1977

A Modification of the Fourier Transform Model in Visual Perception

Bruce Apgar

William Harris

1

A MODIFICATION OF THE FOURIER TRANSFORM MODEL IN VISUAL PERCEPTION

by

Bruce K. Apgar and William D. Harris

A thesis submitted in partial fulfillment of the
requirements for the degree of Bachelors of Science in the
School of Photography in the College of Graphic Arts and Photography
of the Rochester Institute of Technology

May, 1977

Thesis Advisors: Dr. Edward Granger
Dr. G.W. Schumann

Acknowledgements

The authors of this thesis wish to express their gratitude to their advisor Dr. Edward M. Granger who was of invariable help during our course of study.

We would also like to thank Professor Mohammed F. Abouelata, Dr. G.W. Schumann, and Dr. E. Newburg.

The grant from the C.I.A. is gratefully acknowledged.

Abstract

As for photographic systems, the human visual system can also be described by an MTF. However, due to the complexity and non-linearity of the visual system, the MTF alone is a poor predictor of response. We have formulated and tested two mathematical models which modify the MTF of the eye. Effectively, an imperfect integrating function (low pass filter) has been proposed to be operating on stimuli which are sent to the brain. The models were tested and found to be better predictors of visual response than the present MTF alone.

Table of Contents

List of Illustrations

INTRODUCTION

The objective of this thesis is to create a mathematical model of how humans perceive visual objects which correlates better with experimental data than the present MTF model. One application of a new model would be in the diagnosis of cancer. In many types of cancer, the shade of the growth is important in determining if the growth is significant. If a photograph of the growth was necessary so that other doctors or patients could view it, the best reproduction would be one that was the same as the original scene. If a model of the human visual system was known, it would be possible to adjust the photographic system to compensate for the distorted perception caused by the visual system. By compensating for the distortion in the visual system, a true reproduction could be achieved.

We will investigate the MTF as a model of human perception, the theory of the MTF, what specifications must be met so that Fourier Transform techniques are valid, how the MTF is determined for humans, and where the MTF fails to predict what occurs in the visual system. We will propose two different models which work in conjunction with the MTF and perform experiments to determine the coefficients of the models.

THEORY OF THE MTF

The MTF is a way of specifying the quality of an imaging system. The MTF is based on Fourier's theorem which states that any function (satisfying certain mathematical restrictions) can be broken into a summation of sinusoidal functions of different frequencies and amplitudes. If it is known the system images sinusoidal functions at all physically realizable frequencies, then it is possible to predict how the system will reproduce an edge.

For example: if a sine wave is imaged by a lens,the lens will degrade the sine target.image. The amount that the sine target is degraded determines the MTF at that frequency. If the lens is tested over a sufficiently wide range of frequencies and the MTF determined, a curve can be plotted(Figure 1). If the MTF of a system is known by using Fourier analysis,it is possible to predict from the object luminance distribution what the image distribution will be.(assuming there is 0 phase-shift for all frequencies and that the transform of the system spread function is positive and even) For example: an object has a distribution as shown in figure 2. Using Fourier analysis, the transform of the object is taken(Figure 3). This is multiplied times the MTF of the system(Figure 4). The product of the MTF and the object luminance distribution is back transformed and the image distribution is obtained(Figure4a). Therefore, by knowing the MTF of a system and using Fourier analysis techniques, it is possible to predict the way that system will reproduce the object scene.[1]

SPECIFICATIONS FOR USE OF FOURIER TECHNIQUES

Fourier analysis can be used on a system provided the system meets certain conditions. The system must be linear. If a system is linear, when two functions are added in the spatial domain to give a third function, and the transforms can be added in the frequency domain to give the transform of the resultant spatial function.

$$h(x) = f(x) + g(x) \iff \mathcal{F}[h(x)] = \mathcal{F}[f(x) + g(x)]$$

also:

$$c\,h(x) \qquad \Rightarrow \mathcal{F}[c\,h(x)] = c\,\mathcal{F}[h(x)]$$

The system must be homogeneous. A homogeneous system is the same in all locations. When a system is homogeneous, changes in the location of the input pattern will cause corresponding changes in the output pattern. But the pattern will remain the same in all other respects. If a system is not homogeneous, the MTF of the system would be different for each different location. The system must be isotropic. An isotropic system has the same characteristics in all directions. An example of a non-isotropic system is a lens with astigmatism. In this situation, a target in one direction is imaged better than the same target in another direction.

