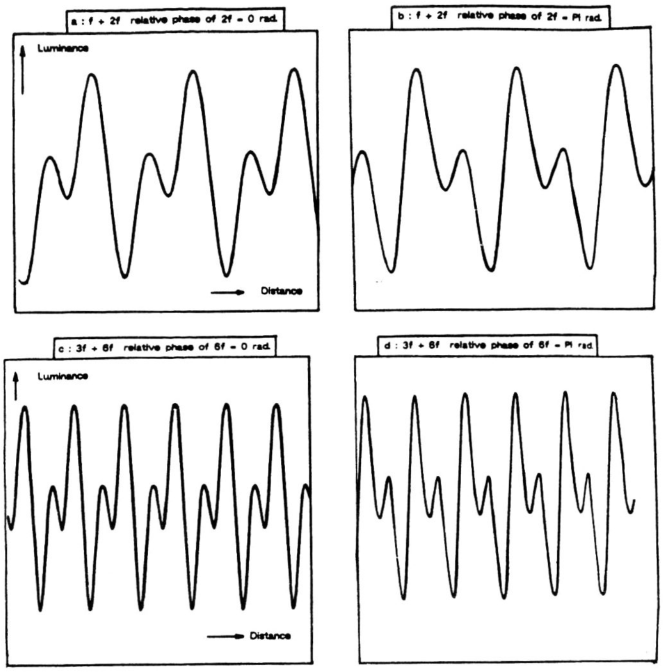

Figure 4 Examples of the luminance profiles of the stimuli used in Experiment 3.

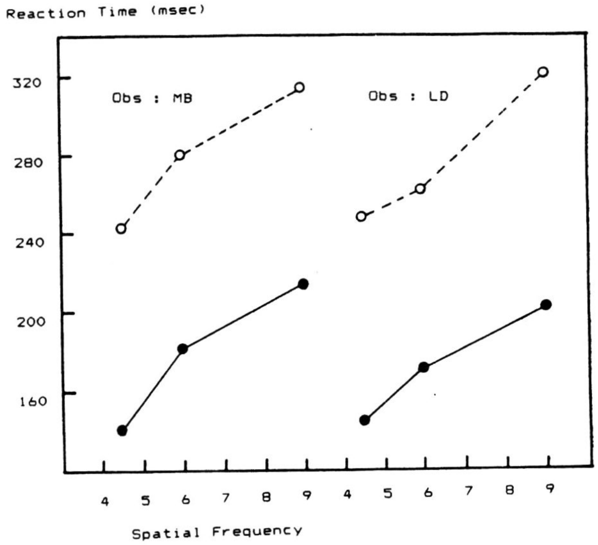

Figure 5. Results of Experiment 3 illustrating Reaction Time as a function of Spatial Frequency for 2 subjects.

Closed circles represent mean simple Reaction Times to single component gratings.

Open circles represent mean choice Reaction Times for correct responses in the speeded phase discrimination task. The symbols are plotted against the frequency of. the second harmonic component of the compound stimulus.

waveforms. Although the contrasts of the two frequency components are the same in both phase relationships, the ratio of the difference of image maxima and minima between the square wave and triangle wave is 1.41. So subjects could have been responding to the stimulus on the basis of global contrast rather than relative phase. Therefore it was decided to repeat the experiment with a condition in which peak to peak contrast is controlled.

Experiment 3

In this experiment the stimulus is a compound of a fundamental and a second harmonic of equal contrast. The the luminance at each point is given by Equation 3 :

$$L(x) = Lm + c \sin(2Pifx + Phi1) + c \sin(2Pi2fx + Phi1) \qquad \text{(Equation 3)}$$

The luminance profile of stimuli in the two phase relationships and two fundamental frequencies are shown in Figure 4. It can be seen that global contrast cannot be used as a cue as the stimuli are related by a reflection. sRT to the second harmonic stimuli presented in isolation was also measured.

The results for two subjects are shown in Fig 5. The increase in cRT is parallel to the increase in sRT with spatial frequency. Thus when global contrast is removed as a cue there does indeed appear to be an increase in the time taken to perform a discrimination task when the decision has to be based upon the high rather than low spatial frequency content of the stimulus.

