Copyright © 1975, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

VISIBILITY OF RANDOM PATTERNS AS RELATED TO IMAGE ENCODING

by

Hassan Mostafavi-Kashani

Memorandum No. ERL-M538

25 August 1975

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Professor David J. Sakrison, for his guidance, support and encouragement throughout the course of this research. Thanks should also go to Matt Halter, Dan Swanson, Zahir Khan and Ron Scudder who served as test subjects for the psychophysical experiments, and to Bruce Romberger and Jeff Schriebman who prepared many of the computer programs used in this research. A special and deep appreciation is reserved for my wife, Mahroo, who was a constant source of encouragement during my graduate studies at Berkeley. I would also like to thank Doris Simpson who expertly typed the final manuscript of this thesis.

Research sponsored by the National Science Foundation Grant ENG72-04198-A01.

Visibility of Random Patterns as Related to Image Encoding

Ph.D. Hassan Mostafavi-Kashani

Dept. of Electrical Engineering and Computer Sciences

Committee Chairman

ABSTRACT

In order to find a distortion measure for image encoding that reflects the properties of the human visual system, a number of psychophysical experiments investigating the detection of pseudo random patterns by human subjects are conducted. A mathematical model for the detection of distortion in still monochromatic images is developed. Estimates for the numerical values of the parameters of the model are found. The model consists of a number of parallel channels each selectively sensitive to a range of spatial frequencies and orientations, with detection taking place when the response of any one of these channels reaches a threshold level. The threshold rms contrast of narrow-band isotropic noise, superimposed on a constant luminance background, is measured as a function of the center frequency of the stimulus. The result is similar to the contrast sensitivity curve for sinusoidal gratings. Maximum sensitivity appears around the frequency 4.5 cycles per degree. In another experiment the visibility of different linear combinations of two narrow-band patterns with different center frequencies is investigated; this confirms the multichannel form of the model. In order to find numerical values

for the parameters of the channel filter tuned to 4.5 cycles/degree, threshold contrast of gaussian pseudo-random patterns that are narrow-band in both spatial frequency and orientation, are measured for various radial and angular stimulus bandwidths. The thereshold contrasts of these patterns as predicted by the model, are computed for various values of the filter radial and angular bandwidths. By comparing the simulation results with experimental results, the radial bandwidth of the filter is found to be around ± 2.5 cycles/degree and the orientational selectivity expressed in terms of the polar angle in the frequency plane is found to be $\pm 10°$. The threshold contrasts of pseudo-random patterns obtained from a family of nonlinear operations on a narrow-band gaussian noise pattern are measured. Using different values of p in the L_p-norm operation, the threshold of these patterns as predicted by the model is computed; the results for p=8 match the experimental data. The threshold contrast of narrow-band noise, multiplicatively combined with a narrow-band background image, is measured for two different background center frequencies as a function of the noise center frequency. The presence of the background image decreases the sensitivity to the noise pattern. This decrease is maximum when the noise and the background have the same center frequency. It is also found that the threshold contrast of the noise is a monotonic increasing function of the background contrast. The effect of the background image is reflected in the model by assuming that the background is processed by a separate channel whose output scales down the response of the channel to the noise. Various forms of the fidelity criteria resulting from this model are discussed. The expression for the resulting distortion measure depends on the image that is to be transmitted as well as the difference of this image and its distorted reproduction.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

CHAPTER 1

INTRODUCTION

1.1. Research Objectives

The transmission of images over communication channels with finite capacity usually results in some form of distortion in the image. This distortion can be due to several factors. Suppose a still monochromatic image is regarded as a function $u(x,y)$, representing the luminance as a function of the spatial variables x and y over some frame $[0,T] \times [0,T]$. The transmission of such an image requires encoding it into a form that is acceptable by the communication channel. For example we may have available a digital communication channel which can transmit up to C binary digits (bits) per second reliably (i.e. with very small probability of error). Therefore the transmission of one image frame within a finite period of time requires that the function $u(x,y)$ be encoded into a sequence of a finite number of bits from which at the receiver an approximation $\hat{u}(x,y)$ of the original image can be reconstructed.

Efficient transmission schemes require that we encode the image into the smallest possible number of bits consistent with some level of the fidelity of the reconstructed image $\hat{u}(x,y)$. Clearly the smaller the number of bits, the coarser will be the quantization of the space of all possible images $u(x,y)$; this in turn results in $\hat{u}(x,y)$ being a poorer approximation of $u(x,y)$. Therefore this process which is referred to as source coding, introduces some inevitable amount of distortion. Further distortion can result from channel errors. Depending on the specific source coding algorithm used, in some cases the channel's error in transmitting a single

bit can cause a substantial amount of distortion in the received image.

In many applications it is necessary to transmit large numbers of images over a communication channel of limited capacity. One such example is in the use of satellite photographs to study the earth's environment. In these applications the efficiency of the source coding algorithm and consequently the trade off between the number of bits per picture frame and the fidelity of the received image plays an important role. Clearly the criteria used in evaluating this fidelity depends on the subjective judgment of the user of the communication system. If we were able to find a non-negative function, $d(u,\hat{u})$, termed the distortion measure, which represents the degree of the dissatisfaction of the user due to transmitting the image $u(x,y)$ and receiving the image $\hat{u}(x,y)$, then the trade off between the number of bits per image frame and the distortion as perceived by the user could be treated quantitatively. This distortion measure enables the designer of the coding and transmission systems to predict the quality of the system from the point of view of the end user rather than building systems and experimentally determining the user's satisfaction.

Such a quantitative measure of distortion would also allow one to apply Shannon's rate distortion theory (Shannon [1]) to the problem of image encoding. Given a distortion measure $d(\cdot,\cdot)$, for an image source with a given probability distribution, Shannon has shown that we can define a function $R(D)$, termed the rate distortion function, with the important property stated by Shannon's source coding theorem. This theorem states that if an encoding and

transmission system yield average distortion less than or equal to D

$$E\{d(U,\hat{U})\} \leq D$$

for some positive D, then regardless of its complexity or the specific encoding algorithm used, it must contain a channel whose transmission capacity is at least $C = R(D)$. Conversely by suitably designing the encoding system and allowing sufficient complexity of the encoder, it is possible to achieve a transmission rate $R(D) + \varepsilon$ with an average distortion of value D, where ε is positive but can be made arbitrarily small.

Although the rate distortion theory does not in general tell us how to design efficient encoding algorithms, it is of extreme importance in that it gives the absolute (but achievable) lower bound to the efficiency of any system and thus provides a point of reference with respect to which the performance of the existing or proposed systems can be measured.

One major problem is to find the distortion measure d(.,.) which does in fact reflect the human subjective sensation of the distortion, and does this consistently for a large class of images and for all levels of distortion. Finding a precise quantitative measure for an imprecise notion such as the dissatisfaction of the user is, to say the least, a very difficult task; we can not even be sure that such a function exists at all.

Suppose we approach the problem in a different way by asking what are the distortions to which the user of the communication system is insensitive. This insensitvity may be due to the very specific objectives of a transmission system, in which the user's

requirements allow certain types of distortion. For these special cases we may very well be able to define a distortion measure which can be used to increase the efficiency of the encoding system appreciably.

More important however is the fact that due to the limitations of the visual perception mechanism of the human being, there is a large class of distortions that are not detectable at all by the user. A trivial example is when the two images $u(x,y)$ and $\hat{u}(x,y)$ differ only in very high spatial frequency components, beyond the resolving power of the visual system. Then no matter how large this difference is it will not be detectable. It is clear that in designing an <u>efficient</u> system, regardless of any other requirements of the user, the limitations of the visual perception mechanism have to be taken into consideration.

The primary objective of this research is to find a mathematical model which can predict the threshold of detection of the distortion in images. In order to construct such a model, a series of psychophysical experiments investigating the detection of pseudo random patterns by human subjects, has been conducted. Based on this model we define a distortion measure which is in agreement with the subjective judgment of the user as far as the limitations of his perceptual mechanism is concerned. It should be pointed out that such a distortion measure is useful as long as the level of distortion is around the threshold of detection; for very visible distortions we are still faced with the problem of subjective judgment and the dependence of user's criteria on the specific goals of the transmission system.

The construction of a mathematical model for the detection of distortion requires a knowledge of the properties of the human visual system. Fortunately the study of both the physiological and psychophysical properties of the visual system started long before the study of image transmission systems. Of specific importance for this research are a number of psychophysical experiments, conducted within the last decade, investigating the visibility of such patterns as sinusoidal gratings, lines and edges. These studies have established some basic facts about the detection of these structured patterns. Based on these results, a survey of which will be conducted in Chapter 2, detection models that use the concept of linear two dimensional filters have been suggested. The application of the results of these experiments to the image encoding problem was suggested by Sakrison [2].

All of the experiments of this research deal with the properties of still monochromatic images. The stimulus patterns used in these experiments were sample functions of random patterns. There are two reasons for choosing pseudo-random patterns. First the distortion caused by transmitting an image over a communication channel often takes the form of an unstructured pattern which does not resemble structured forms such as periodic gratings or lines. The second reason is that for a random pattern, the phase effects of the individual frequency components average out over the frame of the image. Consequently such properties of the image as the ratio of the peak to r.m.s. value are not affected substantially by a change in other parameters such as the bandwidth of the image. While, for example the ratio of peak to r.m.s. value for a compound grating consisting of two sinusoids whose frequencies are close together can be changed

substantially by changing the relative phase of them. Pseudo random patterns were also used by Mitchell [3] who investigated the effect of the spatial frequency components of these patterns on their detectability.

1.2. Overview of the Dissertation

In an effort to find a model for the threshold perception of distortion patterns, we first concentrate on the detection of patterns on a blank (constant luminance) background, thus ignoring the effect of the image that is to be encoded on the visibility of the distortion. In Chapter 2, after a brief description of the processing of images by the visual system, a survey of the past research on the visibility of patterns on a blank background is given. Then based on these results a general structural form for the detection model is presented.

In Chapter 3 the experimental methods and the stimulus images are described. Chapter 4 is devoted to describing a sequence of experiments designed to refine the model that was assumed in Chapter 2. The experiments of sections 4.3, 4.4, 4.5, 4.6 were designed specifically for a quantitative analysis of this model. They provide numerical estimates for the values of the important parameters of the model.

The experiments of Chapter 5 investigate the effect of the background image on the visibility of pseudo random patterns. In Chapter 6 the experimental results and their implications are discussed and a modified version of the model of Chapter 2 which takes the effect of the background image into account is presented. Based on this model a quantitative measure of distortion is defined.

Finally the possible implications of this distortion measure for image encoding are discussed.

CHAPTER 2

VISUAL PERCEPTION

In order to find a distortion measure that can be used successfully in the encoding of still monochromatic images, we need to have a mathematical model for the mechanism of the perception of distortion in images. In this chapter we will first study the formation of the retinal image and the properties of the photoreceptors of the retina to the extent that they relate to the processing of visual patterns by the visual system. Then, after discussing the significance of psychophysical methods, we will make a survey of the past research in psychophysics of vision as related to the visibility of patterns. Based on the results of these studies we will present a basic structural form for the model of the perception of patterns on a constant luminance background. The objectives of most of the experiments that were conducted in this research are based on the assumption of this basic model. These objectives will be discussed at the end of this chapter.

2.1. Processing of Images by the Visual System

The first step in the visual process is the formation of an optical image on the retina of the eye. This step is of interest because it is the retinal image which is the real input to the neural mechanism and it is therefore important to know how it is related to the outside visual scene. Figure 1 is a diagramatic cross section of the human eye. The light entering the cornea is refracted by its surface, then refracted more by the body called the lens, and finally forms an image of the light source at the

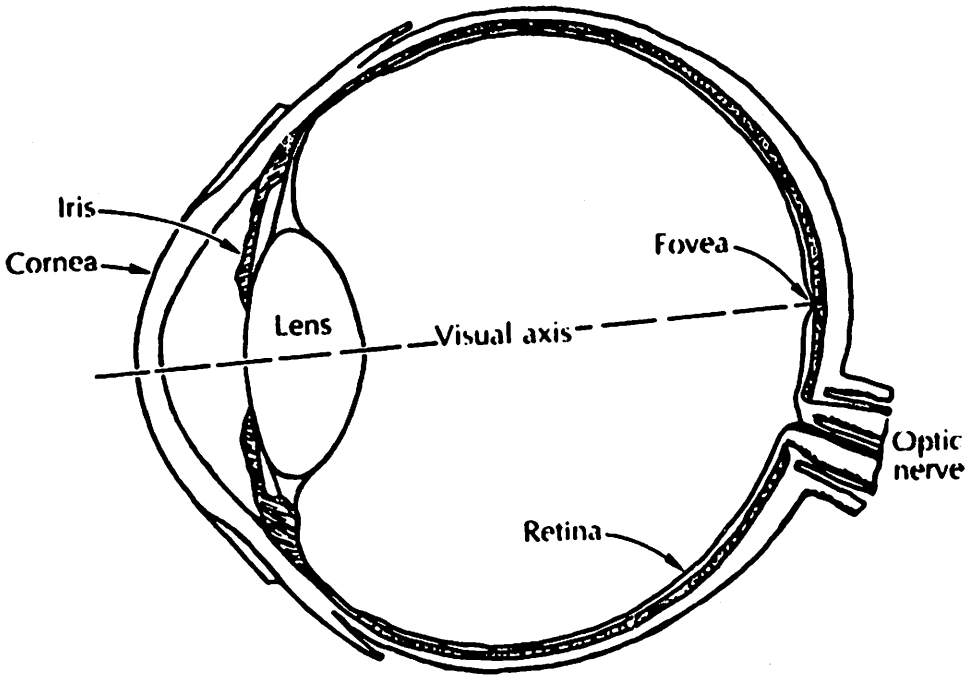


Fig. 1. Diagram of a horizontal cross section of the human eye.

retina. The combination of corneal and lens surfaces all may be considered as an optical imaging device that forms an image of an object on the retina.

Although the retinal image in each eye is usually formed as a projection of three dimensional objects, it is always possible to produce an equivalent light distribution on the retina by an appropriate two dimensional external stimulus. Therefore we shall assume that for each eye our visual stimulus can always be characterized by a function $i(x,y)$ of two spatial variables x and y, which represents the luminance distribution in a plane. We call this the stimulus image. If distances between points in both the object and image planes are expressed as angles subtended at the appropriate nodal

points of the optical system, then both the stimulus image $i(x,y)$ and the retinal image $i_r(x,y)$ can be expressed as the light intensity as a function of the same two variables x and y, which are the orthogonal components of the subtended angle between a point and the origin which is assumed to be on the visual axis.

Partly because of diffraction at the pupil and partly because of focus error, aberrations and imperfections in the optical components of eye, $i(x,y)$ and $i_r(x,y)$ are not in general identical. But like in any imaging device the relation between these two images can be expressed in terms of the point spread function of the optical system, which is the retinal image $h(x,y,u,v)$ in the coordinates x and y, that is produced by a point source of light located at the point (u,v) in the object plane. In its general form the point spread function is a function of four variables. If the optical system has the shift invariance property, namely for all u and v

$$h(x,y,u,v) = h(x-u,y-v,0,0) \qquad (2.1)$$

then the system is called homogeneous or isoplanatic and the point spread function can be identified by a function of two variables, $h(x,y)$, which is the image of a point source of light located on the visual axis. A homogeneous optical system, whose point spread function is circularly symmetric around the origin, is called isotropic. If we identify the stimulus image corresponding to the point source of light by a delta function, $\delta(x,y)$, then we can say that an optical system is perfect if it is homogeneous and its point spread function is a delta function.

Assuming that the stimulus image is incoherently illuminated the illuminance at each point in the retinal image is the sum of the contributions from all the stimulus picture points. Therefore it is possible to calculate the retinal illuminance distribution for any arbitrary luminance distribution in the object plane by the convolution integral

$$i_r(x,y) = \iint h(x,y,u,v)\ i(u,v)\ dudv \qquad (2.2)$$

We can see that the optical system of the eye can be regarded as a two dimensional linear filter. If the assumption of shift invariance is made about this optical system we can also characterize it by its two dimensional frequency response

$$H(f_x,f_y) = \iint \exp[-j2\pi(f_x x + f_y y)]\ h(x,y)\ dxdy \qquad (2.3)$$

which is the two dimensional Fourier transform of the point spread function (impulse response) of the optical system. The variables f_x and f_y are called the spatial frequency variables, and when x and y are subtended angles, they are expressed in cycles/degree (~/0).

From the theory of linear shift invariant filters, we know that if $I(f_x,f_y)$ and $I_r(f_x,f_y)$ are the Fourier transforms of the stimulus and retinal images respectively, then

$$I_r(f_x,f_y) = H(f_x,f_y) \cdot I(f_x,f_y) \qquad (2.4)$$

and the response image can be computed by the inverse Fourier transformation

$$i_r(x,y) = \iint \exp[+j2\pi(f_x x+f_y y)]\ I_r(f_x,f_y)\ df_x df_y \qquad (2.5)$$

In optics the two dimensional frequency response is called the modulation transfer function (MTF). For isotropic imaging devices the MTF becomes real and circularly symmetric, and thus we can identify it by a function of one variable f_r, where

$$f_r = +\sqrt{f_x^2+f_y^2} \qquad (2.6)$$

It is possible to measure the MTF of an optical system by using sinusoidal gratings which are stimulus images generally expressed by

$$i(x,y) = I_o + A \sin[2\pi(\nu_x x+\nu_y y) + \phi]. \qquad (2.7)$$

For example for $\nu_y = 0$, we have a vertical sinusoidal grating with spatial frequency ν_x ~/0. Since i(x,y) is the luminance and hence always non-negative, we have $A \leq I_o$.

The response of a linear shift invariant filter to this stimulus image is a sinusoidal grating with the same spatial frequency components ν_x and ν_y

$$i_r(x,y) = I_o + B \sin[2\pi(\nu_x x+\nu_y y) + \psi] \qquad (2.8)$$

The ratio B/A is equal to the magnitude of the MTF at the point (ν_x,ν_y) in the frequency plane. Therefore if we measure the ratio of the amplitude of the image grating to the amplitude of the stimulus grating for all spatial frequency pairs (ν_x,ν_y), we find the magnitude of the MTF. For isotropic imaging systems, we can

determine the real valued MTF completely by using gratings with one direction only, that is for example let $\nu_y = 0$ and vary ν_x.

Most imaging devices, including the eye, act as low pass filters and above a certain frequency which depends upon the focal ratio of the optical system and the wavelength of the light, the MTF becomes essentially zero. The quality of the optical system of the human eye depends upon the size of the pupil; aberrations being the dominant contributor to the degradation of the image when the pupil is large and diffraction when it is small. At some intermediate pupil size (about 2-3 mm), the optical system produces the least blurred image.

Assuming that the imaging system of the eye is isotropic, its MTF has been measured by, among others, Arnulf and Dupuy [4] and Campbell and Green [5]. The results of Campbell and Green for different pupil sizes are shown in Fig. 2.

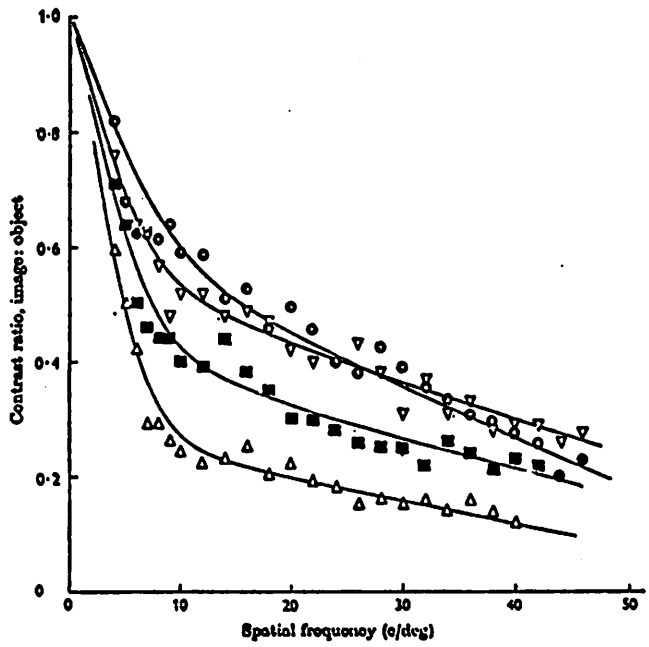


Fig. 2. Contrast transfer functions for the focused eye at four pupil diameters. ● ,2, mm; ▽, 2.8 mm; ■, 3.8 mm; Δ, 5.8 mm (Campbell and Green [5]).

The second step in the visual process is the conversion of the light into signals which can be transmitted and processed by the nervous system. This is done by the photoreceptors of the retina. There are two types of photoreceptors in the retina, rods and cones. In the center of the back of the eyeball, where the visual axis meets the retina, there is a region called the fovea which is slightly depressed. This region contains only cones and the cones there are very densely packed. The density of cones rapidly declines as the distance from the fovea increases. The rods begin to appear about 1° from the center of the fovea, and their density increases to a maximum at about 20°, thereafter falling to a low level. Except for the presence of the optic disk which is the region where the axons from all ganglion cells - which are connected to rods and cones - coverage and exit from the eyeball, the distribution of receptors is radially symmetric about the fovea.

The cones of the fovea are responsible for the high resolution vision at the illumination levels that we normally encounter (photopic vision). The cones at these illumination levels, which range from a modestly lighted room up to a bright sun light, have a lower threshold than rods for responding to the light incident upon them. On the other hand, for a dark adapted eye, the rods have a lower threshold of excitation at very low levels of illumination.

The retinal image is only of functional significance in so far as the retinal photoreceptors respond to the light incident upon them. Since the activity of each photoreceptor depends on the light falling upon it, we can think of the activity pattern of the photoreceptor array as a <u>neural image</u>. This image clearly

differs from the retinal image in that it is spatially quantized. However so long as the retinal image does not contain any spatial frequency component with periods smaller than twice the distance between adjacent photoreceptors, the sampling process does not, per se, result in any loss of information. Indeed in the center of the human fovea, the spacing between adjacent photoreceptors (center to center) is about 2 μm, equivalent to a little less than .5 minutes of are (Polyak [6]). Thus the photoreceptor mosaic in this region is capable of transmitting the spatial frequencies up to 60 ~/0, while the absolute diffraction - determined cutoff frequency for a 2 mm pupil and a light of wavelength 560 nm is 63 ~/0. Therefore the coarseness of the foveal photoreceptor mosaic and the frequecy response of the optical system of the eye are reasonably well matched.

It is well know that the activity of each photoreceptor is a monotonically increasing function of the intensity of the light falling on it. So we might expect that the sensation of brightness produced by any point in the visual field should depend only on the intensity of the light radiating from that point. But in fact this is not the case, and many perceptual phenomena, like the existence of Mach bands (Cornsweet [7]), indicate that the sensation of brightness at each point is determined by the distribution of luminance in a region that includes some of the surrounding area as well as that point.

There is also physiological evidence about the spatial interaction in the visual system of animals. In order to explain this we consider the response of the ganglion cells of the retina.

These are not of course the cells that are immediately affected by signals from the photoreceptors. Rather, they are the output cells of the retina and each one of them is connected to a group of photoreceptors by a rather complex arrangement of intermediate cells. The information in the response of a ganglion cell is conveyed in the form of nerve impulses which the cell generates for transmission along its axon to the brain. Although the occurence of these impulses is random in time, the average frequency of the cell's firing (occurence of an impulse) is a measure of the response of the cell. Kuffler [8] was among the first to study, by electrophysiological techniques, the response of mammalian retinal ganglion cells to patterned visual stimulation. Kuffler found that he could influence the activity of a ganglion cell from points in a relatively extensive area of the retina. This was in fact expected since each ganglion cell is connected indirectly to a number of photoreceptors. There is also evidence that at higher levels of processing, for example in the visual cortex, the response of retinal cell are further combined and actually the process of spatial interaction takes place at several stages of the visual process (Brindly [9]).

The study of the exact nature of the spatial interaction, at all stages of the visual system is a difficult job. So far the amount of information obtained from physiological studies is very small compared to what is actually needed for constructing a mathematical model suitable for the goals of this research. Another approach to the problem of the identification of the visual system is the study of the input-output relationship in this system. The input

in this case is the visual stimulus and the output is the response of the subject which is his psychological reaction to a particular physical characteristic of the stimulus image. The experiments that use this general approach are called psychophysical experiments. The reaction of the subject can take many different forms. For example he can give a grade corresponding to his certainty of seeing a visual pattern after he has been exposed to it for a limited duration of time, or he can give a yes or no answer to a question about his sensation of some characteristic of the stimulus image.