To use one-dimensional Fourier analysis on the human system, the conditions of linearity, homogeneity, and isotropy must be met. The human visual system overall is not a linear system. Therefore, Fourier analysis on the system as a whole would be incorrect. But it is possible to examine small parts of the system over small brightness ranges and obtain results which are usefull. A proposed curve for brightness vs. scene luminance is shown in figure 5. This graph shows the visual system is not linear in luminance but is nearly a logarithmic. If we assume that the human visual system is linear with the exception of the brightness response, it is possible to use Fourier techniques on the system. This could be done by using an input which would compensate for the logarithmic response in the early stages of the system.. An example of this is an antilog sine wave grating. This

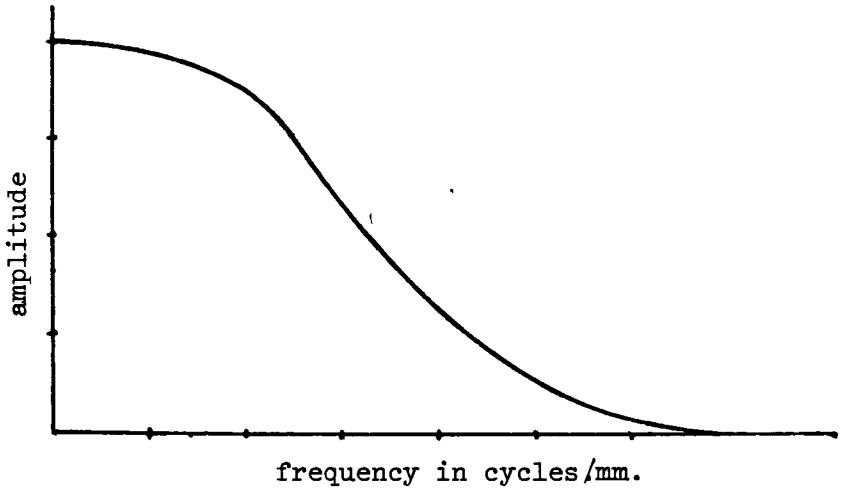


Figure 1 (hypothetical MTF)