Discussion

Experiment 1 provided strong evidence for the idea of asynchronous detection of spatial frequency components when presented in isolation. Low frequency stimuli can be detected up to 100 msec faster than high frequency stimuli of the same contrast. That contrast also increases detection time further supports the idea of global precedence, as the amplitude spectra of most natural stimuli are low pass. Thus the idea of asynchronous channel operation can be viewed as providing a computationally explicit explanation of the intuitive observation of global precedence.

Assuming it is possible to predict the detectibility of any stimulus at a constant mean luminance given a knowledege of the frequency spectrum of the stimulus and the observer's modulation transfer function, it should similarly be possible to predict visual conspicuity and the growth of detail perception over the first few moments of inspection of any static visual stimulus, from a knowledge of its Fourier spectrum and the observer's reaction time performance. The results of Experiment 3 are also encouraging, illustrating that frequency related detection latency differences transfer to a discrimination task.

However the analogy between computer vision and human multi-resolution systems is far from perfect. One major problem is raised by the results of Experiment 2. While there is good evidence for individual spatial frequency components being detected independently at threshold contrast, the evidence is not compelling for suprathreshold stimuli. Assuming that the channels have a bandwidth of approximately 1 octave it should not be possible to integrate energy from two frequency components separated by a factor of 3. That this appears to have occurred in Experiment 2, enabling global contrast to be used as a cue in the discrimination task, implies that channel bandwidth in some way increases with stimulus contrast. If this is so it seriously weakens the idea of independent channels.

References

[1] Rosenfeld, A. (1980) Multi-resolution Image Processing and Analysis. Springer-Verlag, Berlin, FRG.

[2] Hall, E.L., Wong, R.Y., and Rouge, J. (1977) Sequential Scene Matching with Hierarchical Search. Proc. IEEE Southeast Conference, Williamsburg, Va., U.S.A. pp.402-405.

[3] Witkin, A.P. (1984) Scale Space Filtering : A New Approach to Multi-scale Description. ICASSP 39A.1.1-4.

[4] Baker, K.D. & Sullivan, G.D. (1980) Multiple Bandpass Filters in Image Image Processing. IEE Proc. Vol 127 Part E(5) pp.173-184.

[5] Burt, P.J. & Adelson, E.H. (1983) The Laplacian Pyramid as a Compact Image Code. IEEE Trans. Communications Vol COM-31(4) pp.532-540.

[6] Crowley, J.L. & Parker, A.C. A Representation for Shape Based on Peaks and Troughs in the Difference of Low-pass Transform. IEEE Trans Pattern Analysis and Machine Intelligence Vol PAMI-6(2) pp.156-170.

[7] Campbell, F.W. & Robson, J.G. (1968) Application of Fourier Analysis to the Visibility of Gratings. J. Physiology (London) Vol 197, pp.551-566.

[8] Blakemore, C. & Campbell, F.W. (1969) On the Existence of Neurones Selectively Sensitive to the Orientation and Size of Retinal Images. J. Physiology (London) Vol 203 pp.237-260.

[9] Pollen, D.A., Lee, J.R. & Taylor, J.H. (1971) How Does the Striate Cortex Begin the Reconstruction of the Visual World ? Science Vol 173, pp.74-77.

[10] Piotrowski, L.N. & Campbell, F.W. (1982) A Demonstration of the Visual Importance and Flexibility of Spatial Frequency Amplitude and Phase. Perception Vol. 11 pp.337-346.

[11] Marr, D. & Hildreth, E. (1980) Theory of Edge Detection. Proc. Roy. Soc. London B, Vol. 207 pp.187-217.

[12] Hanson, A.R. & Riseman, E.M. (1980) Processing Cones : A Computational Structure for Image Analysis. In Structured Computer Vision by Tanimoto, S. & Klinger, A. Academic Press, New York.

[13] Robson, J.G. (1966) Spatial and Temporal contrast senstivity functions of yje Visual System. J. Opt. Soc. Am. Vol. 56, pp.1141-1142.

[14] Navon, D. (1977) Forest before Trees : The Precedence of Global Features in Visual Perception. Cognitive Psychology, Vol 9, pp.353-383.

[15] Breitmeyer, B.G. (1975) Simple Reaction Time as a measure of the Temporal Response Properties of Transient and Sustained Channels. Vision Research Vol 15, pp.1411-1412.

The work reported here was supported by an S.E.R.C. research studentship.