This kind of approach seems to be more relevant to the objective of this research which essentially requires that the mathematical model for perception of error in images to be in agreement with the subjective judgment of the user of the image transmission system.

Since we are going to concentrate on the threshold perception of distortion patterns, the psychophysical experiments we will be dealing with will be mainly of the type in which the subject's threshold of perception of a pattern is measured. This threshold is usually expressed in terms of some measure of the contrast of the pattern when it is barely visible to the subject.

2.2. A Survey of the Past Research

There is a growing body of knowledge obtained from psychophysical experiments about the visibility of visual patterns. The majority of researchers in this area have concentrated on describing the mechanism of pattern detection, at threshold levels of contrast, by models that contain linear two dimensional filters. The incentive for this approach is to model the spatial interaction that take place

in the retinal cells and higher levels of the visual process, by the same methods that are used in optical imaging devices and were explained in sec. 2.1 where the formation of the retinal image was discussed.

Because of this, many psychophysical experiments for investigating the visibility of sinusoidal gratings have been conducted. A number of researchers have measured the sensitivity of the visual system to sinusoidal gratings of different spatial frequencies with the intention of characterizing the detection process by the frequency response of a linear two dimensional filter. In these experiments the contrast of the grating is defined in terms of the maximum and minimum luminances in the stimulus image by

$$\text{Contrast} = \frac{i_{max} - i_{min}}{i_{max} + i_{min}} \tag{2.9}$$

Therefore for a sinusoidal grating defined by Eq. (2.7) we have

$$\text{Contrast} = \frac{A}{I_o} \tag{2.10}$$

where A is the amplitude of the sinusoid and I_o is the mean luminance of the stimulus image.

The <u>contrast sensitivity</u> is defined as the inverse of the contrast that corresponds to the threshold of perception. The variation of contrast sensitivity with the spatial frequency has been studied; for example the result of the experiment of Campbell and Robson [10] is shown in Fig. 3. The function represented by the curve in Fig. 3 is called the contrast sensitivity function or the visual modulation transfer function. Typically the contrast

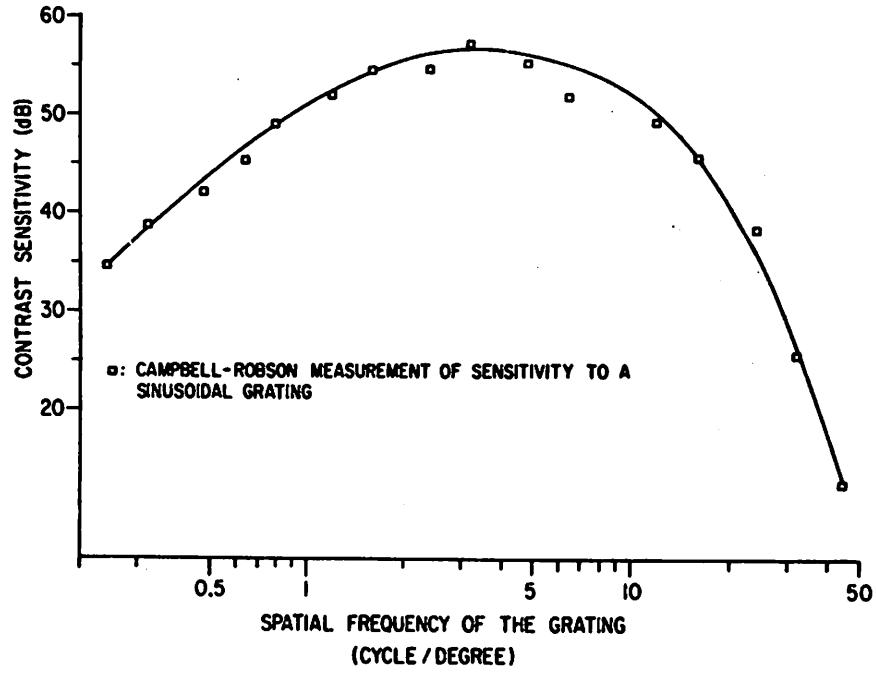


Fig. 3. Contrast sensitivity in dB as a function of the spatial frequency.

sensitivity increases with spatial frequency in the range of frequencies lower than 3 ~/0 and after leveling around 4 ~/0, it decreases exponentially. The rapid decrease of sensitivity at higher frequencies is of course partly because of the low pass filtering that is associated with the formation of the retinal image (Fig. 1). But the additional decrease in contrast sensitivity can be attributed to the process of spatial summation within the neural system. Also since no optical factor can account for the fall-off of the contrast sensitivity at lower frequencies, it is likely to be representative of some kind of inhibitory interaction between the activities of the cells which respond to the retinal image.

The effects of the mean luminance I_o and the wavelength of the light source of the stimulus image on the contrast sensitivity function were studied by Van Nes and Bouman [11]. They found little difference between the contrast sensitivity curves corresponding to different mean luminance levels as long as the mean luminance remains within the range that results in photopic vision. For mean luminances below this range the maximum of the contrast sensitivity curve shifted toward lower frequencies. Furthermore at the photopic luminance levels the same invariance was exhibited with respect to different wavelengths of the light source.

The independence of the contrast sensitivity from the mean luminance means that the amplitude of the grating at the threshold of detection is proportional to the mean luminance. This phenomena is attributed to the results of electrophysiological studies on the retina of vertebrates (Werblin [23]). These results indicate that interactions among various retinal cells adjust the response characteristics of the overall system according to the level of ambient illumination. As indicated previously, each of the ganglion cells of the retina is connected to the photoreceptor cell through a number of intermediate cells (bipolar, horizontal and amacrine cells). Only the photoreceptors contain photopigments and act as transducer, converting the light energy into neural signals. The receptor cell drives the bipolar cell which then passes the signal on to the ganglion cells. The ganglion cells generate the retinal output; their outgoing fibers comprise the optic nerve and carry the neural image to higher centers of the nervous system. The

slope and the offset level of the response curve of the bipolar cell is adjusted within the broad range of the output of the photoreceptor cells. This adjustment is caused by the acitivity of the horizontal cells in the immediate surrounding of the bipolar cell; these cells carry the information about the luminance levels laterally and tune the operating curve of each bipolar cell to the appropriate intensity range. Specifically the above process causes the slope of the linear part of the response curve of the bipolar cell to be inversely proportional to the average surrounding illumination, thus enabling the retina to form a high contrast neural image over a broad range of light conditions. The spatial extent of this interaction among retinal cells is very small compared with the extent of the spatial interactions that take place at higher levels of the nervous system. The overall effect of the above phenomena, when the retinal image consists of a low contrast pattern superimposed on a constant luminance background, is that the neural image is an approximately logarithmic transformation of the retinal luminance levels.

The contrast sensitivity as a function of the orientation of the sinusoidal grating was measured by Campbell, Kulikowsky and Levinson [12]. They concluded that the slope of the decreasing part of the contrast sensitivity curve was maximum for oblique (45° and 135°) gratings. Therefore the maximum resovable spatial

frequency is highest for horizontal and vertical gratings. They also concluded that optical factors such as astigmatism alone can not significantly account for the existence of these preferred directions, and that they are mainly due to some orientational selectivity in the visual nervous system.

The contrast sensitivity curve was initially interpreted in some of the studies (Cornsweet [7]) as the frequency response of a single linear two dimensional filter in the model associated with the perception mechanism. In such a model the stimulus pattern is first processed by this filter and when some measure of the contrast (e.g. rms value or peak value) of the output of the filter reaches a certain threshold level, the subject detects the pattern. But the results of some psychophysical experiments could not be described successfully by this single channel model. Rather they suggested a model in which the stimulus pattern is processed by a bank of processors or channels. The results of some of these experiments will be discussed here.

Campbell and Robson [10] in addition to sinusoids used gratings with square, rectangular and saw-tooth waveforms. They found that over a wide range of spatial frequencies the contrast threshold of the grating is determined only by the amplitude of the fundamental Fourier component of its waveform. Also gratings with complex waveforms cannot be distinguished from sine wave gratings until their contrast has been raised to a level at which the higher harmonic components reach their independent thresholds. Campbell and Robson suggested that these findings can be explained by a model in which several filters, that are selectively sensitive to

limited ranges of spatial frequencies, process the stimulus pattern independently and when the contrast of the output of any one of these channels reaches a threshold level, detection is achieved by the subject.

Blakmore and Campbell [13] arrived at the same conclusion through a set of experiments in which the contrast sensitivity of a grating was measured before and after adaptation to a high contrast (suprathreshold) grating of the same frequency and orientation. They found a five-fold decrease in contrast sensitivity after exposure to the high contrast grating. They also measured the decrease in contrast sensitivity for a range of frequencies when the frequency of the adapting grating was kept constant. This decrease was limited to a band of frequencies, with a half amplitude bandwidth of about one octave, centered about the adapting frequency.

Another result of the experiment of Blakmore and Campbell was that the adaptation of one eye to a high contrast grating reduced the sensitivity of the other eye to the same grating. Through these results they argued that the human visual system may posses neurons selectively sensitive to spatial frequency and that the visual cortex is the site of these neurons.

Saks, Nachmias and Robson [14] conducted a set of experiments using linear combinations of sinusoidal gratings of different frequencies (complex gratings). They considered single channel and multiple channel probabilistic models of detection for the visual system. In their experiment they measured the psychometric function, which is the relative frequency of detection as a

function of contrast, for simple and complex gratings. The resulting psychometric functions were consistent with the multiple channel model but not with the single channel model.

Some experimental results suggest that gratings with different orientations are also detected with separate mechanisms. For example Campbell and Kulikowsky [15] studied the detection of gratings with fixed frequency and varying orientation as they were superimposed on a grating with suprathreshold contrast (masking grating) which had the same frequency and a fixed orientation. They found that the masking effect which is the decrease in contrast sensitivity due to suprathreshold background was maximum when both gratings had the same orientation. The masking effect, as a function of the angle between the threshold grating and the background grating, decreased exponentially by a factor of 2 at 12° on either side of the vertical background grating. This angle was independent of the background contrast. They argued that the high contrast of the background grating reduced only the sensitivity of the mechanism which is tuned to the orientation of background. They also compared the proposed orientationally tuned mechanisms with the orientationally sensitive cells that were electrophysiologically discovered in the visual cortex of the cat by Hubel and Wiesel [16].

Another important aspect of the perception mechanism is the effect that the spatial extent of the stimulus pattern has on its visibility. There are a number of psychophysical results on this question of which we describe that of Robson [17]. He conducted an experiment using sinusoidal gratings with fixed spatial frequency

and varying number of cycles, hence varying area. The contrast sensitivity as a function of the number of cycles displayed was measured for five different spatial frequencies. The family of curves in Fig. 4 show his results. Except for 22 ~/0 and 28 ~/0 the curves are similar in shape and monotonically increasing up to 100 cycles. If we assume that gratings with different spatial frequencies are detected by separate channels, then the similarity and parallelism of the curves in Fig. 4 suggest that the spatial extent of the activity of each channel is different and in fact it is inversely proportional to the frequency to which that channel is sensitive. Also the increasing shape of the curves can give us an idea of the approximate extent of spatial interaction in each channel. For example for the channel which is sensitive to 4.8 ~/0, this extends up to more than 50 periods or about 10 degrees.

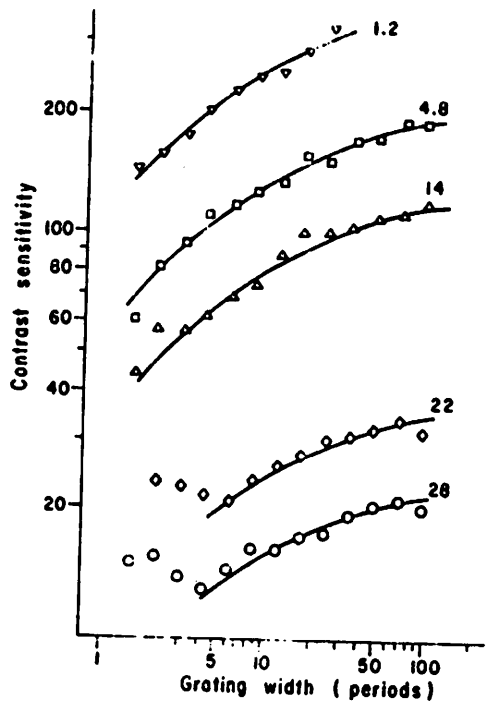


Fig. 4. Contrast sensitivity of sinusoidal gratings as a function of spatial extent for frequencies 1.2, 4.8, 14, 22 and 28 ~/0 (J. G. Robson [17]).

2.3. Assumption of a Model

The experimental results described in the last section suggest a general structural form for the model of perception of the patterns on a constant luminance background. All of the assumptions that we will make here are based on definitive results obtained by other investigators. But some of them will be further supported by the results of the experiments of this research.

The structural form that we assume for the model is shown by the block diagram of Fig. 5. Suppose we consider the detection of the change in the displayed image when it is changed from the background with constant luminance I_o to the image

$$i(x,y) = I_o + \alpha n(x,y) \tag{2.11}$$

$n(x,y)$ is the pattern for which the threshold of detection is to be found and α is the attenuation factor which is varied for different contrast settings. It is assumed that the subject is fixating on the origin of the spatial coordinates which is the same as the center of the display. Under these conditions the input pattern to our model is $g[i(x,y)] - g[I_o]$, where g is a nonlinear function which transforms the luminance of each point and does not involve any spatial interaction. Taking the difference of the patterns after being operated on by the function g is dictated by the results of Van Nes and Bouman and the discussion following it in the previous section. Thus the model of Fig. 5 describes the detection mechanism at levels that are beyond the nonlinearity that exists at the photoreceptor stage of the visual process.

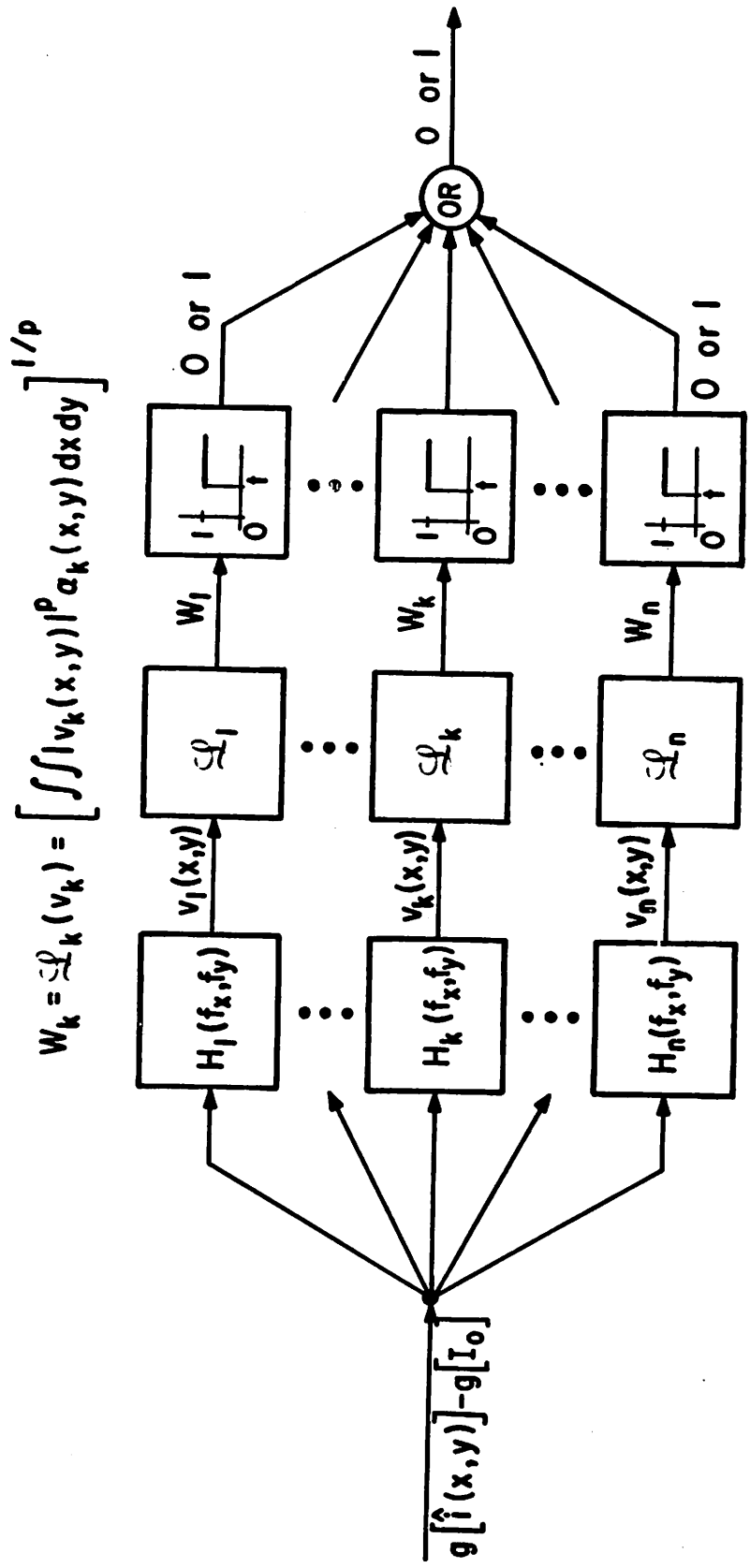


Fig. 5. Block diagram ot the model for detection of patterns on a constant luminance background.

The pattern $g[i(x,y)] - g[I_o]$ is the input to a group of parallel processors that are identical in structural form but each has a different set of parameters. For example in the kth processor the linear two dimensional filter $H_k(f_x,f_y)$ is a band pass filter whose frequency response is peaked at (ν_x^k,ν_y^k) in the frequency plane and is almost zero outside a limited region that corresponds to a particular band of spatial frequencies and orientations. The output of this filter $v_k(x,y)$ is operated on by the functional $\mathcal{L}_k$ which generates the real number W_k. This number is a measure of the response of the kth channel to the pattern $\alpha n(x,y)$.

If W_k is greater than the threshold level t, the output of the threshold device is a 1, which signifies detection by the kth channel. Otherwise the output of the threshold device is 0. The output of each threshold devices is treated as a logical variable and the outputs corresponding to all of the channels are combined by the inclusive OR function to generate a logical 1 or 0 corresponding respectively to detection or no detection by the visual system. This represents the assumption that the presence of some pattern is detected by the subject if one or more of the channels reach their threshold levels.

The linear filter followed by a functional operation in each channel was initially chosen because of the generality of this arrangement for a model using linear operation for describing the spatial interaction in the visual system. But this choice will be further justified when in a coming chapter we compare the experimental results of this research with the results of the simulations that uses this structural form.

Furthermore, at the outset, we can rule out at least one other form of linear operation which seems to be a candidate for this part of the model. In order to explain this we first recall that the experimental results of Robson indicate that the spatial extent of the activity of each channel in terms, of the number of cycles of the grating detected by that channel, is rather large. Given this, we want to see if the response of the channel is the result of a coherent spatial summation of the stimulus grating. If this were indeed the case the obvious choice for the linear operator would have been a linear functional which gives the response W_k directly as the inner product of a fixed pattern $h_k(x,y)$, which should be a tapered sinusoid, with the pattern $g[i(x,y)] - g[I_o]$, namely

$$W_k = \iint h_k(x,y)\{g[i(x,y)] - g[I_o]\}\, dxdy \qquad (2.12)$$

Halter [18] conducted a psychophysical experiment which tests this hypothesis. The stimulus pattern in this experiment consisted of three nonoverlapping segments of sinusoidal gratings with frequency 12.8 ~/0. One patch, of width 2.5 cycles, was fixed in the center of the display, two side patches, of width 2.5 cycles, were displaced from the center patch by gaps of 2.5, 2.25, 2 and 1.75 cycles, so that the phase difference between the center grating and side gratings varied by increments of 90°. The sensitivities for the center patch alone and an uninterrupted grating, whose extent equalled the three patches, were also measured.

The experiment was a rating experiment in which the subject gave grades from 1 to 5, with a 5 indicating complete certainty

of observation of a grating and 1 indicating complete confidence that only a blank screen was shown. The results are shown in Fig. 6. The continuous grating is clearly more visible and the center patch alone clearly less visible than any of the other four patterns, while to within experimental accuracy, all other four patterns, each consisting of three grating segments, are equally visible.

If we assume that all of these patterns are detected by the same channel, namely the channel that is sensitive to frequencies around 12.8 ~/0, then the assumption of coherent summation is rejected because the phase discontinuity in the stimulus did not affect their visibility.

The effect of the area of the stimulus pattern on its detectability is accounted for by incorporating a spatial weighting function in

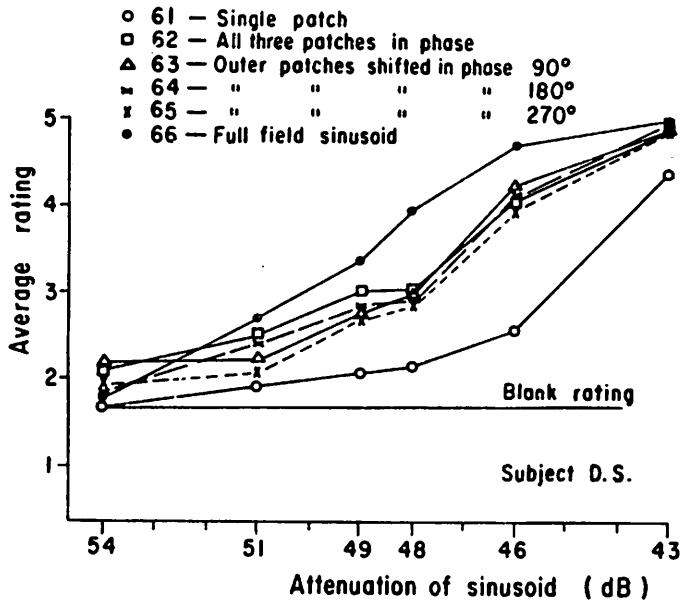


Fig. 6. Visibility of grating segments for various phasings.

the functional operation that follows the filter in each channel. The functional operation should produce some measure of the contrast of the pattern $v_k(x,y)$. The use of L_p-norm functional is useful because for different values of p it includes the usual notions of contrast such as peak value or rms value. Therefore in order to take the extent of the activity of each channel into account, the functional operation will have the form of a spatially weighted L_p-norm, that is

$$W_k = \left[\iint |v_k(x,y)|^p a_k(x,y)\ dxdy\right]^{1/p} \tag{2.13}$$

where the non-negative function $a_k(x,y)$ weights the contribution from different parts of the visual field. From the curves of Fig. 4 we conclude that a(x,y) should be monotonically decreasing from the origin. We also assume that it is circularly symmetric about the origin. The spatial extent over which $a_k(x,y)$ is substantially different from zero varies from channel to channel and, according to the interpretation of the results of Fig. 4, this extent is inversely proportional to the spatial frequency at which the frequency response, $H_k(f_x,f_y)$, is peaked.

We note that this choice of the functional operation causes the response of each channel to be proportional to the value of α in the stimulus pattern given by Eq. 2.11. Therefore we take the threshold level to have the same value, t, for all channels and thus account for different sensitivities of the channels by proper scaling of the response of each filter.

2.4. The Objective of the Experiments

The psychophysical experiments that were conducted in this research can be categorized according to their objectives. Some experiments were designed to provide a test for the hypotheses that we have made so far about the human visual system. Also the results of these experiments, which use sample functions of random patterns as the stimulus pattern, can be compared with the corresponding results obtained by using sinusoidal gratings in other studies.

In order to make this model useful for the eventual application of it, which is in the determination of a measure of distortion, we need to study the model quantitatively. Therefore another set of experiments was designed to estimate the numerical values of some of the parameters which are most important for this application. But the model that we have considered so far has a large number of unknown parameters. This fact, and the assumption that the structural form of different channels is identical, lead us to choose one of these channels and investigate as many parameters corresponding to that channel as possible. Further research combined with the existing results will hopefully provide information as to how each parameter changes as we go to other channels.

The experiments mentioned so far were for detection of the patterns on a constant luminance background. Therefore these experiments do not provide any information about the effect of the image that is to be encoded on the visibility of distortion in that image. In order to study this effect two experiments were

conducted in which the backgrounds themselves were images with contrasts above the threshold of detection. The results of these experiments will provide some information as to how the model that we have defined so far has to be modified in order to take the effect of background image into account.