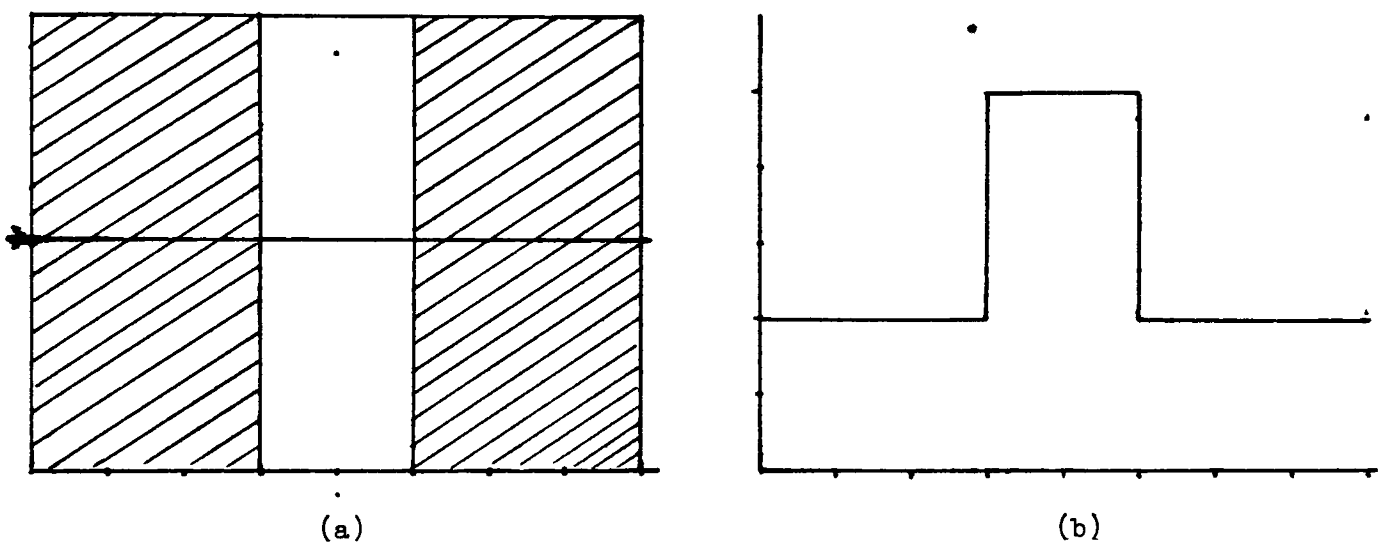


Figure 2

(A two-dimensional image(a) scanned by a micro-densitometer along the red line, the resulting densities transformed into brightness vs. distance(b))

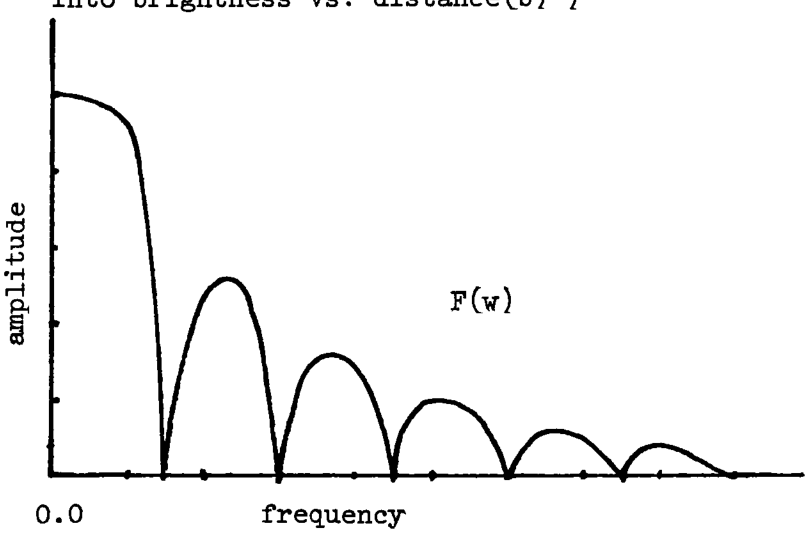


Figure 3

(Modulus of transform of rectangular pulse, positive frequencies plotted)

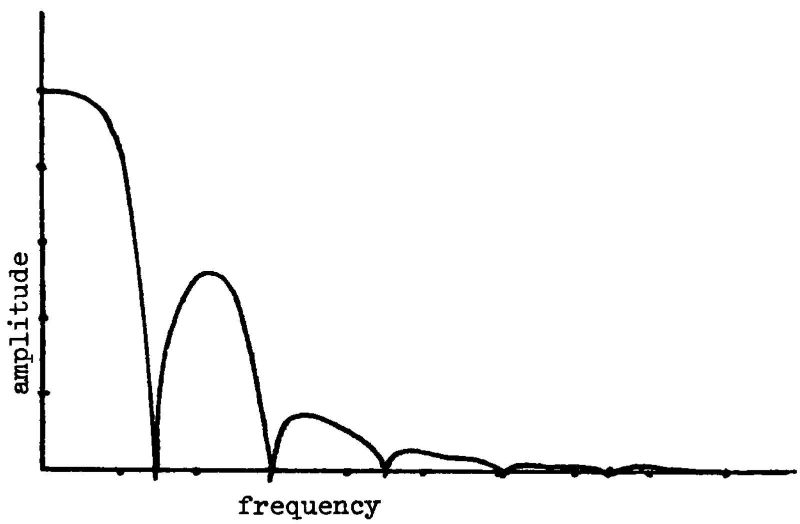


Figure 4

(Modulus of the transform of a pulse multiplied by an MTF, positive frequencies plotted)

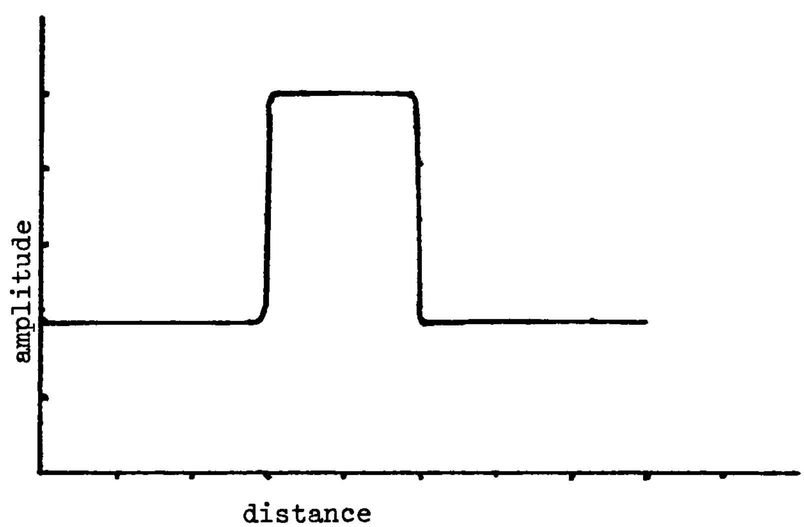


Figure 4a

(Inverse transform of MTF(w)·F(w))

grating would become sinusoidal after going through the first part of the system and Fourier techniques would yield reasonable results. An examination of figure 5 shows that over a small range of intensities, the system is nearly linear. Therefore, if a small intensity range is used, the system is nearly linear and Fourier techniques can be used. Homogeneity and isotropy must be considered when using Fourier analysis but the information available thus far indicates that Fourier analysis can be used on the eye in spite of non-homogeneity and non-isotropy.[2]

DETERMINATION OF THE MTF FOR THE HUMAN VISUAL SYSTEM

The threshold measurement technique is one method which has been used to calculate the MTF of the eye. In this method, sine wave targets are shown to a subject and the intensity variation is adjusted until he can just distinguish the target from a uniform field. The procedure is repeated for a number of different frequencies to obtain a plot. This method gives the MTF of a system it the system is linear and homogeneous. If the amplitudes of the sine waves are quite small, the change in intensity is small and the system is linear.[3]

FAILURE OF MTF

Although the MTF is used as a model for the human visual system , there are several things which the MTF does not show that occur in real life. One of the things that the MTF does not predict correctly is the Mach effect. An example is given in figure 6. The two targets have different luminance distributions but the MTF calculates they will be perceived as the same(Figure 7). Visually, the two targets are seen to be different, therefore the model fails.