CHAPTER 3

STIMULUS PATTERNS AND THE EXPERIMENTAL METHODS

In this chapter, the experimental set up, stimulus image characteristics, and the psychophysical methods that were used in this research will be described. Also the use of the sample functions of random patterns as the stimulus images will be discussed.

3.1. Experiment Set Up and Methods

The still monochromatic images that were used in the psychophysical experiments were generated on the face of an ordinary black and white 8" television monitor whose input signal was supplied by a three-channel video disk unit with 105 tracks (total capacity of 105 color, or 315 black and white frames). When a track was selected, three black and white TV picture frames were available at the three channel outputs. The patterns used in the experiments were generated by an IBM 1800 computer system and stored in TV format on the video disk.

The pale blue (phosphor P4) TV screen had a space average luminance of 860 cd/m^2 in all of the experiments, and it was surrounded by a 81 cm x 111 cm cardboard with approximately the same color and luminance as the face of the TV tube. The TV screen was located just behind a rectangular opening in the middle of the cardboard. This opening subtended 3 degrees and 10 minutes in the vertical side and 3 degrees and 36 minutes in the horizontal side at the viewing distance of 218 cm which was used in our experiments.

Most of the experiments were performed on more than one subject.

These subjects either had hematropic vision, or vision correctable to 20/20 with glasses. Binocular vision with natural pupils was used in all of the experiments.

Each stimulus image, as represented by luminance as a function of spatial variables (which can be translated into subtended angles in the x and y directions), was, in each experiment, a special case of the general expression

$$\hat{i}(x,y) = i(x,y)[1 + \alpha_1 n_1(x,y) + \alpha_2 n_2(x,y)] \qquad (3.1)$$

where $i(x,y)$ is a non-negative function representing luminance variation in an image, $n_1(x,y)$ and $n_2(x,y)$ are zero mean functions and α_1 and α_2 are variable attenuation factors.

Normally the image $i(x,y)$ is displayed on the screen; when the subject pushes a micro switch, the display is switched to $\hat{i}(x,y)$ for a specified period of time that was 0.9 seconds in our experiments. The image $i(x,y)$ is called the background image; it was either a constant function of x and y (860 cd/m^2) for experiments dealing with detection of patterns on a blank background, or an image with a fixed contrast above the threshold of perception. The functions $n_1(x,y)$ and $n_2(x,y)$ are called test patterns and, in our experiments, were sample functions of random patterns.

The process by which the patterns $i(x,y)$, $n_1(x,y)$ and $n_2(x,y)$ were generated by the computer is explained in appendix A. Appendix B explains the equipment set up and the generation of the input signal for the TV monitor from the video disk outputs.

The perturbation in a displayed image when it is switched from $i(x,y)$ to $\hat{i}(x,y)$ is caused by the pattern $\alpha_1 n_1(x,y) + \alpha_2 n_2(x,y)$

In each experiment the experimenter's goal was to find a good estimate for the threshold values of α_1 and α_2; that is, the values of α_1 and α_2 that cause the change from $i(x,y)$ to $\hat{i}(x,y)$ to be barely detectable by the subject.

The multiplicative form of the expression for the perturbed image $\hat{i}(x,y)$, given by Eq. (3.1), was dictated by the fact that the input pattern to the model of Fig. 5 of chapter 2, is taken to be $g[\hat{i}(x,y)] - g[i(x,y)]$. Since the luminance of the background patterns were always in the range of photopic vision, we can assume that the function $g(\cdot)$ is logarithmic, that is

$$g(z) = \ell n(z) \tag{3.2}$$

Furthermore for the values of α_1 and α_2 corresponding to the neighborhood of the threshold, the peak to peak value of $\alpha_1 n_1(x,y) + \alpha_2 n_2(x,y)$ is very small compared with 1. Under these conditions the input pattern to model becomes

$$\begin{aligned} g[\hat{i}(x,y)] - g[i(x,y)] &= \ell n[i(x,y) + \ell n[1+\alpha_1 n_1(x,y)+\alpha_2 n_2(x,y)] \\ &\quad - \ell n[i(x,y)] \\ &\approx \alpha_1 n_1(x,y) + \alpha_2 n_2(x,y) \end{aligned} \tag{3.3}$$

This is important, because in the experiments in which the background $i(x,y)$ is a suprathreshold image, we want to study various effects of $i(x,y)$ on the visibility of the perturbation pattern $\alpha_1 n_1(x,y) + \alpha_2 n_2(x,y)$. But we already know one effect of the background image due to the fact that neural image is taken to

be approximately the logarithm of the retinal image, as explained in sec. 2.2. This causes the sensitivity of the eye to the change in luminance i(x,y) at the point (x,y) to be inversely proportional to the value of i(x,y). The multiplicative form of Eq. (3.1) allows us to investigate, without the interference of this phenomenon, other effects that the background image might have on the detectability of the test pattern. In other words the input to the stages beyond the nonlinear process in the retina is generated only by the test patterns and whatever effect the background image has on the visibility of test patterns is due to the phenomena other than this nonlinearity.

In all of the experiments, except one, the function $n_2(x,y)$ was identically zero and only the threshold value of α_1, which for these experiments we denote by α, had to be measured. The self setting method was used in these experiments and the graded response method was employed in the case where both $n_1(x,y)$ and $n_2(x,y)$ were present.

In the self setting method α was controlled by two attenuators that were connected in cascade, each changing the value of α in 1 dB steps. For each trial the experimenter sets one of the attenuators at a randomly chosen value above the approximate threshold of the subject. Then the subject, who is unaware of this setting, adjusts the other attenuator until he feels that by pushing the micro switch he can just detect the presence of the test pattern. The total attenuation of both attenators at the threshold is averaged over a number of such trials; this average is used as the estimate of the threshold value of α.

In the graded response experiment the values of α_1 and α_2 were

set each time by the experimenter and the subject, after pushing the micro switch, gave a 1 response if he detected any change in the stimulus pattern and a 0 response otherwise. The average of responses over a number of trials for each setting of α_1 and α_2 gives the relative frequency of detection. The values of α_1 and α_2 corresponding to a relative frequency of detection of .5 were chosen as the threshold values. In order to monitor the relative frequency of false alarms, the values of α_1 and α_2 were occasionally set to zero and after giving his response the subject was informed that a blank image had been used.

In all of the experiments the objective was to estimate the threshold values of α_1 and/or α_2 as a function of a particular parameter of the test or background pattern. Therefore in each experiment a sequence of stimulus images, corresponding to different values of the parameter of interest, were stored on consecutive tracks of the video disk. In a typical experiment the threshold value of α for each stimulus pattern is determined in a set of trials that uses each track only once. Then this process is repeated while the order in which different stimulus images are presented to the subject is randomly changed each time by the experimenter. In this way the slow variations of the subject's criterion during the entire course of an experiment affects the estimates of α equally for all of the stimulus patterns and consequently its effect on the shape of the variation of the threshold value of α as a function of the parameter of interest is minimized. The total number of trials for estimating the threshold value of α was fixed within each experiment but varied from 12 to 48 for different experiments.

3.2. Pseudo Random Patterns as Stimulus Images

Sample functions of two dimensional random fields (random patterns) were used both as test patterns and as background images throughout the experiments of this research. These patterns are obtained by two dimensional filtering of a sample function of a Gaussian white noise pattern. The implementation of this process on a digital computer is described in appendix A. The filtering process is for controlling the spatial frequency components of the stimulus image. An example of a sample function of a Gaussian band pass random pattern with radially symmetric power spectral density is shown in Fig. 1a of appendix A.

The use of pseudo random patterns instead of conventional gratings has first of all the advantage of being able to test, by using a different type of test pattern, some of the hypotheses made about the visual system from the past psychophysical experiments. However more important are some inherent characteristics of random patterns that make their sample functions more suitable for the experiments designed to estimate the parameters of the visual perception model. One such characteristic is the following.

In all of the experiments except one, the sample functions that were used corresponded to homogeneous Gaussian random patterns. A property of these sample functions, which is useful in our experiments, is the relation that exists between the L_p-norms of such sample functions. This relation can be explained as follows.

Suppose $\{N(x,y),\ (x,y) \in \mathbb{R}^2\}$ is a collection of random variables defining a homogeneous zero mean Gaussian random pattern. The L_p-norm of a sample function $n(x,y)$ of this random pattern over

the frame $[-\frac{T}{2}, \frac{T}{2}] \times [-\frac{T}{2}, \frac{T}{2}]$ is

$$\|n(x,y)\|_{p,T} \triangleq \left[\frac{1}{T^2}\int_{-\frac{T}{2}}^{\frac{T}{2}}\int_{-\frac{T}{2}}^{\frac{T}{2}} |n(x,y)|^p dxdy\right]^{1/p} \tag{3.4}$$

If we assume the ergodicity condition, then $\|n(x,y)\|_{p,T}$ for large values of T becomes approximately equal to the statistical average $[E\{|N(x,y)|^p\}]^{1/p}$. But for all x and y, N(x,y) is a zero mean Gaussian random variable with standard deviation σ, and for such a random variable we have (Papoulis [19])

$$E\{|N(x,y)|^p\} = \begin{cases} 1.3\ldots(p-1)\sigma^p & \text{for } p = 2k \\ \\ \frac{2}{\pi}\, 2^k k!\sigma^{2k+1} & \text{for } p = 2k+1 \end{cases} \tag{3.5}$$

Therefore for large enough T, $\|n(x,y)\|_{p,T}$ differs from $\|n(x,y)\|_{2,T}$ only by a scale factor that depends on p.

This property will prove useful in estimating the different parameters in a channel of the assumed model. For this estimation we need to design a sequence of experiments whose results are orthogonal to each other. This means that we want to be able to use the results of each experiment for estimating a given parameter in such a way that the values of the unknown parameters of the channel do not affect the measurement of the given parameter. It will become clear, in the description of the experiments, that the above property of Gaussian random patterns enables us to estimate the bandwidth of the channel filter without knowing the value of p

in the L_p-norm functional that follows the filter. On the other hand the non-Gaussian random pattern stimuli which lack the above mentioned property will prove useful in the experiment that is designed to estimate the value of p.

Sample functions of Gaussian random patterns were also used as background images in the experiments desinged to study the effect of the background image on the detectability of distortion patterns. This allows us to simulate a Gaussian image source with a given power spectral density, and thus enables us to study the effect of the spatial frequency contents of the image source on the detectability of the distortion.

CHAPTER 4

EXPERIMENT RESULTS (I)

PSEUDO RANDOM PATTERNS ON A CONSTANT LUMINANCE BACKGROUND

In this chapter the experimental results for detection of pseudo random patterns on a constant luminance background are described. Therefore in all of these experiments the background image i(x,y) is the constant $I_o = 860$ cd/m^2. Except for the experiment described in sec. 4.2, in all of the experiments reported here, the function $n_2(x,y)$ was identically zero, and the perturbed image presented to the subject was

$$\hat{i}(x,y) = I_o[1 + \alpha n(x,y)] \tag{4.1}$$

4.1. Visibility of Narrow Band Noise as a Function of Center Frequency

The test patterns used in the experiments described in this section and section 4.2 are sample functions of random patterns with power spectral densities given by the general expression

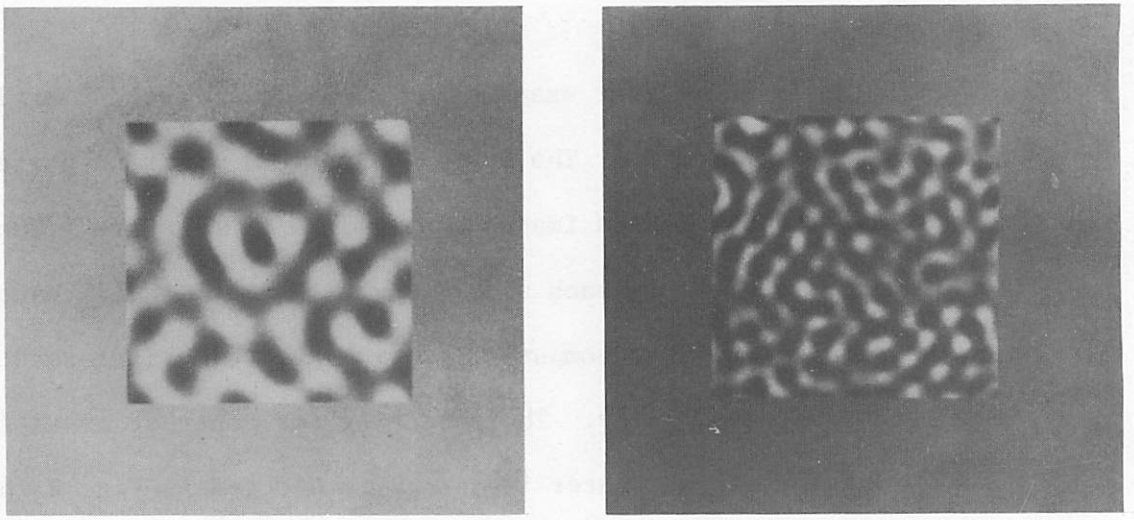
$$S_N(f_x,f_y) = N_o\left\{\exp\left(-\frac{1}{2}\,\frac{\sqrt{f_x^2+f_y^2} - f_o}{\omega}\right)^2\right\} \tag{4.2}$$

where f_x and f_y are in cycles/degree (~/0). We call f_o the center frequency and ω the one-sided bandwidth of these band-pass isotropic random patterns.

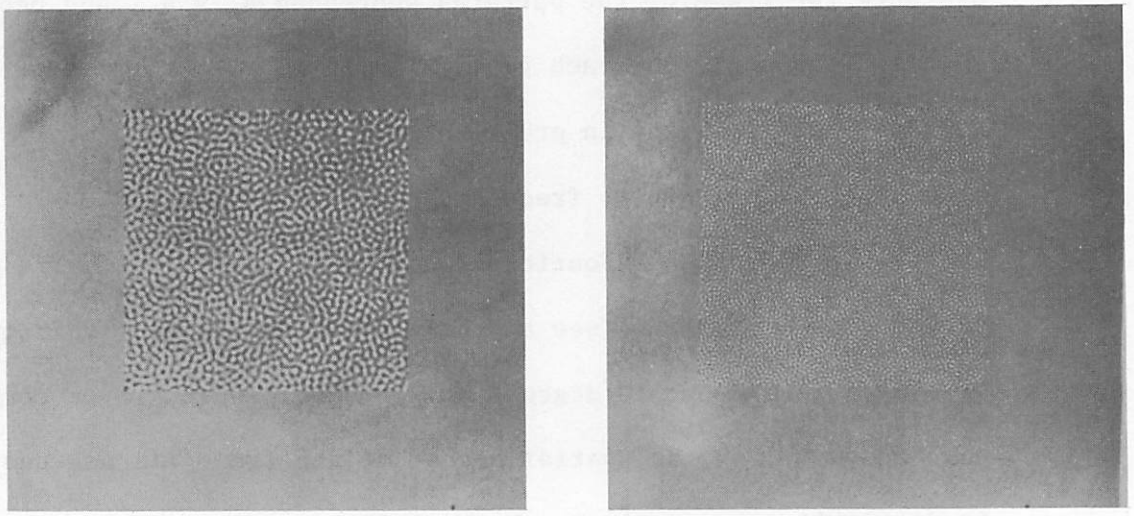
In order to investigate the effect of the center frequency of a narrow-band noise on its detectability, a self setting experiment was conducted in which n(x,y) was a sample function of a narrow band Gaussian random pattern with power spectral density given by Eq. (4.2). The one sided bandwidth ω was 1/5 the center frequency in all patterns. A sequence of stimulus images in which the center

frequency of $n(x,y)$ ranged from 1 ~/0 to 36 ~/0 were presented to the subject and the threshold values of α for these test patterns were measured.

Figure 1 shows four examples of the suprathreshold versions of these test patterns. The spatial extent of the test patterns in all of these stimulus images was limited by a square frame that subtended 2 degrees in each side. For each image the threshold value of α was used to compute $1/\|\alpha n(x,y)\|_2$, which is termed the rms contrast sensitivity. The resulting rms contrast sensitivities as a function of the center frequency is plotted in Fig. 2 for two subjects. The dashed curves correspond to the data taken when the viewing distance was changed from 218 cm to 109 cm. At this viewing distance the frame of the patterns subtended 4° x 4°, and hence the spatial frequencies of each pattern in terms of cycles/degree were halved. This was done in order to make the measurement for very low values of the center frequency feasible. The discrete finite two dimensional Fourier transform is used to generate these pseudo random patterns (see appendix A); when the center frequency of a pattern (in cycles/degree) is not sufficiently large compared with $\frac{1}{T}$, where T is the spatial extent of the frame (in degrees), it is difficult to approximate the radially symmetric power spectral density of Eq. (4.2) in the discretized transform domain. The contrast sensitivity curve obtained by Campbell and Robson using sinusoidal gratings (Fig. 3 of chapter 2) is plotted on the same coordinate axes with a downward shift of 15 dB. We can see that the rms contrast sensitivity curve for narrow band pseudo random patterns is very similar in shape to that of sinusoidal gratings.



f_o = 2.5 ~/0 f_o = 4.5 ~/0



f_o = 18 ~/0 f_o = 36 ~/0

Fig. 1. Narrow band isotropic pseudo random patterns with different center frequencies; one sided bandwidth is 1/5 the center frequency f_o.

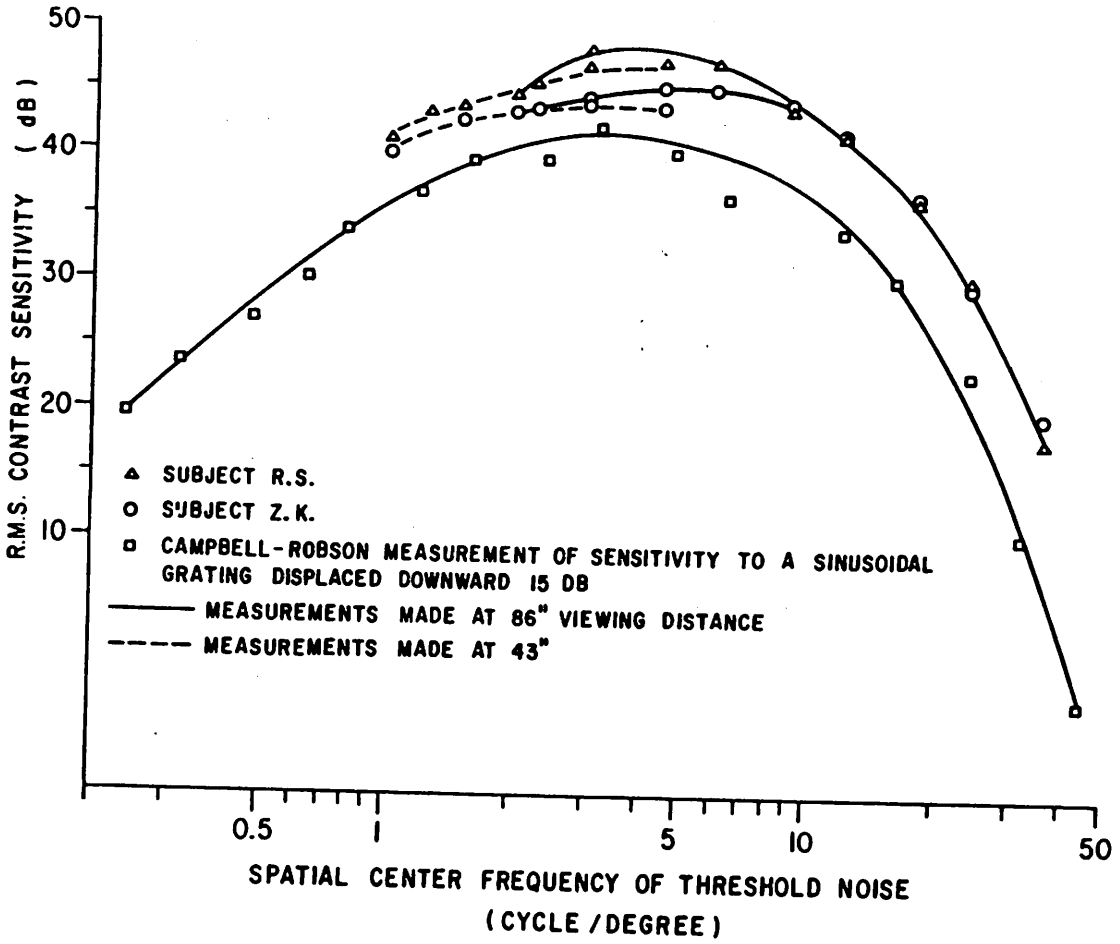


Fig. 2. rms contrast sensitivity curves for narrow band noise and sinusoidal gratings.

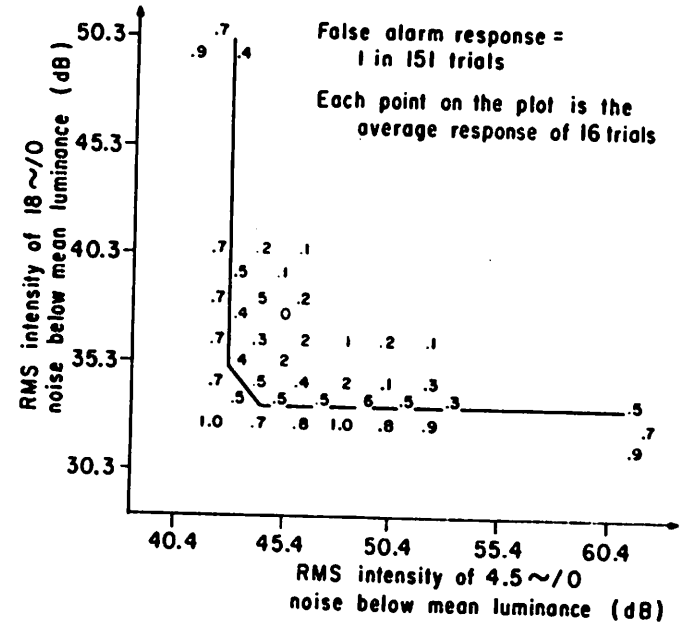
4.2. Detection of Two Narrow Band Pseudo Random Patterns

In order to give more support to the parallel and independent channels hypothesis, shown by the model of Fig. 5 in chapter 2, we report in this section the results of an experiment that was designed by Sakrison [20]. In this experiment $n_1(x,y)$ and $n_2(x,y)$ were sample functions of two independent narrow band random patterns with power spectral densities given by Eq. (4.2). Each of the patterns $n_1(x,y)$ and $n_2(x,y)$ was generated by filtering a different white noise sample function. This cuased the two patterns to be statistically independent.

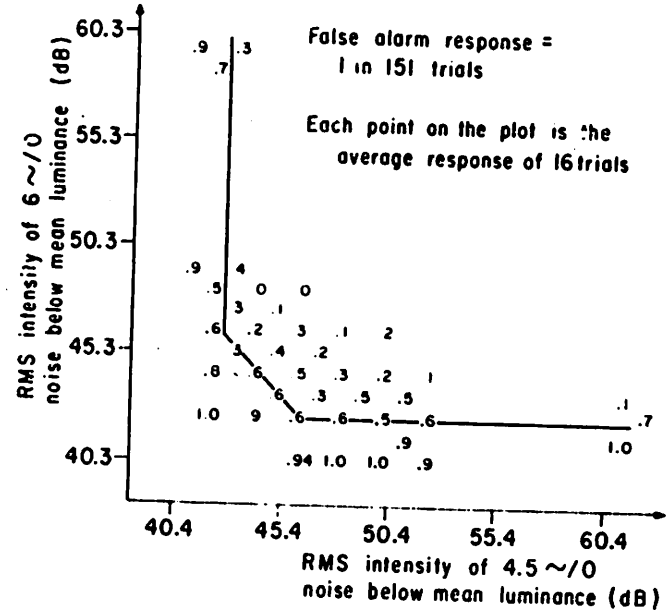
Two different stimulus images were presented to the subject. The center frequencies of $n_1(x,y)$ and $n_2(x,y)$ were 4.5 ~/0 and 18 ~/0 respectively in one image and 4.5 ~/0 and 6 ~/0 in the other. The one-sided bandwidth of each random pattern was 1/5 its center frequency. The spatial extent of the smaple functions was limited to a square frame that subtended 2 degrees in each side.

Using the graded response method, the relative frequency of detection for different combinations of the values of α_1 and α_2 was measured for each of the two cases. From the attenuations α_1 and α_2, $\|\alpha_1 n_1(x,y)\|_2$ and $\|\alpha_2 n_2(x,y)\|_2$, which are the rms values of $\alpha_1 n_1(x,y)$ and $\alpha_2 n_2(x,y)$, were computed and the values $\frac{1}{\alpha_i \|n_i(x,y)\|_2}$ (rms contrast) in dB were used for plotting the results shown in Fig. 3. Figure 3a is for the image with 4.5 ~/0 and 18 ~/0 center frequencies and Fig. 3b is for 4.5 ~/0 and 6 ~/0 center frequencies. In each plot the relative frequencies of detection are shown at the points with corresponding rms contrast coordinates.

Lines are drawn through the points with .5 relative frequency



(a)



(b)

Fig. 3. Relative frequency of detection for combination of two narrow band random patterns. (a) Center frequencies 4.5 ~/0 and 18 ~/0. (b) Center frequencies 4.5 ~/0 and 6 ~/0.

of detection which are assumed to correspond to the threshold of detection. When the center frequencies of $n_1(x,y)$ and $n_2(x,y)$ are widely spaced (Fig. 3a) these lines are vertical and horizontal over a broad range of contrsts, indicating that the detection of one of the test patterns occurs independently of the amplitude of the other pattern. In other words, detection occurs when either one of the patterns reaches its threshold. But when the center frequencies of $n_1(x,y)$ and $n_2(x,y)$ are close together (Fig. 3b), for the contrasts where both patterns are on the verge of being detected, the visibility of the combined pattern is increased beyond the visibility of either pattern alone, and it seems that the combination of both patterns is being detected by a single mechanism tuned to a frequency intermediate to the frequencies of the two components.

4.3. Visibility of Isotropic Noise as a Function of its Bandwidth

In the model discussed in sec. 2.3, one of the most important parameters is the bandwidth of the filter in each channel. According to the results of sec. 4.2 and those referred to in chapter 2, it is reasonable to presume that each channel is selectively sensitive to a band of spatial frequencies. In this and the next two sections we describe a sequence of experiments that were conducted in order to estimate the selectivity of one of the channels.

In choosing a suitable set of stimulus patterns for the experiments that estimate such parameters as the bandwidth of a channel one must make sure that the detection of all of these patterns is always achieved by the single channel that we have in mind. This suggests that the channel chosen for investigation has to be the one with a frequency response peaked at the spatial frequency to which

the visual system is most sensitive. The reason for this can be explained by noting that although the pass bands of different channels may substantially overlap in the frequency domain, it is very likely that a narrow band pseudo random pattern with center frequency equal to that of the most sensitive channel, will always be detected by this channel.

The rms contrast sensitivity curve of Fig. 2 shows that the channels with center frequency 4.5 ~/0 are among the most sensitive ones. The six test patterns $n_{\omega_i}(x,y)$, $i = 1,2,\ldots,6$ that were used in this experiment were sample functions of isotropic random patterns with power spectral density given by Eq. 4.2, with center frequency $f_o = 4.5$ ~/0 and one sided bandwidth ω ranging from $\omega_1 = .3$ ~/0 to $\omega_6 = 2.25$ ~/0. Figure 4 shows examples of $n_\omega(x,y)$ for four different values of ω. As can be seen the sample functions were multiplied by a weighting function that caused the test patterns to be slightly tapered on the edges of a 2° x 2° square frame. This was done for all of the remaining experiments of this chapter in order to eliminate sharp edges that result in unwanted frequency components which may in turn cause the detection to take place through a channel different from the one being investigated. The above patterns were scaled so that the value of $\max_{x,y}|n_{\omega_i}(x,y)|$ was identical for all six test patterns.

The experiment consisted of measuring the threshold value of α in Eq. (4.1) as a function of the bandwidth ω. The self setting method was used for these measurements. The threshold value of α in dB attenuation (i.e. $-20 \log \alpha$) as a function of ω is plotted for two subjects with circles at the top of Fig. 5. Since the value

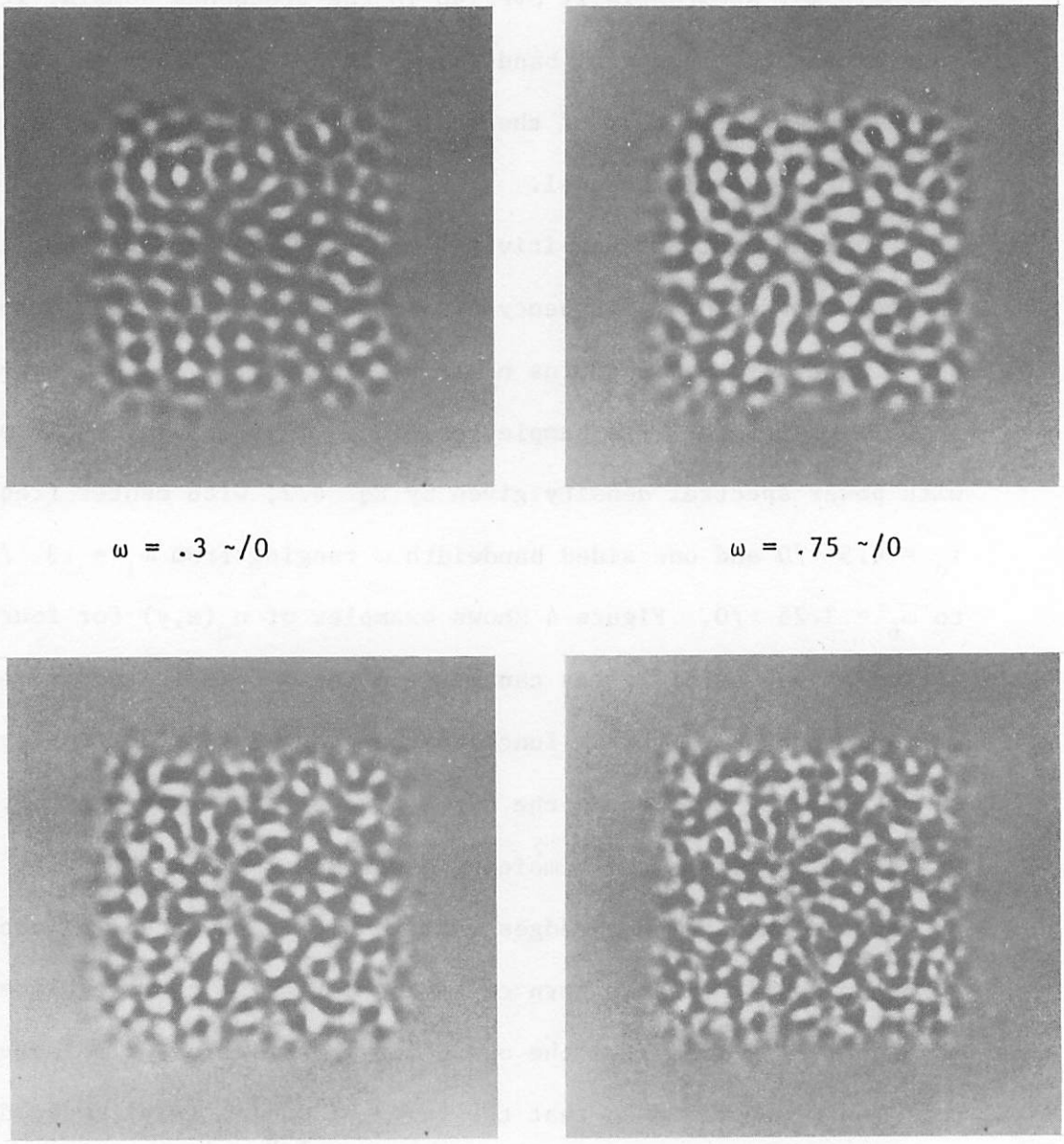


Fig. 4. Narrow band isotropic pseudo random patterns with different one-sided bandwidths, ω.

of $\max_{x,y} |n_{\omega_i}(x,y)|$ was the same for all of the stimulus patterns, the shape of the variation of α as a function of ω does not have any particular meaning in itself.

The estimation of the selectivity of the channel filter consisted of first assuming a certain shape for the frequency response of the band pass filter, with the bandwidth of the filter being a parameter that can be changed. For each bandwidth of the filter and each value of p in the L_p-norm functional that follows the filter (see Fig. 5 of chapter 2), the detection process in the channel can be simulated by computing the response of the filter $v(x,y)$ to the same patterns $n_{\omega_i}(x,y)$ that were used in the experiment, and performing the L_p-norm operation to obtain the response of the channel W_k. This process was implemented on a digital computer using two dimensional digital filtering.

We note that in the model the response of the channel is proportional to the value of α for the input $\alpha n(x,y)$. Therefore hypothesizing that $\|v\|_p$ should be a constant value at the threshold of detection implies that the threshold value of α should be inversely proportional to $\|v\|_p$, or that $20 \log_{10} \|v\|_p$, the model response in dB, should differ from the threshold value of α in <u>dB attenuation</u> only by an additive constant.

The general procedure followed in the remaining experiments described in this chapter is to compute the response of the channel, $\|v\|_p$ for all of the stimulus patterns, using different combinations of the bandwidth parameters and the parameter p of the model. For each set of model parameters we plot the response in dB as a function of the parameter of the stimulus pattern (i.e. the bandwidth

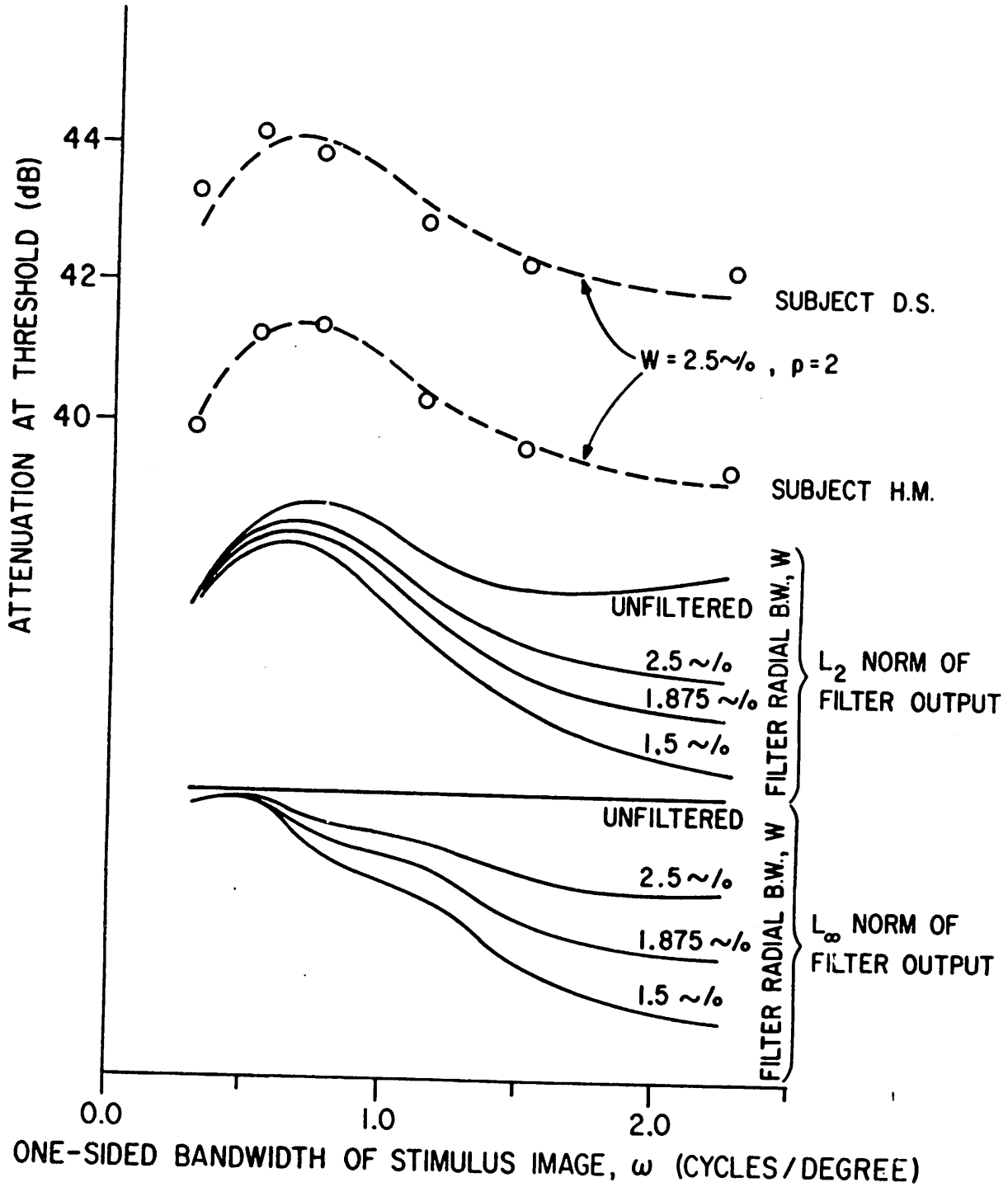


Fig. 5. Response of visual system as a function of the bandwidth of isotropic narrow band pseudo random pattern; experimental results and simulation results.

for this experiment); the curve corresponding to the right values of channel parameters should match the curve of α (in dB attenuation), as a function of the stimulus parameter only by a vertical shift.

For this experiment the filter transfer function was assumed to have the form

$$H(f_x,f_y) = A \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{f_x^2+f_y^2} - 4.5}{w}\right)^2\right\} \qquad (4.3)$$

which corresponds to an isotropic filter with band pass characteristics (see Fig. A2, appendix A); w is termed the (one-sided) radial bandwidth of the filter.

Assuming this filter shape, the response of the channel for different values of w and the parameter p were computed. These results are shown by two groups of curves at the bottom of Fig. 5. One group corresponds to p = 2 and the other to p = ∞, where L_∞ norm of the filter output is $\|v(x,y)\|_\infty = \max_{x,y}|v(x,y)|$. These curves are drawn with the same vertical dB scale as the experimental data shown at the top of the figure. The vertical position of each group is arbitrarily chosen.

Each curve within a group corresponds to a different value of w. For the value of p from 2 to 16 the shape of the curves remained almost unchanged, change was noticed for p = 32 for which value the curves bore resemblance to those for p = ∞. Therefore the p = 2 group of curves can actually represent the results for small values of p(2 to 16). We can see that the curve corresponding to p = 2 (or 2 to 16) and w = 2.5 ~/0 fits the experimental data points for

both subjects. A dashed replica of this curve is drawn through the experimental points for each subject. We can also see that the shape of the curves show good sensitivity to the value of bandwidth w around the values 2 to 4. This allows a relatively accurate estimation of the bandwidth for the assumed isotropic filter.

It is noted that the experimental data can not be fitted with any of the curves in the $p = \infty$ group. Therefore assuming that the filter has the isotropic form expressed by Eq. (4.3) our estimate of its one-sided bandwidth becomes 2.5 ~/0. As for the value of p although 32 and higher values are ruled out, it remains undetermined.

The insensitivity of the shape of the curves to the value of p, for small values of p, is to be expected. This is due to the fact that n(x,y) and v(x,y), which is the result of linear homogeneous filtering of n(x,y), are sample functions of Gaussian random patterns. According to the discussion of sec. 3.2, the L_p norm of each sample function differs from its L_2 norm only by a scale factor which depends on the value of p. Therefore on a logarithmic scale the curves corresponding to the same value of bandwidth parameter but different values of p differ only by a vertical shift. The changes that do occur for large p are partly due to the finiteness of the frame size, and partly because of the fact that the tails of the distributions of the random patterns involved depart from Gaussian.

4.4. Orientational Selectivity of the Channel Filter

The experimental results of Campbell and Kulikowsky, described

in sec. 2.2 indicate the existence of channels in the visual system that are selectively sensitive to different orientations. In the model presented in sec. 2.3 this was accounted for by assuming that each channel is sensitive to the frequencies within a limited region of the frequency plane, corresponding to a range of spatial frequencies and orientations.

An experiment similar to that described in sec. 4.2 was designed by Sakrison [20] to further test this hypothesis. The pseudo random patterns used in this experiment were band limited in both f_x and f_y directions in the frequency domain. In one stimulus image the pass band of $n_1(x,y)$ (see sec. 4.2) was on the f_x axis and the pass band of $n_2(x,y)$ was on the f_y axis, both centered around 4.5 ~/0. In the other stimulus image, the pass band of $n_1(x,y)$ was on the f_x axis and that of $n_2(x,y)$ was located on a line making a 45° angle with the f_x axis. As in the experiment of sec. 4.2, the relative frequency of detection for different combinations of the contrasts of $n_1(x,y)$ and $n_2(x,y)$ was measured for both pair of images. The results were similar to those of Fig. 3a in both cases, indicating that detection of one pattern occurs independently of the contrast of the other pattern when the two patterns have different orientations.

Given that channels are orientationally selective, the objective of the experiment that we are going to describe in this section is to estimate the orientational selectivity of one of these channels, i.e. the angular bandwidth. In order to do this we assume that the filter in each channel has a frequency response that in terms of the polar coordinates in the frequency plane can be characterized by the

separable form

$$H(f,\theta) = H_r(f) \cdot H_a(\theta) \tag{4.4}$$

$$f = \sqrt{f_x^2+f_y^2} \quad ,\theta = \tan^{-1}\frac{f_y}{f_x}$$

where the radial part is given by

$$H_r(f) = A \exp\{-\frac{1}{2}\left(\frac{f-f_o}{w}\right)^2\} \tag{4.5}$$

and the angular part is given by

$$H_z(\theta) = B\{\exp[-\frac{1}{2}\left(\frac{\theta-\theta_o}{b}\right)^2] + \exp[-\frac{1}{2}\left(\frac{\theta-\theta_o-\pi}{b}\right)^2]\} \tag{4.6}$$

This means that the frequency response is band pass in both the spatial frequency variable f and the orientation variable θ. The frequency f_o, termed the center frequency, and the orientation θ_o, specify the locations at which the frequency response of the filter is peaked. The parameters w and b are termed the radial bandwidth and angular bandwidth respectively and specify the width of the pass band of the channel.

According to the results of Campbell, Kulikowsky and Levinson (see sec. 2.2) the channels that correspond to $\theta_o = 0^o$ (vertical gratings) are among the most sensitive ones. Therefore the experiment of this section was designed for the $f_o = 4.5$ ~/0 and $\theta_o = 0^o$ channel. The objective was to estimate the angular bandwidth b.

The five test patterns, $n_{\beta_i}(x,y)$, i = 1,...,5 that were used in this experiment, were sample functions of random patterns with power spectral densities given in terms of polar frequency coordinates by

$$S_{N_\beta}(f,\theta) = N_o \exp[\frac{1}{2}(\frac{f-4.5}{2.25})^2]\{\exp[-\frac{1}{2}(\frac{\theta}{\beta})^2] + \exp[-\frac{1}{2}(\frac{\theta-\pi}{\beta})^2]\} \quad (4.7)$$

in which the value of β, the angular bandwidth of the random pattern, ranges from 3.5° to 28°. Figure 6 shows examples of $n_\beta(x,y)$ for four different values of β. The value of $\max_{x,y} |n_\beta(x,y)|$ was the same for all five test patterns.

The experiment consisted of measuring the threshold value of α in Eq. (4.1) as a function of β, the angular bandwidth of the stimulus patterns. The threshold values of α in dB attenuation for two subjects are plotted by circles and squares in Fig. 7.

In order to estimate the angular bandwidth, the frequency response given by Eqs. (4.4), (4.5) and (4.6) was assumed for the filter. The parameter b is to be estimated; the other two parameters are the radial bandwidth w and p in the L_p norm functional. As in the previous section, for each set of values of b, w, and p the detection process in the channel was simulated by computing the response of the filter to the same patterns $n_{\beta_i}(x,y)$ $i = 1,\ldots,5$ that were used in the experiment, and performing the L_p norm operation to obtain the response of the channel.

Two groups of curves obtained from such a computation are shown in Fig. 7. One group corresponds to p = 2, w = 2.5 ~/0 and the other to p = ∞ and w = 2.5 ~/0. These curves are drawn with the same vertical scale as the experimental data in Fig. 7; the vertical position of each curve is arbitrary and is adjusted so that the response corresponding to β = 3.5° coincide for all filter parameters. Each curve within a group corresponds to a different

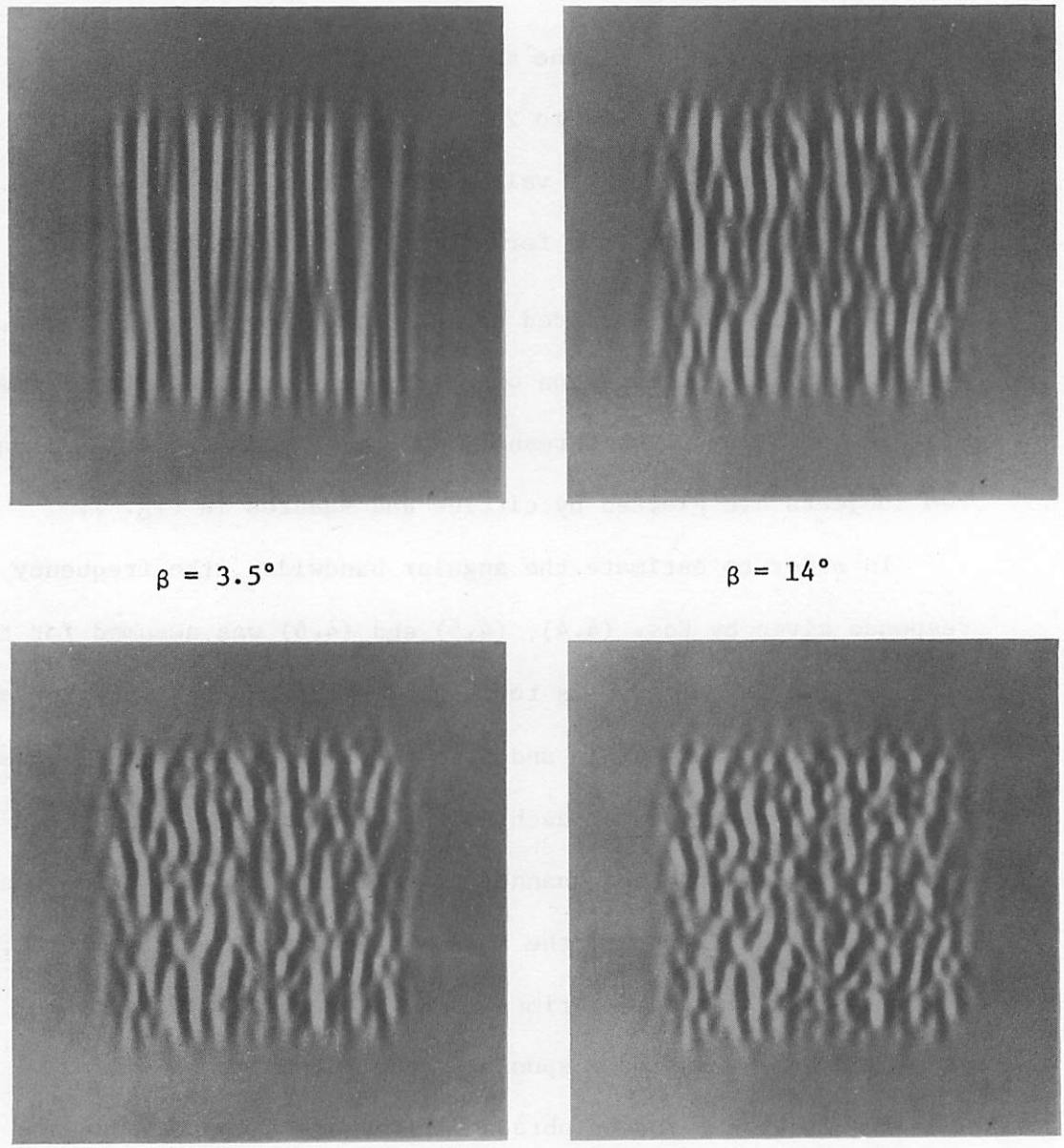


Fig. 6. Four examples of narrow band pseudom random patterns with different angular bandwidths, β.