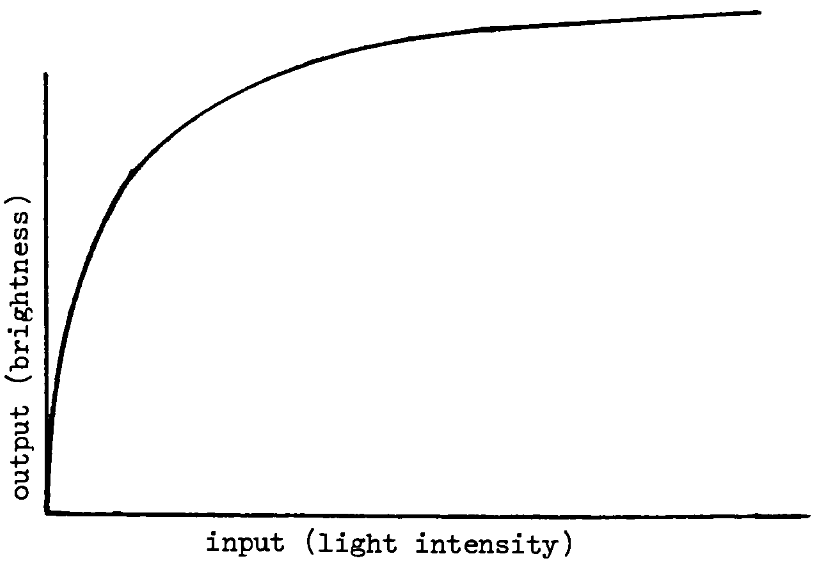


Figure 5

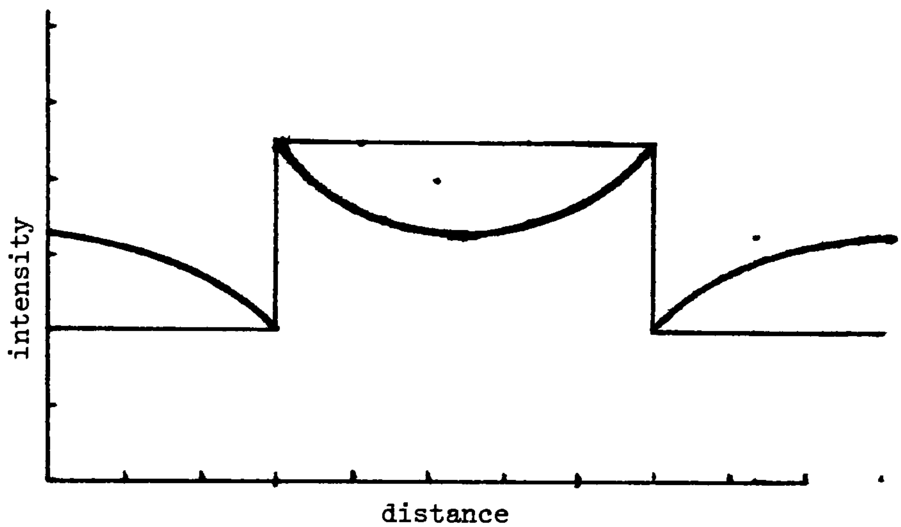


Figure 6

(two different luminance distributions)

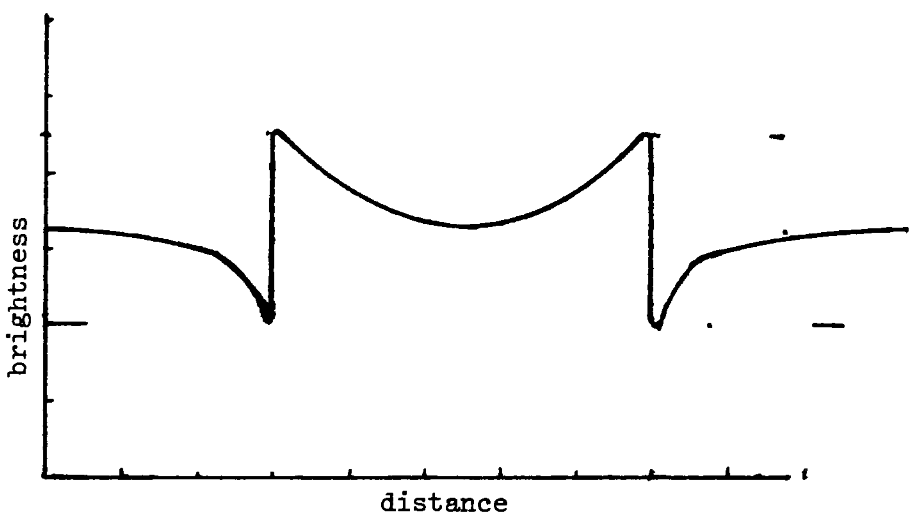


Figure 7

(predicted response for both distributions using the MTF)

THEORY

The organization and function of the nervous systems of higher order animals such as the cat as well as the human being, have recently been studied. It is known that there are at least nine neuron types and signals in the visual processing center of the cat brain.[4] They range from constant 'on' neurons to delayed action neurons which deliver an 'on' signal after illumination of the visual receptors is turned off. The resulting perception is thus some form of summation of a complicated combination of signals.[5] In the human, many more than nine types of neurons probably exist, and the summation is then further processed by other parts of the brain. These other parts might represent learned responses. In attempting to understand and model human vision, the simplest formulas which are looked for in other branches of science are not necessarily the path to follow with visual networks. An interesting quote to this effect is:

> "The principle of sloppy workmanship states that it is dangerous to postulate a neural structure that is precisely arranged in detail. The nervous system is the product of a superb architect and a sloppy workman, and in his plans the architect took into account the fact that the workman would not get all the terminal boutons where they belonged.....
>
> The principle of diversity states that the nervous system often hedges. Instead of presenting a single transform of the peripheral stimulation to the higher centers, the ...(afferent) tract may present a number of transforms."[6]

A phenomenon of the nervous system which is common to tactile sensations in fingers, hearing, and the visual system is that a sharp stimulus is accentuated by increasing its perceived intensity at its incidence accompanied by negating the perceived intensity about that place of incidence. This phenomena as noticed in the visual system was first recognized and studied by Ernst Mach in 1865. These were labeled "Mach Bands" in recognition of his work.

In proposing models, many previous researchers have hypothesized that the phenomena of accentuating differences in stimuli was due to two separate operations: excitation and inhibition. Cornsweet as well as Ratliff have mentioned models of neural networks located directly below the receptor layer in the retina, which predict an MTF similar to that which has been experimentally found. Basically, the models suggest that there are inhibitory synapses which interconnect nerve cell bodies and that the strength of the inhibitory effect on cell(a) is dependent on the strength of the incident signal on cell(b) as well as the distance it is from(b)(approximately inverse square).[7,8] In addition, due to the spreading of incident illumination by the eye optics as well as at the nerve cell level in the retina, a smearing effect may give rise to an MTF similar to MTF's of photographic systems. But due to the complexity of the visual system, a spreading of the excitatory signal may also occur in the brain's processor. We are choosing a model which says that this is what occurs and we are calling it an 'integrator.' Effectively, it represents the way the brain operates on signals from the retina. Thus, it operates on the MTF affected signals(if the models of Cornsweet and Ratliff are assumed correct). It has the function of compensating for the low frequency response of the standard MTF. It is important, however, to note that our model does not represent a perfect integrator. Rather, it accounts for the relative low-frequency processing which the visual system is experimentally seen to have. But it is imperfect in that extremely low frequency information is lost. This is consistent with experimental findings.

One interesting mathematical operation has previously been suggested. It is that the eye-brain processor takes a 2nd order derivative and adds this to some other response to give the final visual response of the eye. Thus, the Mach bands as seen in experiments may be the result of a pseudo-differentiator in the system. If even functions are assumed, this model predicts MTF's similar to those which have been determined by experiment.

It is important to note that the MTF does not give any phase-shift information and that current methods cannot determine it. Under the assumption of isotropy and homogeneity, the resultant spread function is assumed even and real. Also, to accomodate the obvious non-linearity of brightness vs. luminance, a logarithmic transformation of luminance to brightness is made. It is then assumed that the system is linear with respect to brightness. Some models would not be linear, even at this point, and thus, transform analysis would be invalid for them.

For a review of convolution and MTF, see Appendix .

An illustration of the application of a hypothetical excitatory-inhibitory spread function is shown below. By application of the convolution theorem, the same operations can be done in the frequency domain as shown.