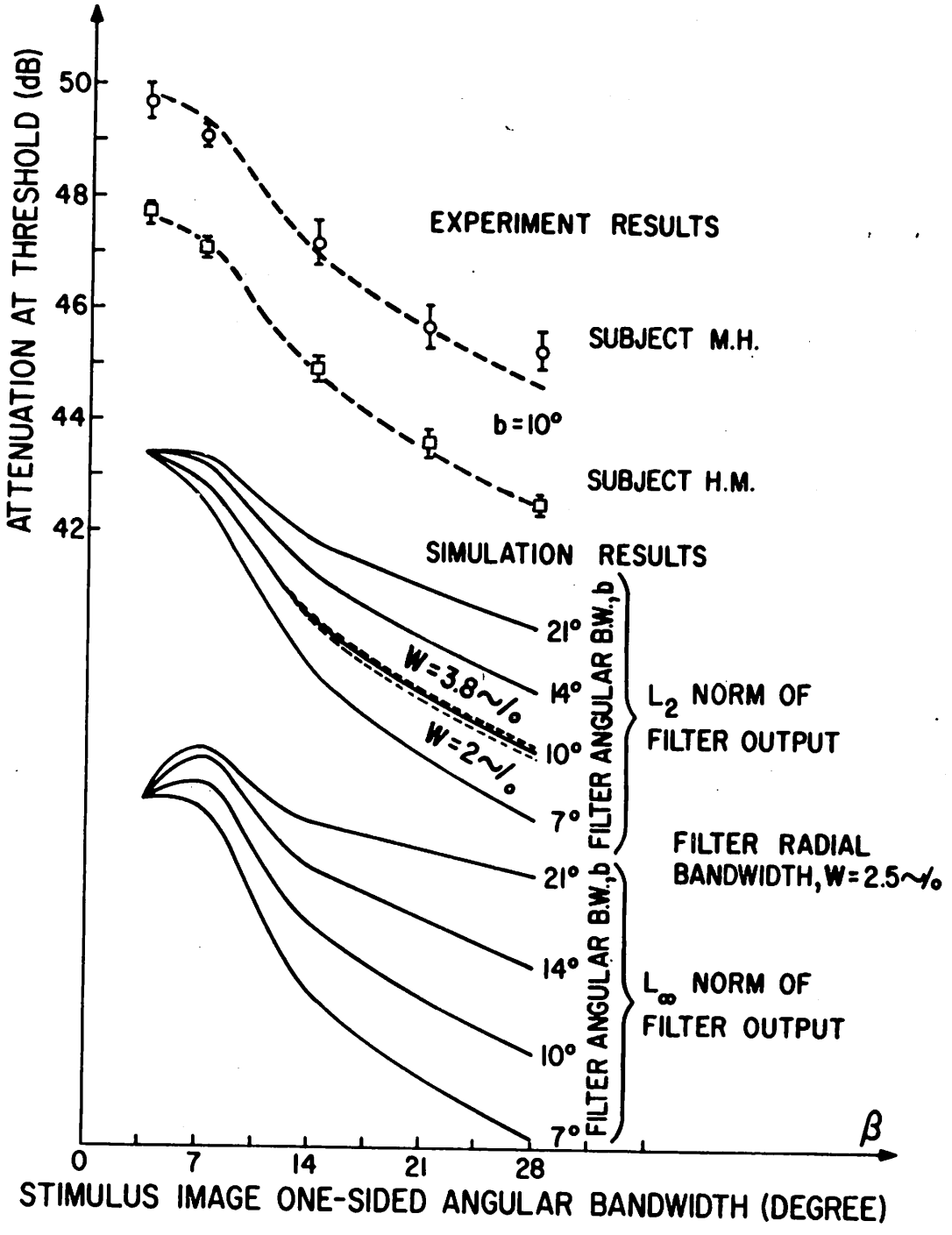


Fig. 7. The response of the visual system versus the angular bandwidth of the stimulus; experimental data; and simulation results.

value of b. Again the shape of the curves is insensitive to the value of p for small p (2 to 10) for large p (16 and higher) the curves show a maximum around $\beta = 7°$.

In computing these simulation results the radial bandwidth of the filter, w was assumed to have the value 2.5 ~/0. This choice was based on the findings of sec 4.3. Since the assumption of orientational selectivity may change this figure, we should also examine the sensitivity of the shape of the curves to the value of w. This was done by computing the curves corresponding to $p = 2$, $b = 10°$, $w = 2$ ~/0 and $p = 2$, $b = 10°$, $w = 3.8$ ~/0. These are shown by dashed curves in the $p = 2$ group. As it can be seen, the shape of this curve is almost insensitive to changes in the parameter w.

We note that the curve corresponding to $p = 2$ (or $p = 2$ to 10) and $w = 2.5$ ~/0 (or $w = 2$ ~/0 to 3.8 ~/0) and $b = 10°$ fits the experimental data points for both subjects. A dashed replica of this curve is drawn through the data points of each subject in Fig. 7. We can also see that the experimental data can not be fitted with any of the curves in the $p = \infty$ group. Thus our estimate of the one-sided angular bandwidth of the channel becomes 10°.

Because of the insensitivity of the simulation results to the value of w, this estimate of the bandwidth is independent of the value of w. Therefore in the remaining experiments of this chapter we can fix the value of this parameter at 10°.

4.5. Estimation of the Radial Bandwidth

In sec. 4.3 we described an experiment in which the radial bandwidth of the channel filter was estimated. However, that result

was based on the assumption that the channel filter has a radially symmetric frequency response, thus ignoring the orientational selectivity of the channel. In this section we describe the experiments that were conducted to estimate the radial bandwidth, w, of the filter with frequency response given by Eqs. (4.4), (4.5) and (4.6).

First an attempt was made to use the experimental results of sec. 4.3 by conducting the simulation with the orientationally tuned filter. Therefore the isotropic stimulus patterns used in that experiment were processed by the filter given by Eqs. (4.4), (4.5) and (4.6), with f_o = 4.5 ~/0, b = 10° and various values of w. Since these test patterns are isotropic there is an uncertainty as to which one of the channels corresponding to f_o = 4.5 ~/0 and different values of θ_o reaches its threshold.

Initially θ_o was taken to be 0° and the response of the channel was computed for different values of the radial bandwidth w. Then the same process was repeated for θ_o = 84° which corresponds to the direction in which the particular sample functions used in the experiment had the highest energy concentration. This direction was found by looking at the magnitude of the Fourier transform of these smaple functions. However the shape of the curves, which represent the relative variation of the channel response, exhibited little sensitivity to changes in the value of θ_o.

In order to obtain a good estimate of the radial bandwidth from these curves, the shape of the curves has to be sensitive to the choice of this parameter. However within the range of values of w that resulted in curves that fitted the experimental results,

this sensitivity was small and taking the experimental error into account the value of w could be anywhere between 2 ~/0 and 4.5 ~/0. This is due to the fact that the maximum bandwidth in the six stimulus patterns used in the experiment, $\omega_6 = 2.25$ ~/0 was not large enough to cause a substantial variation in the response of the filter, as w varied within the above range.

In order to obtain a more accurate estimate of the radial bandwidth, another experiment was conducted in which the six stimulus images were sample functions with power spectral densities given by

$$S_{N_i}(f,\theta) = N_o \exp[-\frac{1}{2}(\frac{f-4.5}{\omega_i})^2]\{\exp[-\frac{1}{2}(\frac{\theta}{10})^2] + \exp[-\frac{1}{2}(\frac{\theta-180}{10})^2]\} \qquad (4.8)$$

$$i = 1,\ldots,6$$

which are band-pass around $f_o = 4.5$ ~/0 and $\theta_o = 0°$, 180° with the angular bandwidth fixed at 10°. The radial bandwidth ranged from $\omega_1 = .75$ ~/0 to $\omega_6 = 4.25$ ~/0. The value of $\max_{x,y}|n_{\omega_i}(x,y)|$ was the same for all six patterns. Four examples of these patterns are shown in Fig. 8.

The threshold value of α in dB attenuation as a function of the stimulus radial bandwidth ω was measured on two subjects. The results are plotted in Fig. 9.

The simulation results were obtained by processing the test patterns $n_{\omega_i}(x,y)$, $i = 1,\ldots,6$ with the filter given by Eqs. (4.4), (4.5) and (4.6) ($f_o = 4.5$ ~/0 and $\theta_o = 0°$) using various values of w while b was fixed at 10°. The response of the channel is shown by the curves of Fig. 9. Since these results were insensitive to the value of p, only the curves for p = 4 are shown. By comparing

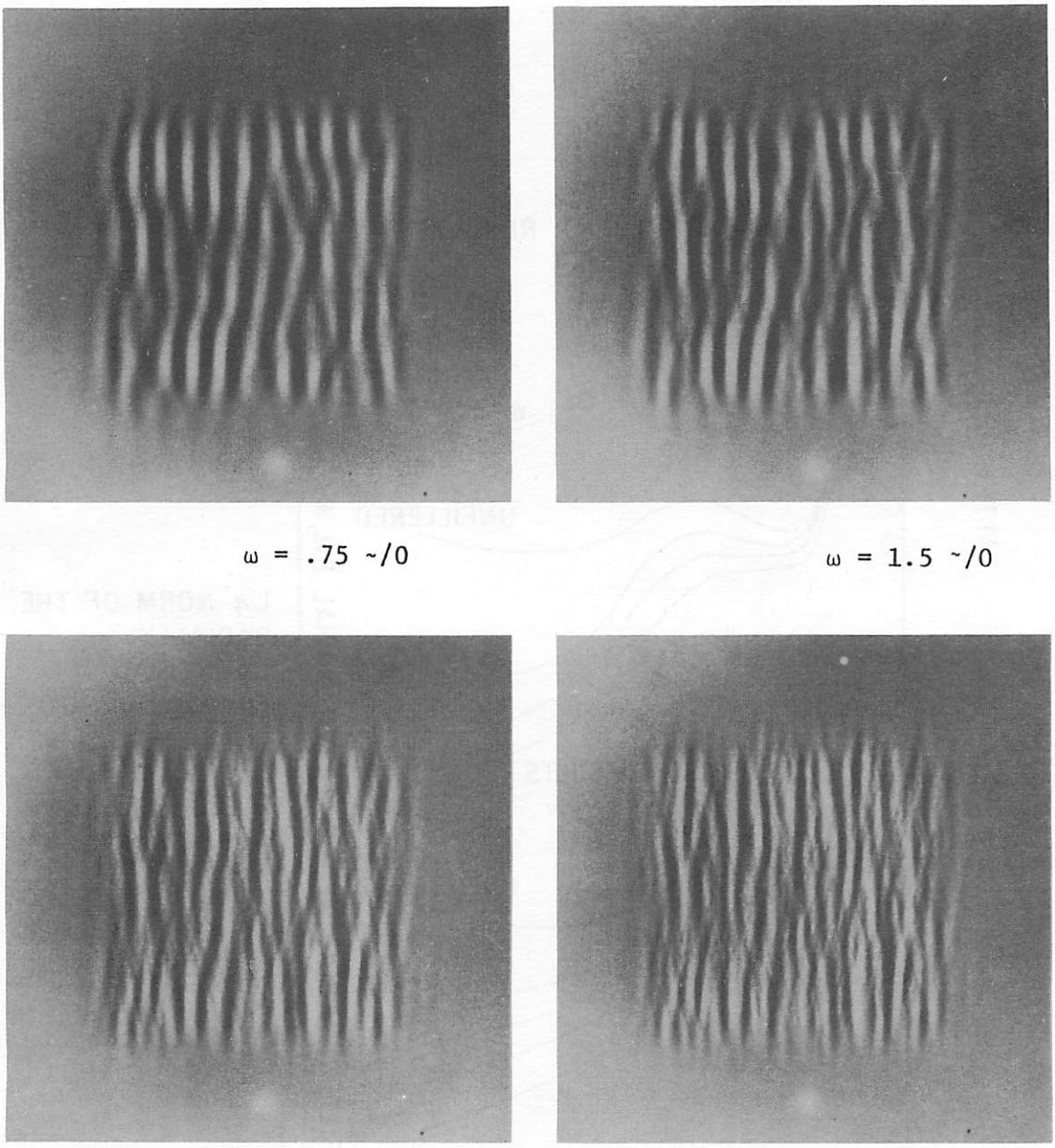


Fig. 8. Narrow band random patterns with four different radial bandwidths ω.

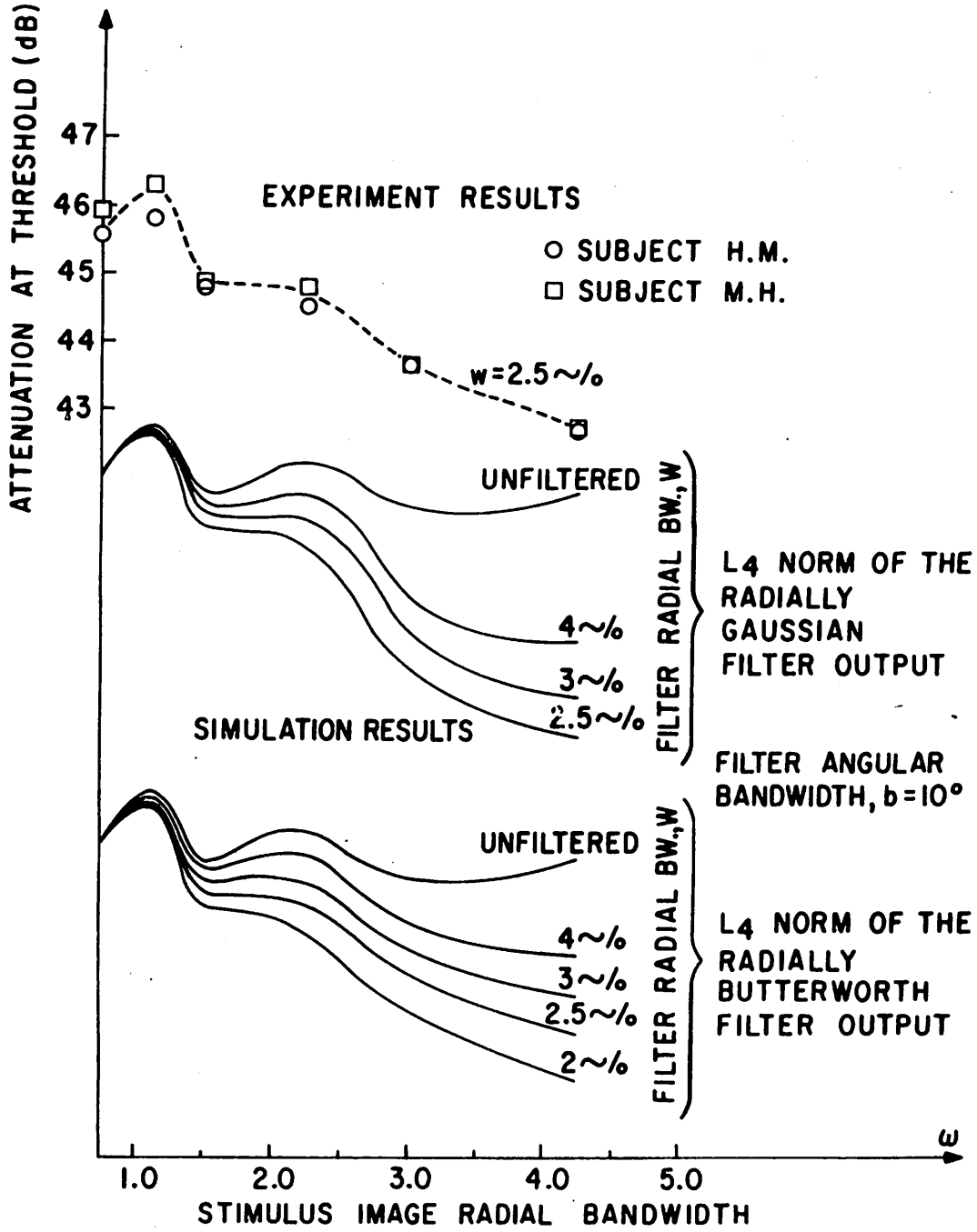


Fig. 9. The response of visual system versus the radial bandwidth of the stimulus pattern; experimental results and simulation results for both radially Gaussian and Butterworth filters.

the experimental results and simulation results we notice that the actual relative response of the visual system for large stimulus bandwidths drops off more slowly than that predicted by any of the filter bandwidths used in the simulation. If we simply increase the value of w,we still cannot get a good fit because the experimental data fits the simulation results for w between 2.5 ~/0 and 3 ~/0 for stimulus bandwidths up to ω = 2.25 ~/0. It is only for larger values of ω that the model response decreases faster than the visual system response. Clearly this can be due to the particular shape of the radial filter that we have chosen. The Gaussian radial filter has tails that drop off very rapidly.

These observations suggested the use of a different radial filter which has a frequency response that approximates the Gaussian shape for small values of $f - f_o$, while for large values of $f - f_o$ drops off more slowly than the Gaussian filter. One such radial filter which produced satisfactory results has the transfer function

$$H_r(f) = \frac{1}{\sqrt{1 + \left(\frac{f-f_o}{w}\right)^2}} \qquad (4.9)$$

Using the filter

$$H(f,\theta) = H_r(f)\ H_a(\theta)$$

where $H_r(f)$ is given above and $H_a(\theta)$ is given by Eq. (4.6), simulation results were computed for values of w ranging from 2 ~/0 to 4 ~/0. These are shown by the curves at the bottom of Fig. 9. The curve for w = 2.5 ~/0 matches the experimental data.

A dashed replica of this curve is drawn through the experimental points. Thus we take the shape of the radial filter to have the form given by Eq. (4.9) (Butterworth filter) with w = 2.5 $\pi/0$.

We note that although we used a Gaussian shaped radial filter in estimating the angular bandwidth b in sec. 4.4, that estimate is still valid because the simulation results were insensitive to the width of the radial filter, hence the change from Gaussian to Butterworth cannot cause any significant change in the simulation results.

4.6. Estimation of the value of p

In order to determine the value of p in the L_p norm functional that operates on the output of the channel filter, an experiment was conducted where as in the previous experiments, the threshold value of α in Eq. (4.1) for a set of test patterns was measured and the results were compared with the simulation results that use various values of p in the model. In these simulations the transfer function of the filter was assumed to have the form found by the last two experiments, independent of the value of p.

For this experiment we need to have a set of stimulus patterns that, unlike the sample functions of Gaussian random patterns, make the variation of the channel response to different test patterns sensitive to the value of p. For this reason non-Gaussian narrow band random patterns were used in this experiment.

The five test patterns $n_q(x,y)$, q = 1, 1.5, 2, 3, 5 were generated as follows:

A sample function n(x,y) of a narrow band zero mean Gaussian

random pattern, with the radially symmetric power spectral density of Eq. (4.2) (f_o = 4.5 ~/0 and ω = 1.25 ~/0) was chosen. For each value of q given above, the function $m_q(x,y)$ was computed by the nonlinear operation

$$m_q(x,y) = |n(x,y)|^q \text{sgn}[n(x,y)] \tag{4.10}$$

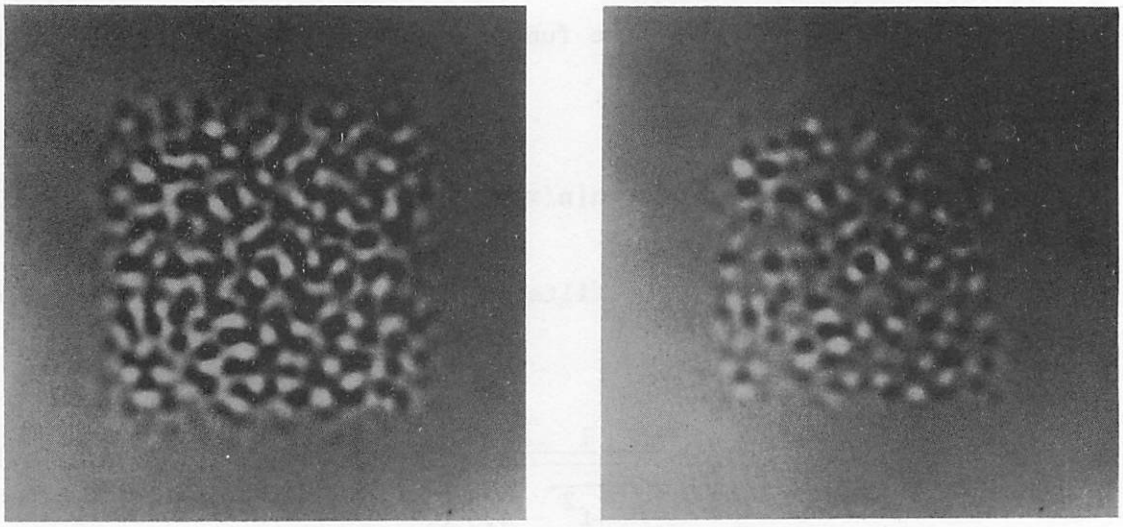
Then $n_q(x,y)$ was found by filtering $m_q(x,y)$ by an isotropic filter defined by

$$G(f_x,f_y) = \frac{1}{\sqrt{1 + \left(\frac{\sqrt{f_x^2+f_y^2} - 4.5}{3.5}\right)^8}} \tag{4.11}$$

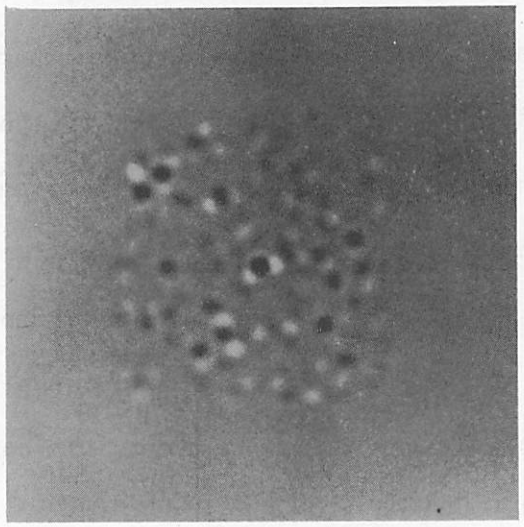
We note that $m_q(x,y)$ for q = 1, 1.5, 2, 3, 5 can be regarded as a sample function of the zero-mean random pattern $M_q(x,y)$ which is non-Gaussian for q = 1.5, 2, 3, 5. But for these values of q it will not be quite narrow band either. Therefore each $m_q(x,y)$ was filtered by $G(f_x,f_y)$ in order to remove part of the frequency components introduced by the nonlinear operation and keep the spatial frequency components of $m_q(x,y)$ in a band which ensures the excitation of only the channel that is under study. Since the bandwidth of $m_q(x,y)$ is not very large compared with the bandwidth of the filter $G(f_x,f_y)$, $n_q(x,y)$ remains non-Gaussian for q = 1.5, 2, 3, 5.

The test patterns were scaled so that the value of $\max_{x,y}|n_q(x,y)|$ was the same for all of them. Three examples of these patterns are shown in Fig. 10.

The threshold value of α(in dB attenuation) as a function of q is plotted in Fig. 11 for two subjects.



q = 2 q = 3



q = 5

Fig. 10. Three examples of the pattern $n_q(x,y)$ used for determining the value of p.

The simulation results were obtained by filtering the same five test patterns that were used in the experiment by the filter given by Eq. (4.4), (4.9) and (4.6). The bandwidth parameters of this filter were those found in previous experiments, namely $b = 10°$ and $w = 2.5$ ~/0 (for Butterworth radial filter). The center frequency of the filter was $f_o = 4.5$ ~/0. However since the stimulus patterns of this experiment are sample functions of isotropic random patterns, the choice of the orientation θ_o depends on what channel is actually detecting each of the test patterns. Therefore θ_o was treated as a parameter to which the sensitivity of the simulation results had to be investigated.

Using various values of θ_o and p the response of the channel was computed. It was noticed that the simulation results were almost insensitive to the value of θ_o; therefore the value of $\theta_o = 0°$ was used. Each of the curves in Fig. 11 show, for a fixed value of p, the variation of channel response as a function of q which is a measure of the deviation of the stimulus probability distribution from Gaussian. The curves corresponding to nine different values of p ranging from 2 to ∞ are shown.

The simulation results show a good sensitivity to changes in the value of p and it can be seen that the curve corresponding to $p = 8$ matches the experimentally measured response of both subjects. This estimate of the value of p is within the range of low values (1-10) that was consistent with the previous experiments.

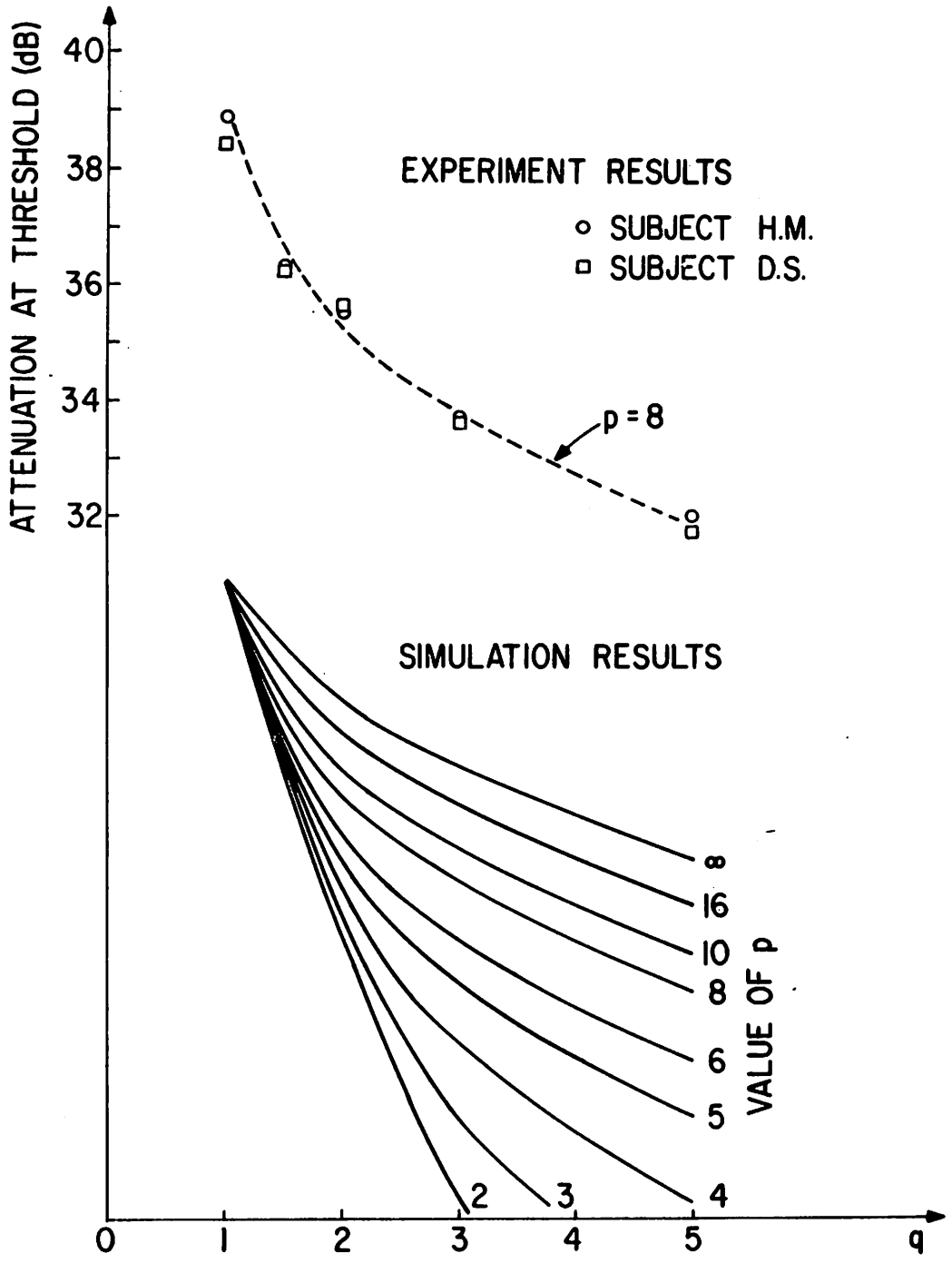


Fig. 11. Response of the visual system as a function of the distribution of the pseudo random patterns; experimental data and simulation results.

CHAPTER 5

EXPERIMENT RESULTS (II)

VISIBILITY OF PSEUDO RANDOM DISTORTION IN NARROW BAND BACKGROUND IMAGES

The experiments of the previous chapter dealt with the detectability of pseudo random patterns on a constant luminance background. It is well known that the visibility of distortion in images depends strongly on the characteristics of the image itself as well as the nature of the distortion. The experiments of this chapter were designed to investigate the effect that various characteristics of a background image have on the threshold detection of pseudo random perturbations in that image.

The self setting method is used to measure the threshold value of α for detecting the difference between the image $i(x,y)$ and the distorted image

$$\hat{i}(x,y) = i(x,y)[1 + \alpha n(x,y)] \tag{5.1}$$

The background image $i(x,y)$ is a non-negative function given by

$$i(x,y) = I_o + u(x,y) \tag{5.2}$$

where I_o is the mean background luminance and the pattern $u(x,y)$, which has a contrast well above the threshold of perception, represents the background image information. In most of the experiments $u(x,y)$ was a sample function of a narrow band random pattern with the power spectral density of Eq. (4.2).

In the experiments of this chapter we want the perturbation pattern $n(x,y)$ and the background pattern $u(x,y)$ correspond to two independent random patterns. Therefore in each stimulus image $n(x,y)$

and u(x,y) were generated by filtering two different white noise sample functions.

5.1. The Effect of Background Image Center Frequency

In order to investigate the effect of the spatial frequency components of a background image on the visibility of narrow band random perturbations, the measurements of the experiment of sec 4.1 were repeated for background images that were narrow band.

Four different background images were used in this experiment. In two of them u(x,y) was a narrow band pseudo random pattern with center frequencies 4.5 ~/0 and 13 ~/0; the bandwidth of each was 1/5 the center frequency. In the other two u(x,y) was given by

$$u(x,y) = c\, J_o(2\pi f_o \sqrt{x^2+y^2})$$

where $J_o(\cdot)$ is the Bessel function obtained by the Hankel transform of the delta function $\delta(f-f_o)$, with the center frequency f_o equal to 4.5 ~/0 for one pattern and 12 ~/0 for the other. The rms value of u(x,y) for the case of pseudo random patterns was 1/3 the value of the mean luminance I_o, and the peak amplitude of u(x,y) for the radial Bessel functions was 1/3 the value of I_o. In Figs. 1a and 1c two examples of these background patterns are shown. Figures 1b and 1d show the perturbed versions of these images as given by Eq. (5.1) for a narrow band n(x,y) with a visible contrast.

As in the experiment of sec. 4.1, for each background image the threshold value of α as a function of the center frequency of n(x,y) was measured. The upper curves of Fig. 5.2 show the rms contrast sensitivity $1/\|\alpha n(x,y)\|_2$, in dB, as a function of the center

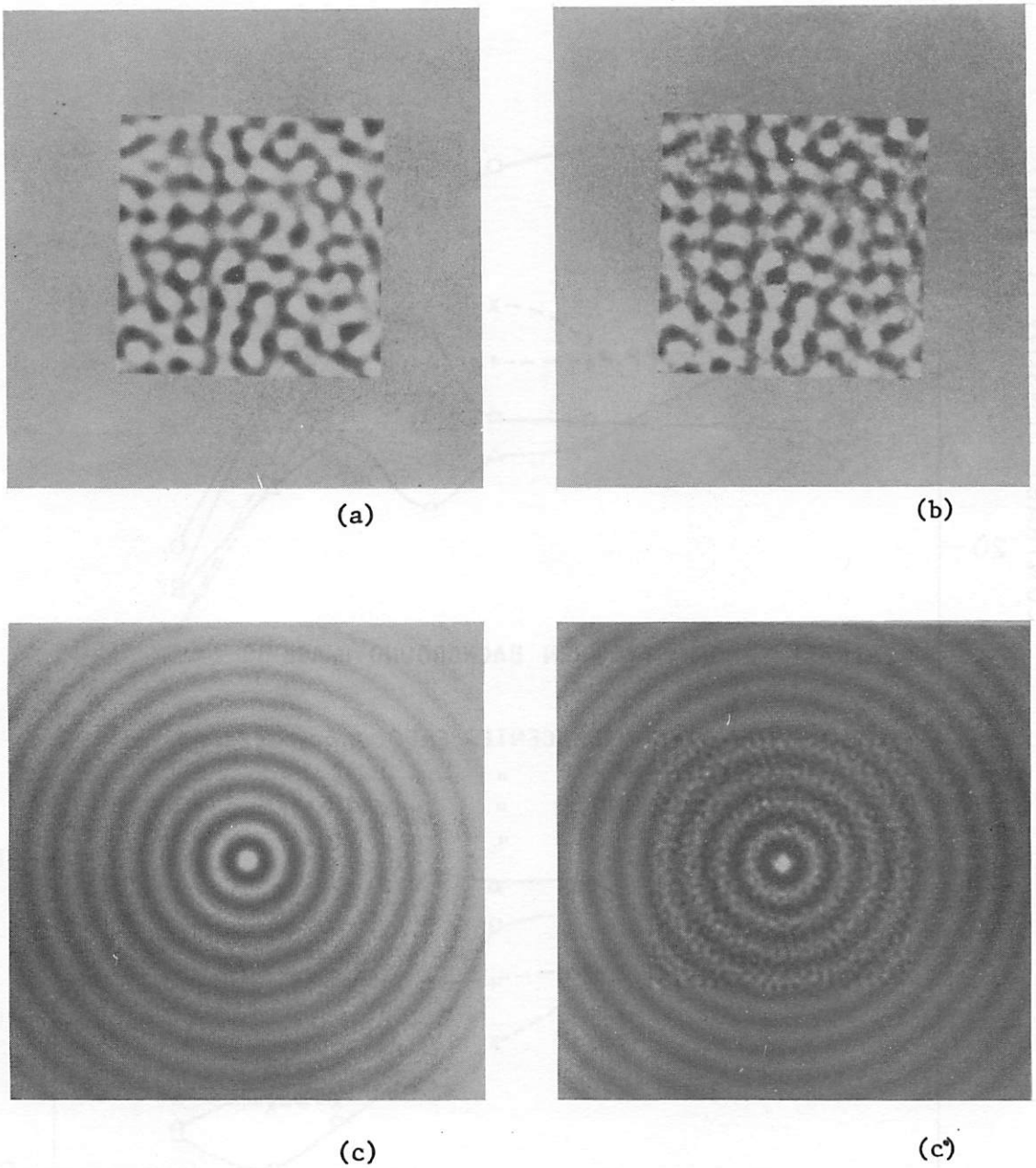


Fig. 1. Examples of background images and distorted background images (see the text).

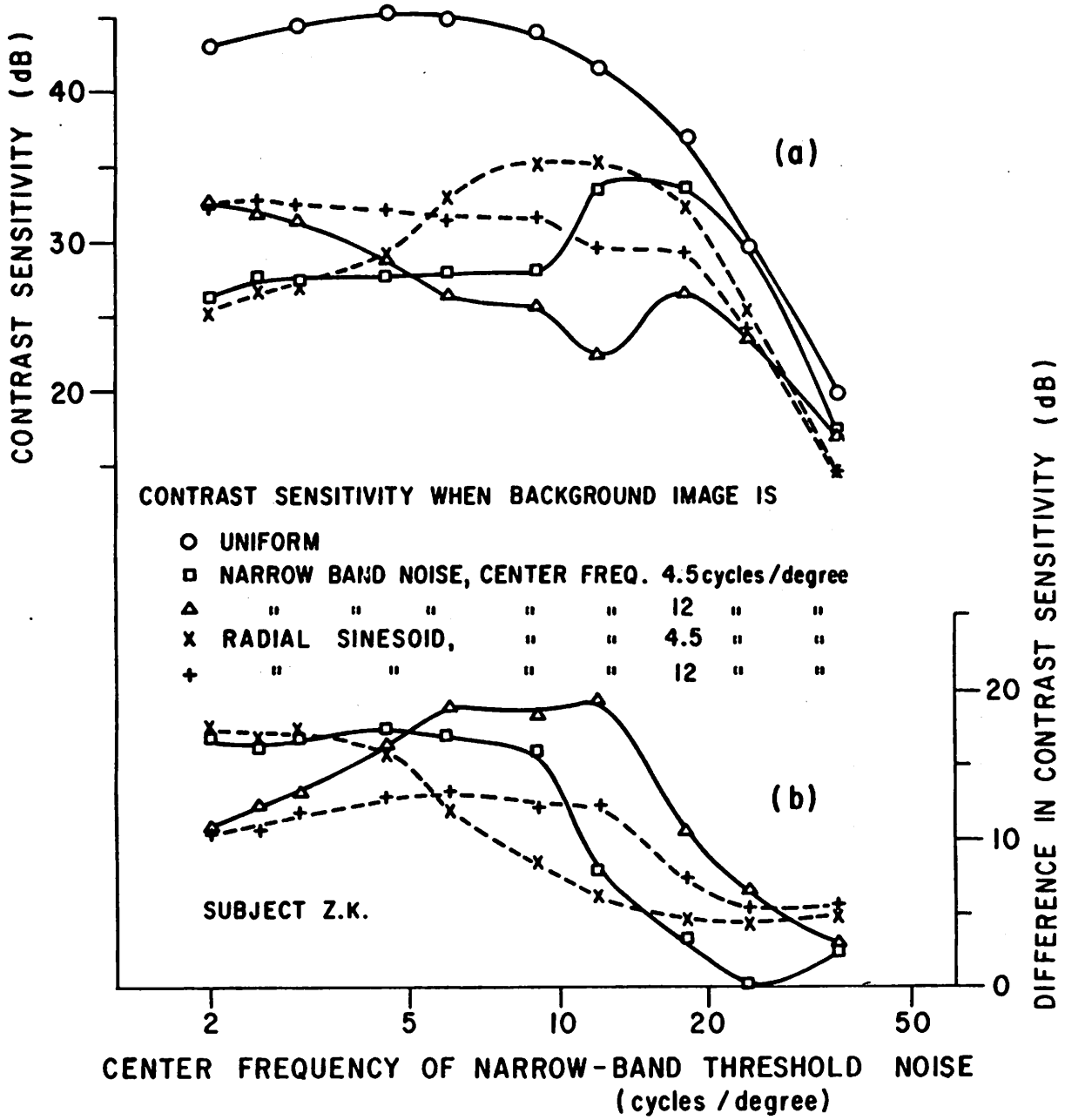


Fig. 2. Contrast sensitivity as a function of center frequency for different background images.

frequency of n(x,y) for different background images, including the curve corresponding to the blank background which is replotted from Fig. 4.3. The lower curves in Fig. 5.2 show the difference in contrast sensitivity between the constant luminance background and different narrow band background images.

The depression in the contrast sensitivity curve is maximum when the center frequency of the test pattern n(x,y) is the same as that of the background pattern u(x,y), but this depression extends over a range of frequencies that is large compared with the bandwidth of the background image.

5.2. The Effect of Background Contrast

In order to investigate the effect of the contrast of a narrow band background pattern on the visibility of a narrow band test pattern an experiment was conducted in which u(x,y) was a narrow band pseudo random pattern with different contrast values. The center frequencies of the test pattern and the background patterns were identical.

By the self setting method the threshold rms contrast of the test pattern as a function of the background rms contrast was measured for two subjects. These measurements were carried out for center frequencies 4.5 ~/0 and 13 ~/0; the results are plotted in Fig. 3.

We can see that one of the subjects (D.S.) has in general a higher sensitivity. But aside from that we can conclude that the threshold contrast of the test patterns increases with the contrast of the background pattern. Also the rise in threshold, due to the

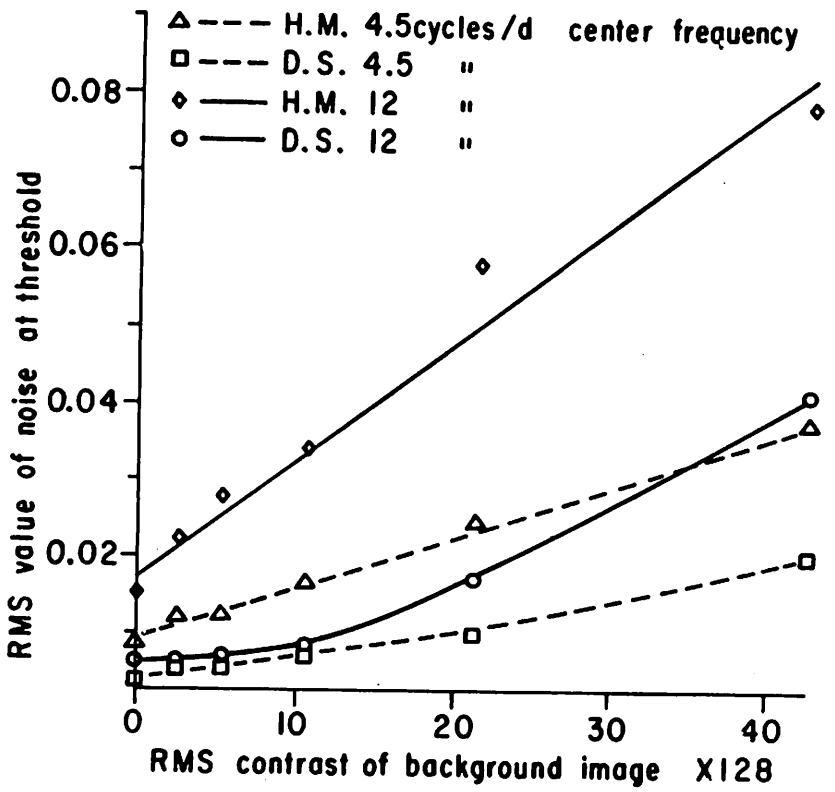


Fig. 3. Threshold contrast of narrow band noise as a function of the contrast of a background with the same power spectral density.

background, is almost a linear function of the background contrast.

The experimental results of this chapter indicate that, in general the presence of the background image considerably reduces the visibility of the test pattern. According to the discussion of sec. 3.1, the multiplicative form of Eq. (5.1) causes this reduction in sensitivity to be due to reasons other than the local spatial interaction of the retinal cells that was described in sec 22. In sec. 6.2 we will present a modified version of the perception model which takes the phenomena observed in the above experiments. into account.

CHAPTER 6

CONCLUSIONS REGARDING THE MODEL AND THE DISTORTION MEASURE

In this chapter the experimental results and their implications will be discussed and a modified version of the model of Chapter 2 which takes the effect of the background image into account will be presented. Finally the possible distortion measures that are suggested by this model will be discussed.

6.1. Implications of the Experimental Results

The experimental results of sec. 4.2 support the multichannel hypothesis about the model given in Fig. 5 of Chapter 2. These results indicate the existence of channels that are selectively sensitive to different spatial frequencies and as indicated by the OR function in the model, the detection of a given pattern takes place when the response of any one of the channels reaches its threshold. The experiment described at the beginning of sec. 4.3 indicates that the same statement is true about the channels that are sensitive to different orientations.

The results of the experiments of secs. 4.3, 4.4 and 4.5 provide us with quantitative information about the filter in the channel which is sensitive to frequencies around $f_o = 4.5$ ~/0 and orientations around $\theta_o = 0°$ i.e. along the f_x axis in the frequency domain. In other words this is the channel that is most sensitive to a 4.5 ~/0, vertically oriented grating. For the frequency response given by Eqs. 4.4, 4.6 and 4.9 the radial bandwidth w and the angular width b of this filter were found to be $\pm$ 2.5 ~/0 and $\pm$ 10° respectively. The magnitude of the frequency response of

this filter is illustrated by a three dimensional plot in Fig. 1.

The center of this plot corresponds to the origin of the frequency plane and it shows the magnitude of the transfer function over the range of frequencies $-32\ ^{\sim}/0 \leq f_x \leq 32\ ^{\sim}/0$, $-32\ ^{\sim}/0 \leq f_y \leq 32\ ^{\sim}/0$. All of the filters used in the simulation studies of Chapter 4 have real valued transfer functions. No attempt was made to find the phase response of the channel filter. Fortunately since the stimulus images were Gaussian random patterns, and the output of the filter was operated on by a L_p-norm operation, it is unlikely that the simulation results will change significantly by changing the phase response of the filter while the magnitude of the frequency response is kept fixed. This is due to the fact that the L_2-norm of the output of the filter is approximately equal to the integral of the product of the power spectral density of the noise pattern and the square of the magnitude of the frequency response of the filter. Thus the L_p-norm of the output for a Gaussian random pattern should depend only on the magnitude of the transfer function.

Therefore the use of Gaussian random patterns has enabled us to determine the magnitude response of the filter independently from the shape of the phase response of the filter. It should be pointed out that in order to apply this model to the detection of patterns that are not random, in most cases the shape of the phase response has to be known.

The point spread function of the filter i.e. the inverse transform of the frequency response can be computed under the assumption of a real frequency response. This is shown by the three dimensional plot of Fig. 2. The center of this plot corresponds

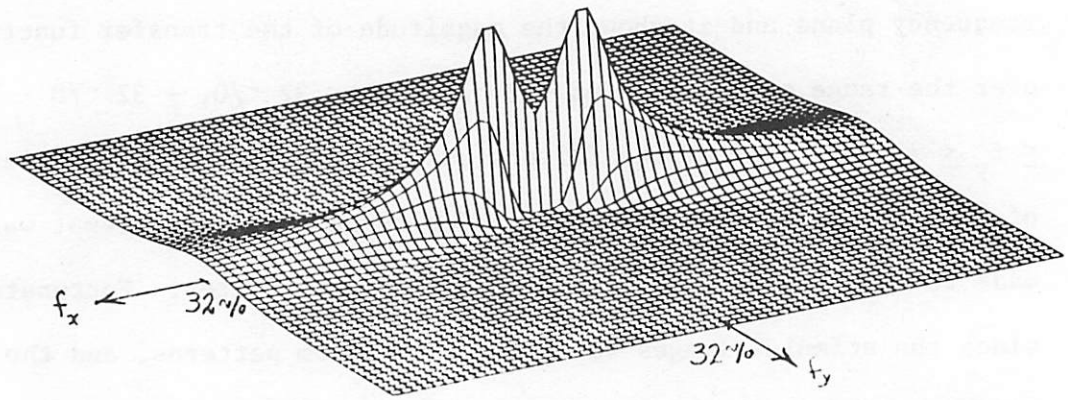


Fig. 1. Magnitude of the frequency response of the filter with $f_o = 4.5$ c/o and $\theta_o = 0°$.

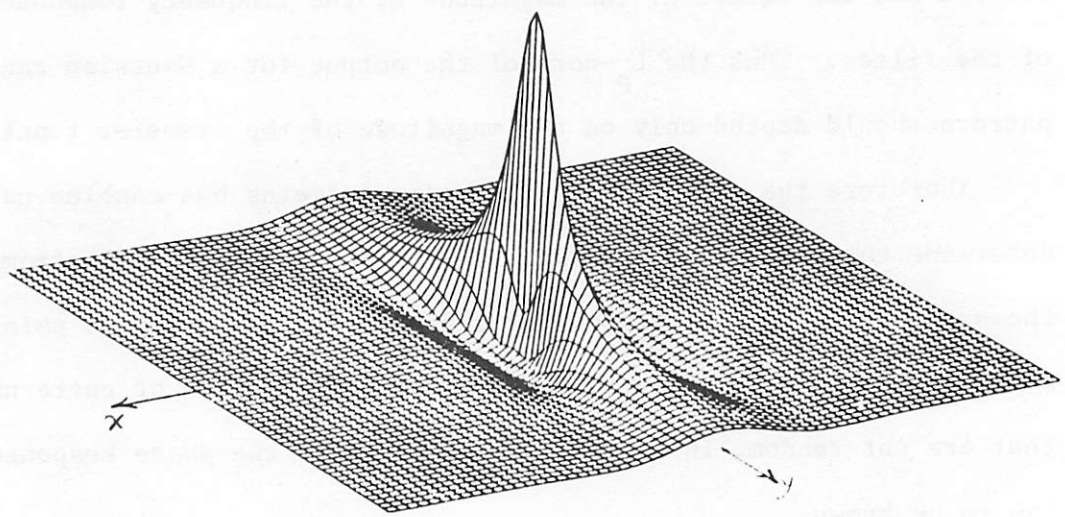


Fig. 2. Point spread function for the real-valued transfer function.

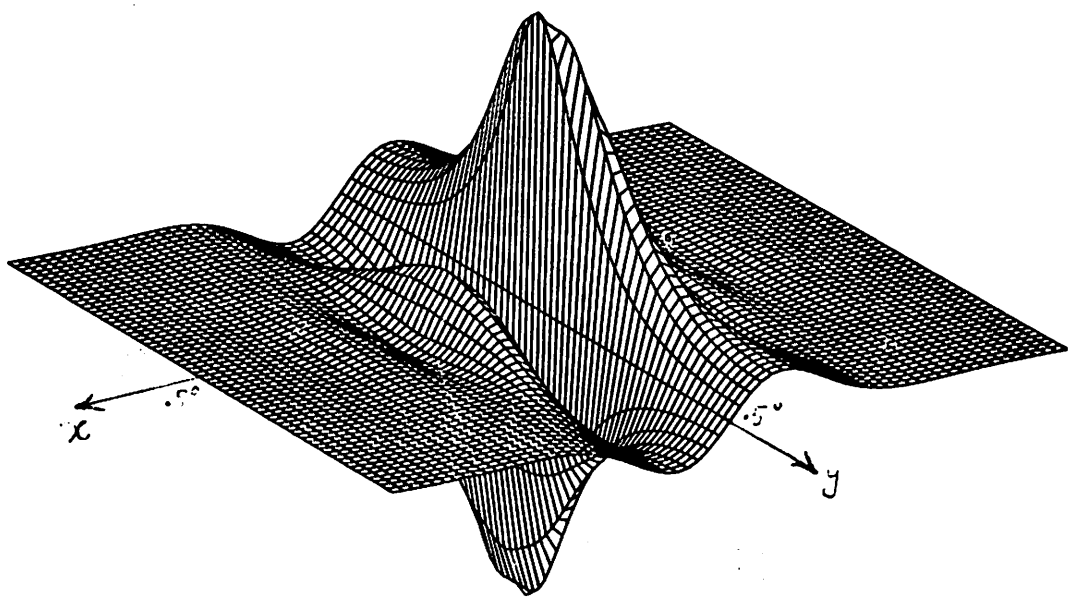


Fig. 3. Point spread function with the assumption of symmetry in the y direction and antisymmetry in the x direction.

to the origin of the spatial coordinates and it shows the point spread function over the range $-.5° \leq x \leq .5°$, $-.5° \leq y \leq .5°$. Notice that the assumption of a real transfer function has resulted in having symmetry in both x and y directions.

If we assume that the frequency response of Fig. 1 corresponds to a point spread function that is symmetric in the y direction and antisymmetric in the x direction, then it can be computed by assuming an imaginary frequency response with proper conjugate symmetries. This point spread function is shown in Fig. 3. Clearly the point spread function can have a form different from these two extreme cases and in genral corresponds to a complex-valued frequency

response. A more plausible alternative is to hypothesize that there exist channels whose filters have the same magnitude response, but different phase responses. Then the two extreme cases shown in Fig. 2 and Fig. 3 can be identified with the line detectors and edge detectors, respectively, that have been investigated by Kulikowsky and Kingsmith [21]. In fact the variation of these point spread functions along the x axis is both quantitatively and qualitatively similar to the line spread functions that Kulikowsky and Kingsmith have found for these detectors.

A few observations about the physilogical correlates of the model can be made at this point. The linear filtering and the norm operation can correspond to spatial interaction at all levels of the visual nervous system, from the retina to the cortical cells. The retinal cells have receptive fields that are almost isotropic. The angular selectivity must correspond to the spatial interaction at higher levels of the visual process. In fact the cortical cells of the cat, studied by Hubel and Wiesel [16] have simple and complex receptive fields that show such angular selectivity. Receptive fields with shapes corresponding to both point spread functions of Figs. 2 and 3 have been discovered in these studies.

The value of $p = 8$ in the L_p-norm operation that follows the filter seems to be in agreement with subjective experience. When a spot with large amplitude dominated the test pattern, subjects detection was almost entirely based on detecting this peak; this corresponds to a value of p that is higher than 2 or 3. On the other hand when no such dominant peak was present, as in the case of Gaussian patterns, the detection was due to contributions from all parts of the image rather than the point with maximum amplitude;

this corresponds to a value of p that is less than infinity.

In estimating the numerical values of the model parameters, we have concentrated only on the channel with center frequency f_o = 4.5 ~/0 and orientation θ_o = 0°. However some observations about the parameters of other channels can be made. First the result of the experiment of sec. 4.1 shows how the sensitivity of the channels tuned to different spatial frequencies depends on the center frequency of the channel. As for the bandwidth of the other channels, the experimental results of Blackmore and Campbell and several other investigators indicate that within a large range of spatial frequencies the ratio of the center frequency to the bandwidth of the filters are constant. Also the orientational selectivity for gratings, expressed in terms of the polar angle in the frequency plane, has been found to be constant for different grating frequencies (Kulikowsky, Abadi and Kingsmith [22]). The implication of these results in our model is that while the angular width b remains fixed, the radial bandwidth w is proportional to the center frequency. Therefore channels with larger center frequency have wider bandwidth in both f_x and f_y directions in the frequency plane.

Another parameter of the model that was not investigated in our experiments is the shape and effective support of the spatial weighting functions $a_k(x,y)$ in the L_p-norm operation. In our simulation studies this function was assumed to be a constant and hence was ignored. This was due to the fact that Robson's data (Fig. 4 of Chapter 2) indicates that the spatial extent of a(x,y) for the channels around 4.5 ~/0 is greater than 10°.

In fact these experimental results show that for the 4.8 c/o grating, the response of the visual system increases drastically by increasing the number of cycles displayed, up to 45 cycles which is equivalent to 10°. Since our stimulus patterns in all cases extended over a 2° x 2° frame and the channel filter is not very narrow band we can assume that over the area in which the output pattern of the filter is non zero, this function is almost constant.

6.2. The effect of the background image

The block diagram of Fig. 4 shows a typical channel of a modified version of the model given by Fig. 5 of Chapter 2. In this model the effect of the background image is taken into account by introducing a separate channel that processes the background image $i(x,y)$ only and whose output U_k scales down W_k the response of the channel to the perturbation pattern $\ell n[i(x,y)] - \ell n[i(x,y)]$. This structural form is in agreement with the experimental results of Chapter 5. The quantity U_k is assumed to be a spatially weighted norm of the output of a filter $G_k(f_x,f_y)$. We assume that the d-c response of this filter, i.e. $G(0,0)$ is zero. This is because we are only interested in the effect of the contrast of $i(x,y)$ on the detectability of the perturbation. It is clear that for a zero background contrast this model reduces to the one given in Chapter 2.

The results of sec. 5.1 indicate that the depression of the contrast sensitivity curve is maximum at the center frequency of the background image, suggesting that this filter is tuned to the center frequency of the filter $H_k(f_x,f_y)$. However since the

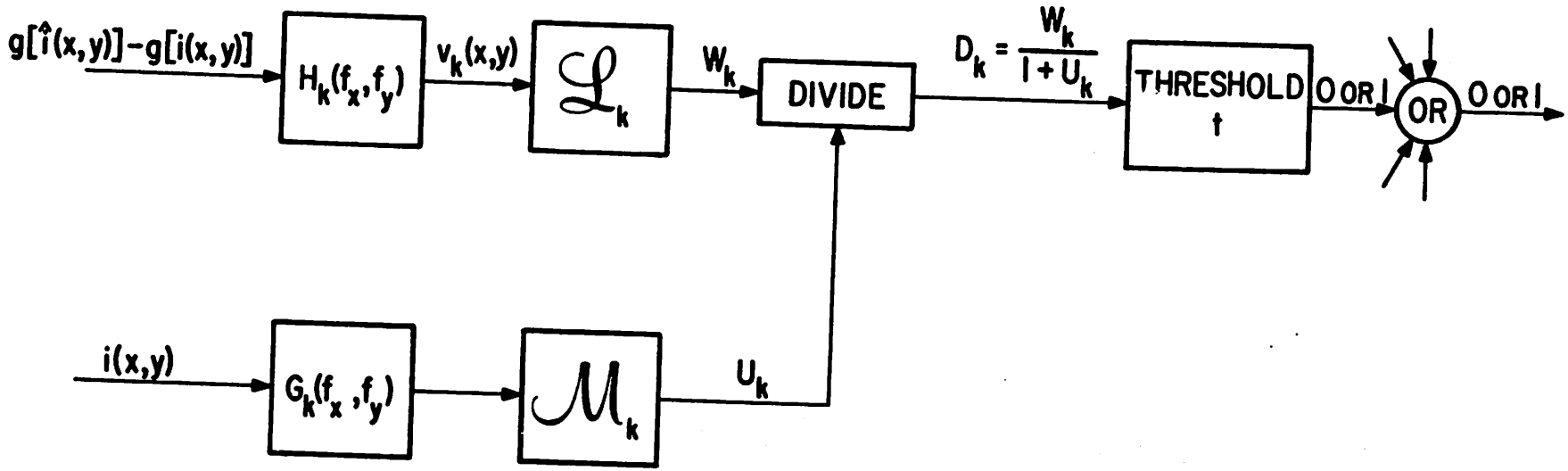


Fig. 4. Block diagram of a typical channel of the model of detection of perturbations in background images.

depression of the contrast sensitivity function (Fig 2 of Chapter 5) due to a narrow band background (masking effect) extends over a wide band of frequencies, the bandwidth of $G_k(f_x,f_y)$ has to be large compared with that of $H_k(f_x,f_y)$. It seems plausible to assume that the spatial weighting function for the channel that processes i(x,y) is the same as that of the channel which processes the perturbation pattern.

The experimental results of sec. 5.2 suggest that the overall channel response D_k should be related to W_k and U_k by

$$D_k = \frac{W_k}{1+U_k} \tag{6.1}$$

To see this we note that the assumption of a linear filter makes the value of U_k, for a given Gaussian background pattern, proportional to the r.m.s value of the background contrast, c_b

$$U_k = A\ C_b \tag{6.2}$$

where A depends on the relation between the background image bandwidth and the frequency response $G_k(f_x,f_y)$ as well as the value of p in the L_p-norm operation which generates U_k. Also W_k is proportional to c_n, the rms contrast, of the perturbation pattern which in this experiment was

$$\ell n[i(x,y)] - \ell n[\hat{i}(x,y)] \approx \alpha n(x,y) \tag{6.3}$$

Thus we have

$$W_k = B_n \tag{6.4}$$

If the value of the threshold is taken to be t, at the threshold we have

$$t = \frac{BC_{n,t}}{1+C_b}$$

or

$$C_{n,t} = \frac{t}{B} + \frac{tA}{B} C_b \qquad (6.5)$$

where $C_{n,t}$ is the threshold r.m.s. contrast of the noise pattern. This relation is consistent with the experimental results of sec. 5.2 (Fig. 6 of Chapter 5).

We notice that the perturbation pattern is taken to be $n[\hat{i}(x,y)] - \ell n[i(x,y)]$ and that U_k does not depend on the mean luninance of the background image. These are suggested by the fact that the perturbed images used in the experiments of Chapter 5 had the multiplicative form of Eq. 5.1 and consequently the results of these experiments correspond to phenomena different from the adaptation of the photoreceptors of the retina.

Unfortunately the experimental results of Chapter 5 are not adequate for making a definitive statement about the relation between the background image i(x,y) and the value of U_k. Also we have used only a limited number of background images, and there are undoubtedly images for which this model will not accurately predict the detection of perturbations. This model does however predict some of the masking phenomena observed in practice. For example it is known that if the background image has a sharp transition in luminance along a boundary, then the perturbations are less visible around this boundary. In our model this corresponds to the response

U_k that the "edge" in the background image generates and the subsequent decrease in the responses D_k of the channels over a wide band of spatial frequencies.

6.3. The Distortion Measure

Based on the model presented in the last section we are going to define several distortion measures and consider their implications. The distortion measures will be for the encoding of still monochromatic images.

Let the non-negative function $u(x,y)$, which is nonzero only on the subset $[-\frac{S}{2}, \frac{S}{2}] \times [-\frac{S}{2}, \frac{S}{2}]$ of $\mathbb{R}^2$, represent an image. Suppose S is small enough so that the variables x and y correspond to subtended angles in the horizontal and vertical direction from the point of view of a human observer. We assume that the fixation point of the visual system of the observer coincides with the origin (0,0). Therefore each image corresponds to a specific fixation point of the observer as well as his distance from the scene.

Let $\hat{u}(x,y)$ be a reconstructed version of $u(x,y)$ at the receiver. We first ocnsider a vector-valued function $\underset{\sim}{D}(u,\hat{u})$ defined in terms of its non-negative components $D_k(u,\hat{u})$, $k = 1,2,\ldots,n$ by

$$D_k = \frac{W_k(u,\hat{u})}{1+U_k(u)} \qquad k = 1,2,\ldots,n \qquad (6.6)$$

where $W_k(u,\hat{u})$ and $U_k(u)$ are the same as W_k and U_k shown in Fig. 4, with $i(x,y) = u(x,y)$ and $\hat{i}(x,y) = \hat{u}(x,y)$. (Note that the notation $i(x,y)$, previously used for an image has been changed to $u(x,y)$ in this section) We are assuming that there is a one to one correspondence

between the components of $\underset{\sim}{D}(.,.)$ and the channels of the model. From the last section the values of $W_k(u,\hat{u})$ and $U_k(u)$ are given by

$$W_k(u,\hat{u}) = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_k(x,y)\left|[(\ell nu-\ell n\hat{u})*h_k](x,y)\right|^p dxdy\right]^{1/p} \tag{6.7}$$

$$U_k(u) = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_k(x,y)\left|[u*g_k](x,y)\right|^{p'} dxdy\right]^{1/p'} \tag{6.8}$$

Notice that if we require that u(x,y) represent all of the visual information perceived by the human observer for a given fixation point, then the size of the frame $[-\frac{S}{2}, \frac{S}{2}] \times [-\frac{S}{2}, \frac{S}{2}]$ needs to be chosen large enough so that the support of all of the spatial weighting functions $a_k(x,y)$, $k = 1,2,\ldots,n$ is covered by this frame.

Suppose the threshold t is taken to be identical for all of the channels, thus accounting for different sensitivities of the channels to distortion by proper scaling of the transfer function $H_k(f_x,f_y)$ in each channel. With this in mind the prediction of the model of Fig. 4 is that if

$$D_k(u,\hat{u}) < t \qquad \text{for all } k = 1,2,\ldots,n$$

then the difference between u(x,y) and $\hat{u}(x,y)$ is not detectable. On the other hand if for some $k, D_k \geq t$, then this difference becomes visible.

Based on the function $\underset{\sim}{D}(.,.)$ we can define various fidelity criteria for encoding and transmission of monochromatic pictures. In what follows we will study some of these fidelity criteria and

discuss their possible implications. We start by choosing a mathematical representation for the image sources that are often encountered in practice .

Let $\mathcal{U}$ be the set of all functions u(x,y), each representing an image in the sense defined at the beginning of this section. We assume that all of the images $\hat{u}(x,y)$, that are reconstructed at the receiver also belong to this set. We consider an information source whose output is a randomly chosen sequence of images from the set $\mathcal{U}$. An image source may therefore be characterized by a probability measure that is defined on the set $\mathcal{U}$. The random output of the source is denoted by U, where U takes on values from the set $\mathcal{U}$. Also the random image reconstructed at the receiver is denoted by $\hat{U}$.

We recall that each image u(x,y) was defined with reference to a specific fixation point of an observer. The sequence of images generated by the source can therefore be a mathematical representation of various real situations. For example the images in this sequence may correspond to a sequence of adjacent fixation points of an observer, as his visual system scans a frame of a picture or an actual scene. In this case the whole picture frame or scene is represented by a sequence of finite length of the source outputs. A sequence of source outputs can also correspond to a fixed observer and a picture that varies with time. In this case any finite segment of the sequence represents a finite time history of the moving picture.

A statistical model in which picture frames are represented by finite sequences requires that we have a probability distribution

defined on the set of all sequences of finite length of images $u(x,y)$. The set of all such sequences of length N is denoted by $\mathcal{U}^N$. The random sequence of images of length N is denoted by $\underset{\sim}{U}^N = (U_1, U_2, \ldots, U_N)$, where $\underset{\sim}{U}^N$ takes on values from the set $\mathcal{U}^N$. The random sequence of images reconstructed at the receiver is denoted by $\hat{\underset{\sim}{U}}^N = (\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_N)$

Some practical situations may be modeled by a sequence of independent and identically distributed random images $U_1, U_2, \ldots$ from the source $\mathcal{U}$. This can, for example, correspond to transmitting a sequence of images that result from widely spaced fixation points of the observer, with the scene being represented by a two dimensional homogeneous random field.

It should be pointed out that many practical situations do not lend themselves to statistical modeling, and even if they do, the probability distribution of the sequences of images may be very difficult to find.

Suppose the probability distribution of a source is known. The sequence of images generated by this source is encoded and transmitted over a communication channel and at the receiver a corresponding sequence of images is reproduced. The randomness of the source outputs and the random errors in the channel cause the reconstructed image sequence to be also random. Thus the combination of the source, the encoder and the channel give a joint probability distribution, $P_{\underset{\sim}{U}^N, \hat{\underset{\sim}{U}}^N}$ for the sequences $\underset{\sim}{U}^N$ and $\hat{\underset{\sim}{U}}^N$.

Since a picture is represented by a finite sequence of source outputs $\underset{\sim}{u}^N = (u_1, u_2, \ldots, u_N)$, the fidelity criteria must be defined

on the set of all pairs of sequences $\underset{\sim}{u}^N$ and $\hat{\underset{\sim}{u}}^N$. However the function $\underset{\sim}{D}(.,.)$ given by Eqs. (6.6-6.8) is defined for a single image pair $(u,\hat{u})$. Therefore any <u>single image distortion measure</u> $d(u,\hat{u})$ defined on the basis of $\underset{\sim}{D}(.,.)$ has to be extended to a distortion measure $d_N(\underset{\sim}{u}^N,\hat{\underset{\sim}{u}}^N)$ defined on pairs of sequences of length N. We note that when $\underset{\sim}{U}^N$ and $\hat{\underset{\sim}{U}}^N$ are random sequences, $d_N(\underset{\sim}{U}^N, \hat{\underset{\sim}{U}}^N)$ is a random variable.

Suppose the user of the image transmission system requires that with probability greater than $1-\delta$, $0 < \delta < 1$, the distortion in a reproduced picture (i.e. a N-long sequence of images $\hat{u}(x,y)$) be below the threshold of detection. Clearly this requirement will be satisfied if we let the single image distortion measure be

$$d(u,\hat{u}) \triangleq \max_{k=1,2,\ldots,n} D_k(u,\hat{u}) \tag{6.10}$$

and extend this to sequence of images by letting

$$d_{max}(\underset{\sim}{u}^N,\hat{\underset{\sim}{u}}^N) = \max_{j=1,2,\ldots,N} d(u_j,\hat{u}_j) \tag{6.11}$$

and then design our system in such a way that

$$P_{\underset{\sim}{U}^N,\hat{\underset{\sim}{U}}^N}\{d_{max}(\underset{\sim}{U}^N, \hat{\underset{\sim}{U}}^N) \geq t\} < \delta \tag{6.12}$$

This fidelity criteria simply specifies an upper bound on the probability with which the encoding and transmission system introduces visible distortion in the pictures transmitted, thus implying that all pictures with visible distortion are judged as being equally distorted without giving more weight to a particular form of distortion or a particular part of the picture frame.

Another possible requirement may be expressed in terms of the average of $d(u,\hat{u})$ over the picture frame, that is we define

$$\bar{d}(\underset{\sim}{u}^N,\hat{\underset{\sim}{u}}^N) = \frac{1}{N}\sum_{j=1}^{N} d(u_j,\hat{u}_j) \tag{6.13}$$

and require that the probability of this average increasing beyond the threshold t be less than δ

$$P\{\bar{d}(\underset{\sim}{u}^N,\hat{\underset{\sim}{u}}^N)\}_{>t}\} < \delta \tag{6.14}$$

This is less restrictive than the fidelity criteria defined by $d_{max}(.,.)$ and an encoding system that satisfies 6.12 also satisfies 6.14.

The model of Fig. 4 and the fact that the threshold level, t, in 6.12 and 6.14 is taken to be constant, imply that for any single image u(x,y) the set of all possible reconstructions, $\hat{u}(x,y)$ is divided into two disjoint sets of <u>distorted</u> and <u>undistorted</u> images. This means that there is no ordering defined on the set of <u>distorted</u> reproductions and thus they are all judged equivalent in quality. The fidelity criteria that we have discussed so far puts restrictions on the relative frequency with which <u>distorted</u> images are reproduced at the receiver.

Although the model of Fig. 4 is for threshold detection of images, it is useful to discuss possible extensions of this model that can be used in evaluating the quality of the images in which the distortion is at a low but detectable level. We start by commenting about the manner in which we can define an ordering on the set of reproductions of an image.

Using the function $\underset{\sim}{D}(.,.)$ one such ordering is defined in the following way. Suppose $\hat{u}_1(x,y)$ and $\hat{u}_2(x,y)$ are two reproductions of the image $u(x,y)$ at the receiver, resulting from two different encoding algorithms. Then we say that $u_1(x,y)$ is a better (less distorted) reproduction if

$$d_{max}(u,\hat{u}_1) < d_{max}(u,\hat{u}_2) \tag{6.15}$$

where $d_{max}(.,.)$ is given by Eq. (6.10). Also when we have

$$d_{max}(u,\hat{u}_1) = d_{max}(u,\hat{u}_2) \tag{6.16}$$

then the two reproductions are taken to be equally distorted. For each transmitted image $u(x,y)$, this gives a total ordering on the set of all reproduced images. Note that the distortion in either or both of these reproductions can be above the threshold of detection. Although they are judged as being equal, the nature of the distortion of $\hat{u}_1(x,y)$ can be quite different from that of $\hat{u}_2(x,y)$. For example the distortion can be due to high frequency noise in our image and low frequency noise in the other.

The user of the transmission system may specify his fidelity criteria for levels of distortion that are slightly beyond threshold in various ways, all of them using the above ordering as the underlying criteria.

For example the fidelity criteria may be specified by requiring that

$$P\{d_{max}(\underset{\sim}{U}^N,\underset{\sim}{\hat{U}}^N) \geq \tau\} < \delta \tag{6.17}$$

where τ is now a positive number that depending on the user's

specification takes on different values greater than or equal to the fixed threshold t.

Another way of specifying the fidelity criteria is by requiring that the statistical average of the distortion $d_N(\underset{\sim}{U}^N, \hat{\underset{\sim}{U}}^N)$ be less than a given positive number Δ.

$$E\{d_{max}(\underset{\sim}{U}^N, \hat{\underset{\sim}{U}}^N)\} \leq \Delta \tag{6.18}$$

Notice that in both of these criteria the extension of the single image distortion measure to pairs of sequences can also be done by $\bar{d}(.,.)$ as defined in Eq. (6.13).

It can be argued that by using the above criteria for values of τ and Δ greater than t, we are extending the domain of the application of the model for threshold detection of distortion to describe the suprathreshold behavior of the visual perception mechanism. On the other hand however, the assumption of an exact and fixed threshold, t, for the human visual system is in itself quite unrealistic. Therefore it seems that at least for the values of τ (Eq. 6.17) within the range that corresponds to the fuzziness of the threshold, we should be able to use the above criteria. For larger values of τ, corresponding to a distortion that is clearly visible, the possibility of using the above criteria needs to be investigated.

We used the value of $\max_{k=1,2,\ldots,n} D_k(u,\hat{u})$ to define an ordering of the set of reproductions $\hat{u}(x,y)$, of the image $u(x,y)$. For suprathreshold distortion we are not restricted to this form, and other functions such as the L_p-norm of the vector $\underset{\sim}{D}(u,\hat{u})$, for some finite p, can be used to define the ordering. Psychophysical

experiments can be conducted to determine if there is a value of p that is in agreement with subjective judgement and, if there is, whether its value depends on the level of distortion.

Another possibility is that for suprathreshold distortion, each component of $\underset{\sim}{D}(u,\hat{u})$ assumes a different weight in contributing to the subjectively judged degradation of the image. Futhermore the relative weight of the components of $D(u,u)$ may change as the level of distortion is increased.

In the above discussion we indicated some of the possible ways in which the user of the transmission system may specify his fidelity criteria. However in the actual design of an encoding algorithm in addition to user's specification we are also concerned about the rate (number of bits per picture frame) that will result from the source coding process. For example the capacity of the channel used in conjunction with the source encoder determines an upper bound on the number of bits that can be used for each picture frame. Therefore the tradeoff between the user's satisfaction represented by his fidelity criteria and the channel capacity requirements plays an important role in designing image source encoders.

Suppose we decide on one of the fidelity criteria discussed above and ask what is the lowest rate with which an image source can be encoded consistent with this fidelity criteria; also how does this minimum rate change if the fidelity criteria is varied. As indicated in Chapter 1 this question can be answered by Shannon's [1] rate distortion theory.

In order to obtain a quantitative answer, however, we need to compute the rate distortion function for the image source. This

requires a knowledge of the probability distribution of the source as well as the distortion measure used. Suppose we use one of the fidelity criteria discussed above in conjunction with an image source with a knwon probability distribution to compute the rate distortion function. We note that all of these fidelity criteria are defined in terms of the function $\underset{\sim}{D}(u,\hat{u})$ which is a rather complicated function (Eqs. 6.6-6.8) of the source output $u(x,y)$ and its reproduction $\hat{u}(x,y)$. To find an analytic expression for the rate distortion function using such a distortion seems to be a very difficult task.

This difficulty seems to be mainly due to the following two characteristics of the functions $D_k(u,\hat{u})$, defined by Eq. (6.6). First, the L_p-norm of the filtered difference of image and its reproduction, for $p = 8$, does not have the mathematical tractability of the L_2 norm functionals. Secondly, from Eq. (6.6) we see that the distortion measure depends on the background image as well as the difference of the two images. Analytically, the distortion measures that depend only on the difference of the two images are more easily tractable.

However, the problem can be approached by trying to find good upper and lower bounds for the rate distortion function using proper approximation of the above distortion measure. If such bounds can be found, we can answer questions concerning the possibility of improving the efficiency of the encoding algorithm by using the distortion measure prescribed by the visual system.

For example we may come to the conclusion that if for a given rate, encoding is performed to minimize the distortion as defined

by a simpler distortion measure, then the resulting distortion in the sense of the measures defined by the function $\underset{\sim}{D}(.,.)$, may be very close to what the rate distortion function predicts as the smallest possible distortion for that rate.

Therefore it seems that the study of the properties of the rate distortion function and finding a set of good lower and upper bounds for it, is a potentially useful area for analytic research.

APPENDIX A

GENERATION OF PSEUDO RANDOM PATTERNS

The pseudo random patterns used in the experiments were generated on the IBM 1800 computer system using the picture processing system PPSYS . On this system picture files with sizes $2^n \times 2^n$, $n = 4, 5, 6, 7, 8, 9$ can be generated and processed. These picture files were either in unpacked form or in packed form. In unpacked form each 16 bit computer word is assinged to the value of one picture element (PEL). In packed form each computer word contains two PELs.

All of the images used in this research have the size 512 x 512. A sample function of a random pattern with a given power spectral density is generated as follows.

The output of an analog Gaussian noise generator is sampled and quantized using the analog to digital converter of the computer. The resulting numbers with values in the range [-32767,32767] are stored in an unpacked 512 x 512 picture file. The sampling frequency is chosen sufficiently small compared to the bandwidth of the noise generator so that the samples are uncorrelated. In this way we generate a 512 x 512 point sample function $n_w(x,y)$ of a white Gaussian noise $N_w(x,y)$ whose power spectral density $S_{N_w}(f_n,f_y)$ is the constant N_o. An example of such a pattern is shown in Fig. A1a.

If this white noise pattern is passed through a linear homogeneous (shift invariant) filter with frequency response $H(f_x,f_y)$, the power spectral density of the output random pattern, $N(x,y)$, is given by

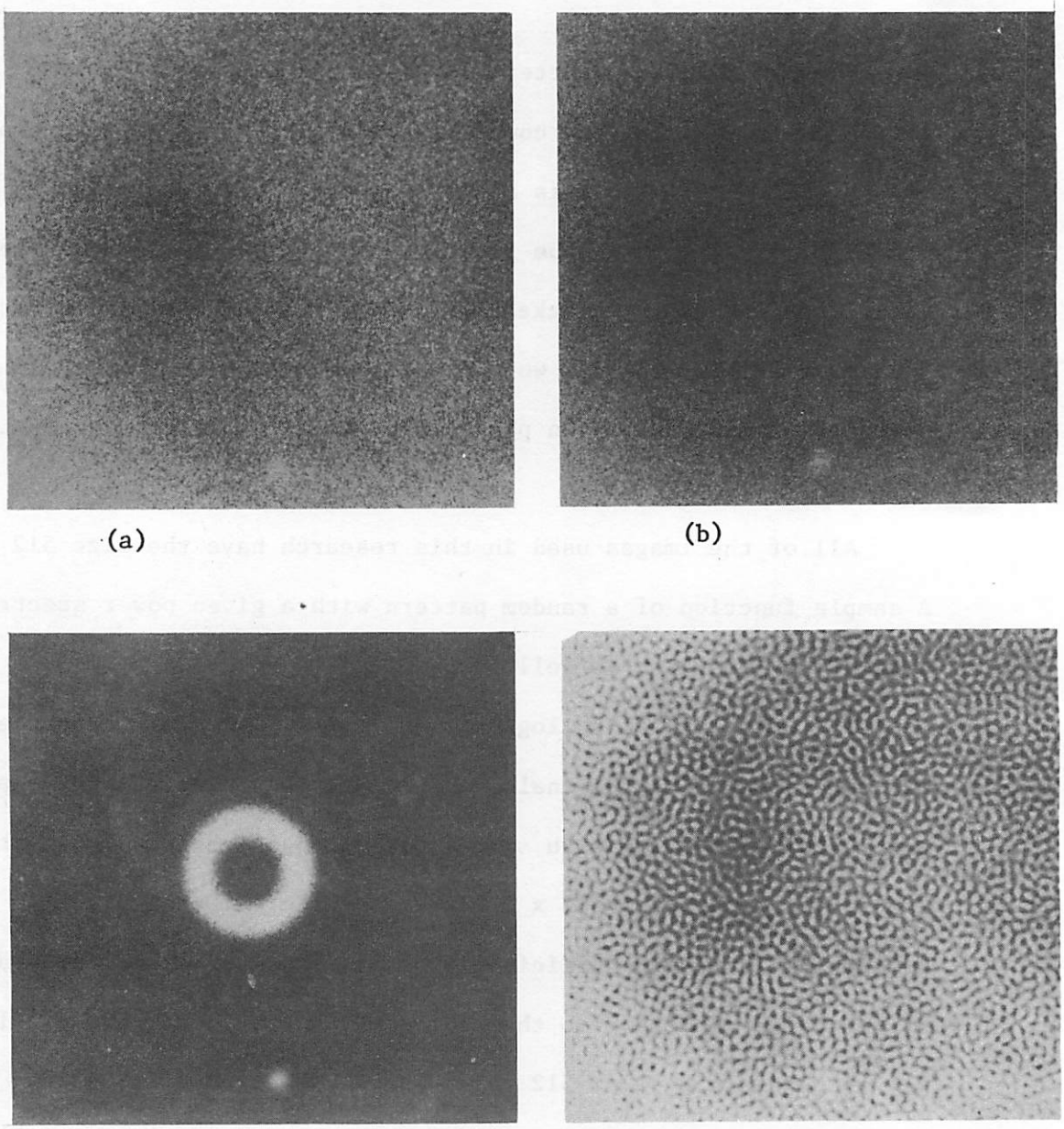


Fig. A1. (a) Sample function of a white noise pattern;
(b) Fourier transform of the white noise sample function;
(c) After multiplication by the frequency response of a narrow band isotropic filter
(d) Inverse Fourier transform of the image in (c).

$$S_N(f_x,f_y) = N_o|H(f_x,f_y)|^2 \qquad (A.1)$$

Therefore in order to generate a sample function of a random pattern with power spectral density $S_N(f_x,f_y)$ we filter the white noise sample function $n_w(x,y)$ with a filter whose frequency response satisfies

$$|H(f_x,f_y)| = [S_N(f_x,f_y)]^{1/2} \qquad (A.2)$$

The linear homogeneous filtering operation is performed on the computer using the two dimensional discrete Fourier transform. The fast Fourier transform (FFT) algorithm is used for computation of discrete finite transforms. Using this algorithm, the Fourier coefficients of the 512 x 512 white noise pattern are computed. Then they are multiplied by the corresponding frequency domain coefficients of the filter. The inverse Fourier transform of the resulting coefficients is a 512 x 512 point sample function of the random pattern with power spectral density given by the square of the magnitude of the frequency response of the filter (this ignores edge effects, which were negligible for the filters used).

The magnitude of the Fourier transform of the sample function of Fig. Ala is shown in Fig. Alb. Figure Alc shows the magnitude of the Fourier transform after filtering with a radially symmetric narrow band filter. The inverse transform which is a sample function of a band pass random pattern is shown in Fig. Ald. The magnitude of the frequency response of the filter used in this process is illustrated by the three dimensional plot of Fig. A2.

In both the generation of stimulus images and simulation studies, the filters had a real valued frequency response which

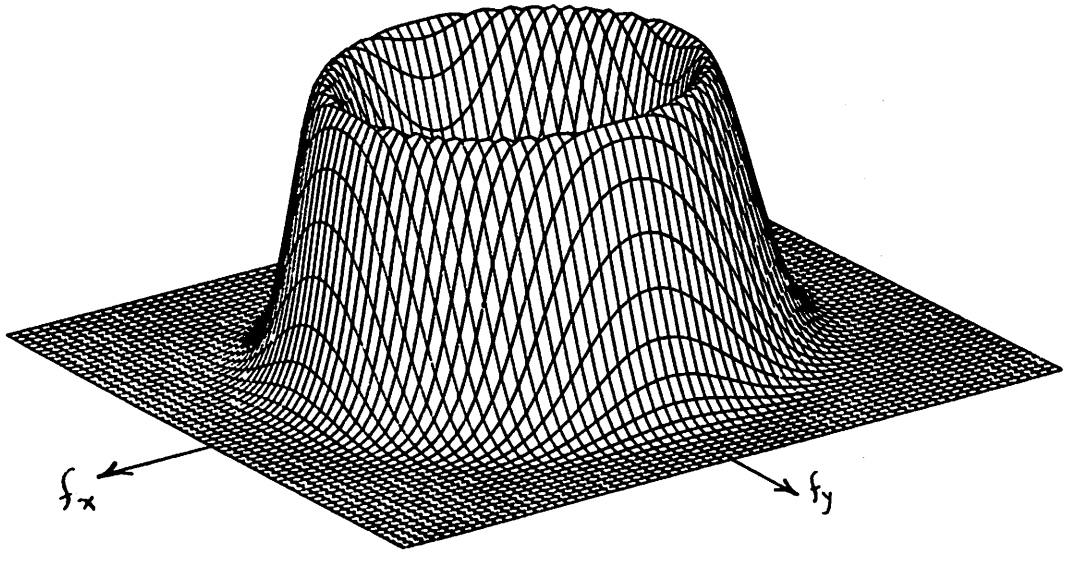


Fig. A2. Magnitude of the frequency response of a band pass isotropic filter.

corresponds to zero phase shift for all spatial frequencies.

The computations performed on picture files on the IBM 1800 computer use 16 bit integer arithmetic. In order to keep the percentage round-off error within an acceptable range, before each filtering operation the picture element values are scaled so that the maximum absolute value in the picture file is 32767 the largest positive decimal integer for 16 bits).

Before storing an image on the video disk for display on the TV monitor (see Appendix B), it has to be packed, which means that two adjacent picture elements are stored in a 16 bit word, each occupying 8 bits. Therefore in the process of packing, the PEL values are scaled so that they fall between 0 and 255. In the experiments of secs. 4.1 and 4.2 and all experiments of Chapter 5, the packed pseudo random files were such that the mean of the PEL values was 128 and the rms value was $\frac{128}{3}$. In the experiments of secs. 4.3, 4.4, 4.5 and 4.6 the mean was 128 and $\max_{x,y}|n(x,y)|$ had the value 127 for all pseudo random files n(x,y).

APPENDIX B

EQUIPMENT SET UP FOR THE EXPERIMENTS

The block diagram of Fig. B1 shows the equipment set up for the psychophysical experiments.

The visual patterns used in the experiments are generated as two dimensional arrays on the IBM 1800 computer system. With the use of the interface unit VIBE these digital pictures are converted into analog signals in standard TV format and are stored on the video disk. Each picture file is stored as a single monochromatic picture frame on any of the three channels of a track which is selected from the 105 tracks of the video disk.

VIBE accepts 8 bit picture elements with values in the range [0,255] as the input. The non-negative video signal voltage at the output of each channel is a linear function of the PEL values with value 0 generating 0 volts and value 255 generating 3 volts. For a specific setting of the CONTRAST and BRIGHTNESS controls on the TV monitor, which was maintained throughout the experiments, the luminance i on the face of the picture tube is related to the video input voltage and hence the PEL values in the computer by a monotonically increasing nonlinear function that we denote by g(.) This function was found by measuring the luminance for PEL values ranging from 0 to 255.

The inverse of this function $f = g^{-1}$ was made into a table mapping the values [0,255] onto [0,255]. In order to make the luminance at each point on TV screen a linear function of the corresponding PEL value in the computer, the background image values are transformed with this table before recording on the video disk.

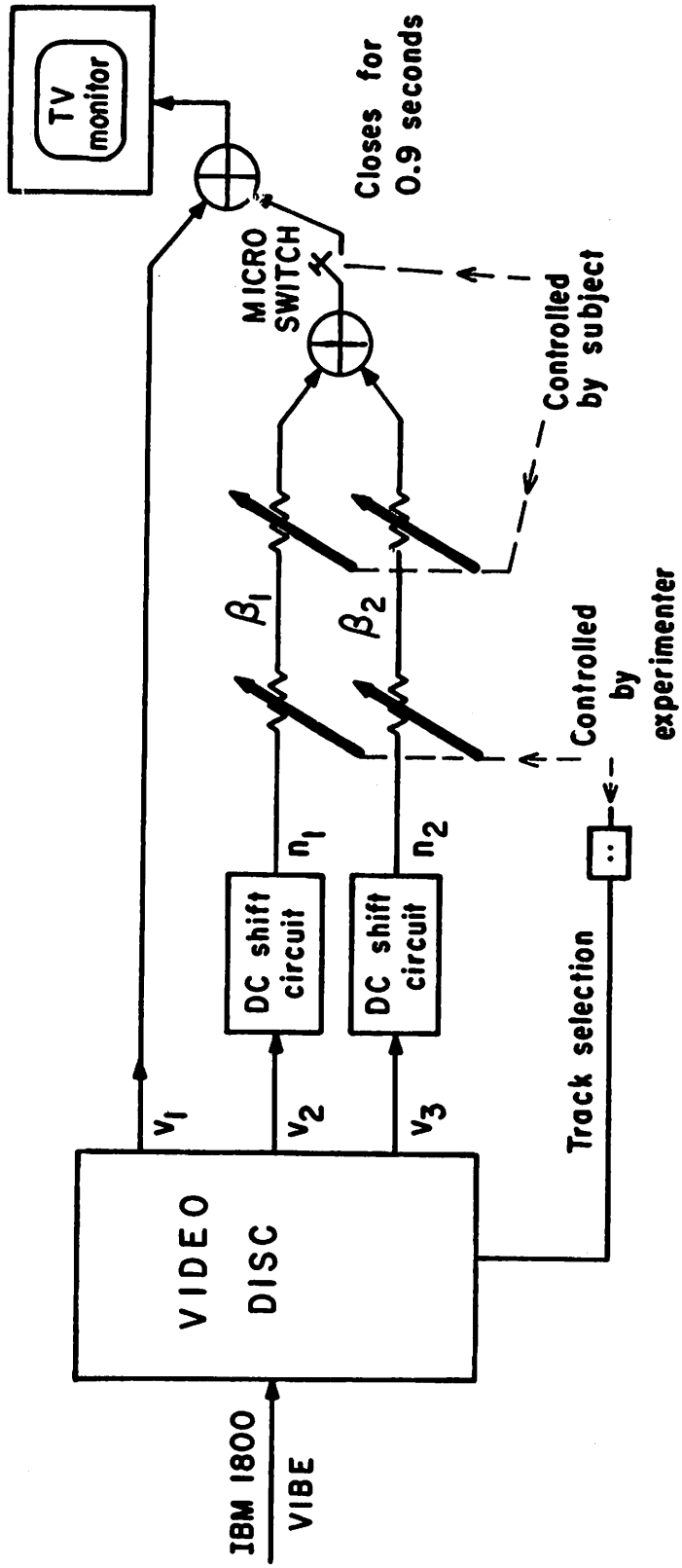


Fig. B1. Block diagram of the equipment set up.

The background image which is a non-negative pattern $i(x,y)$ is stored on channel 1 of each track. The perturbation patterns $n_1(x,y)$ and $n_2(x,y)$ are stored on channels 2 and 3 respectively. Since VIBE accepts only non-negative values in the range [0,255], these patterns are transformed in such a way that their mean value becomes 128 (see Appendix A) which corresponds to 1.5 volts in the output. However since the video signals corresponding to these patterns have to be zero mean, the output signal voltages of channel 2 and 3 are shifted down 1 volts by the circuits shown in Fig. B1.

In the experiments that use a constant luminance background, the luminance of the pattern on the TV screen has to be I_o = 860 cd/m^2 (=270 millilamberts) when the micro switch (Fig. B1) is open and

$$i(x,y) = I_o[1 + \alpha_1 n_1(x,y) + \alpha_2 n_2(x,y)]$$

$$= I_o + I_o\,\alpha_1\,n_1(x,y) + I_o\,\alpha_2\,n_2(x,y) \qquad (B.1)$$

when the micro switch is closed. In order to achieve this, a constant picture file with the value $f(I_o) = g^{-1}(I_o)$ was stored on channel 1 of a track. The pseudo random patterns $n_1(x,y)$ and $n_2(x,y)$ were stored on channels 2 and 3 of the same track. From the diagram of Fig. B1 we can see that the luminance is

$$i(x,y) = g(g^{-1}(I_o)) = I_o \qquad (B.2)$$

when the micro switch is open and

$$\hat{i}(x,y) = g[g^{-1}(I_o) + \beta_1 n_1(x,y) + \beta_2 n_2(x,y)] \qquad (B.3)$$

when the micro switch is closed. The constants β_1 and β_2 are the scalar factors corresponding to the attenuators of Fig. B1. For the contrast levels that are around the threshold of detection of the patterns $n_1(x,y)$ and $n_2(x,y)$, the values of β_1 and β_2 are such that the peak amplitude of $\beta_1 n_1(x,y) + \beta_2 n_2(x,y)$ is very small compared with $g^{-1}(I_o)$ and we have

$$\hat{i}(x,y) \approx g[g^{-1}(I_o)] + g'[g^{-1}(I_o)][\beta_1 n_1(x,y) + \beta_2 n_2(x,y)]$$

$$= I_o + C\beta_1 n_1(x,y) + C\beta_2 n_2(x,y) \qquad (B.4)$$

where

$$C = g'[g^{-1}(I_o)] \qquad (B.5)$$

Therefore for constant background experiments the attenuations α_1 and α_2 are related to β_1 and β_2 by

$$\alpha_j = \frac{C}{I_o}\beta_j , \qquad j = 1,2 \qquad (B.6)$$

In the experiments using a non-constant image i(x,y) as the background, the picture file $g^{-1}[i(x,y)] = f[i(x,y)]$ is stored on channel 1 of the track. The perturbed image must be

$$\hat{i}(x,y) = i(x,y)[1 + \alpha n(x,y)]$$

In order to have a luminance distribution given by this multiplicative form, the derivative of the function f, f', was made into a table and using this table the image

$$n(x,y) \cdot i(x,y) \cdot f'[i(x,y)]$$

was formed by the computer and it was stored on channel 2 of the track. In these experiments channel 3 was not used. Thus the luminance on TV screen becomes

$$g[f[i(x,y)]] = i(x,y) \tag{B.7}$$

when the micro switch is open and

$$\hat{i}(x,y) = g[f[i(x,y)] + \beta n(x,y)\ i(x,y)\ f'[i(x,y)]] \tag{B.8}$$

when the micro switch is closed. Since around the threshold of detection of perturbations the amplitude of $\beta n(x,y)\ i(x,y)\ f'[i(x,y)]$ is small compared to $f[i(x,y)]$ we have

$$\hat{i}(x,y) \approx g[f[i(x,y)]] + g'[f[i(x,y)]]\beta\ n(x,y)\ i(x,y)\ f'[i(x,y)] \tag{B.9}$$

However since $f = g^{-1}$ we have

$$g'[f[i(x,y)]] \cdot f'[i(x,y)] \equiv 1 \tag{B.10}$$

Therefore the luminance of the perturbed image becomes

$$\begin{aligned}\hat{i}(x,y) &\approx i(x,y) + i(x,y)\ \beta n(x,y) \\ &= i(x,y)[1 + \beta n(x,y)]\end{aligned} \tag{B.11}$$

which has the desired form.

12. Campbell, F.W., Kulikowsky, J. J. and Levinson, J., "The Effect of Orientation on the Visual Resolution of Gratings," J. Physiol., London, 187, 427-436.

13. Blakmore, C. and Campbell, F.W., "On the Existence of Neurones in the Human Visual System Selectively Sensitive to the Orientation and Size of Retinal Images," J. Physiol., London, 203, 237-260, 1969.

14. Sachs, M.B., Nachmias, J. and Robson, J.G., "Spatial frequency Channels in Human Vision," J. Opt. Soc. Am., 61, 1176-1186, 1971.

15. Campbell, F.W. and Kulikowsky, J.J., "Orientational Selectivity of the Human Visual System," J. Physiol., London, 187, 437-445, 1966.

16. Hubel, D.H. and Wiesel, T.N., "Receptive Fields of Single Neurones in the Cats' Striate Cortex," J. Physiol., London, 148, 574-591, 1959.

17. Robson, J.G., private communication, August 1972.

18. Halter, M., private communication, August 1974.

19. Papoulis, A., Probability, Random Variables and Stochastic Processes, McGraw-Hill, 1965.

20. Sakrison, D.J., Halter, M. and Mostafavi, H., "Properties of the Human Visual System as Related to the Encoding of Images," to appear in New Directions in Signal Processing in Communications and Control, NATO Advanced Studies Institute Series, Noordhoff Int. Pub. Co., 1975.

21. Kulikowsky, J.J. and King-Smith, E., "Spatial Arrangement of Line, Edge and Grating Detectors revealed by Subthreshold Summation," Vision Res., vol. 13, 1455-1478, Pergamon Press, 1973.

22. Kulikowsky, J.J., Abadi, R. and King-Smith, E., "Orientational Selectivity of Grating and Line Detectors in Human Vision," Vision Res., vol. 13, 1470-1486, Pergamon Press, 1973.

REFERENCES

1. Shannon, C.E., "Coding Theorems for a Discrete Source with a Fidelity Criterion," Information and Decision Processes, R.E. Machol, Ed., McGraw Hill, New York, 1960, pp. 93-126.
2. Sakrison, D.J., "Factors Involved in Applying Rate Distortion Theory to Image Transmission," Proceedings of the UMR-Mervin J. Kelly-Communications Conference, Rolla, Mo., 5-7, Oct. 1970.
3. Mitchell, O.R., Jr., "The Effect of Spatial Frequency on the Visibility of Unstructured Patterns," Ph.D. Thesis, M.I.T., Sept. 1972.
4. Arnulf, A. and Dupuy, Odette, "La transmission des contrastes par le systeme optique de L'oeil et les seuils des contrastes retiniens," C. R. Acad. Sci. Paris, 250, 2757, 1960.
5. Campbell, F.W. and Green, D.G., "Optical and Retinal Factors Affecting Visual Resolution," Journal of Physiology, vol. 181, pp. 567-592, 1965.
6. Polyak, S.L., The Veterbrate Visual System, University of Chicago Press, 1957.
7. Cornsweet, T.N., Visual Perception, New York: Academic Press, 1970.
8. Kuffler, S.W., "Discharge Patterns and Functional Organization of Mammalian Retina," J. Neurophysiol., 16, 37-68, 1953.
9. Brindley, G.S., Physiology of the Retina and the Visual Pathway, Edward Arnold, London, 1970.
10. Campbell, F.W. and Robson, J.G., "Applications of Fourier Analysis to the Visibility of Gratings." J. Physiol., London, 197, 551-566, 1968.
11. Van Nes, F.L. and Bouman, M.A., "Spatial Modulation Transfer in the Human Eye," J. Opt. Soc. Am., 57, 401, 1967.

23. Werblin, F.S., The Control of Sensitivity in the Retina," Scientific American, vol. 223, 70-79, Jan., 1973.