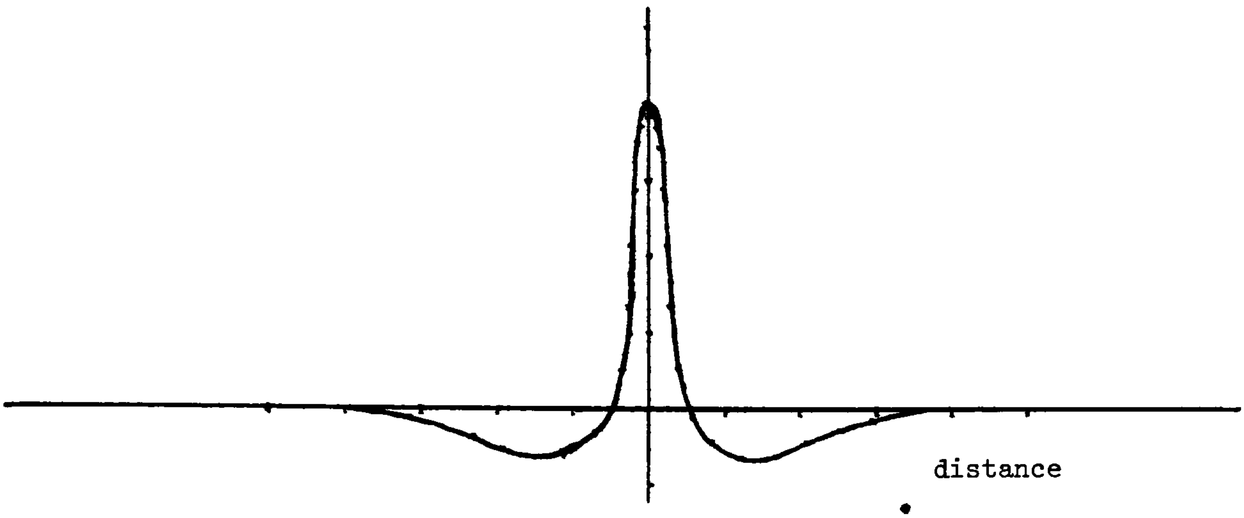


Figure 8

(hypothetical spread function with inhibition)

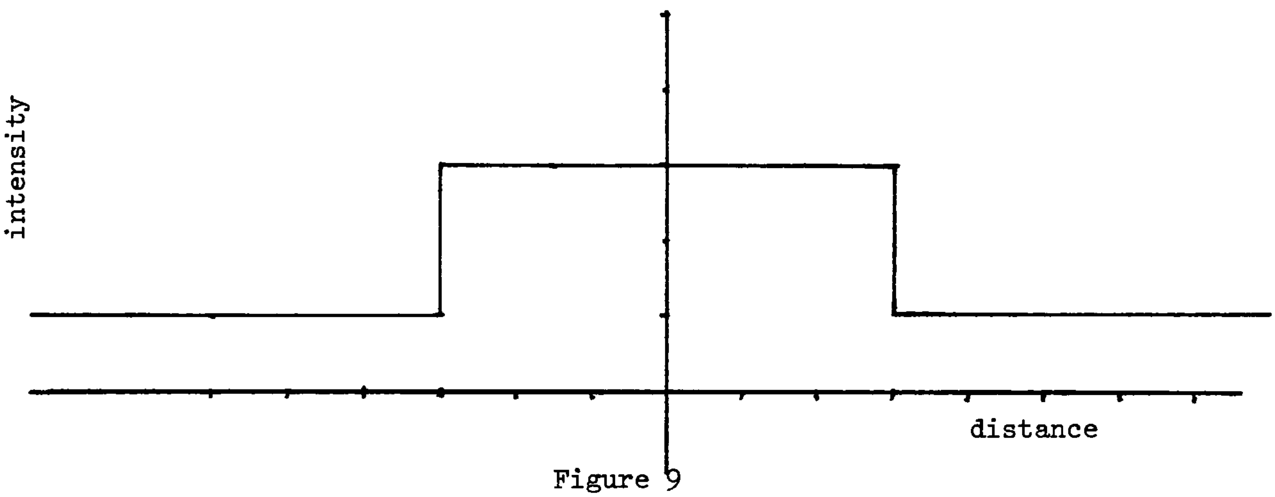


Figure 9

(rectangular pulse stimulus distribution)

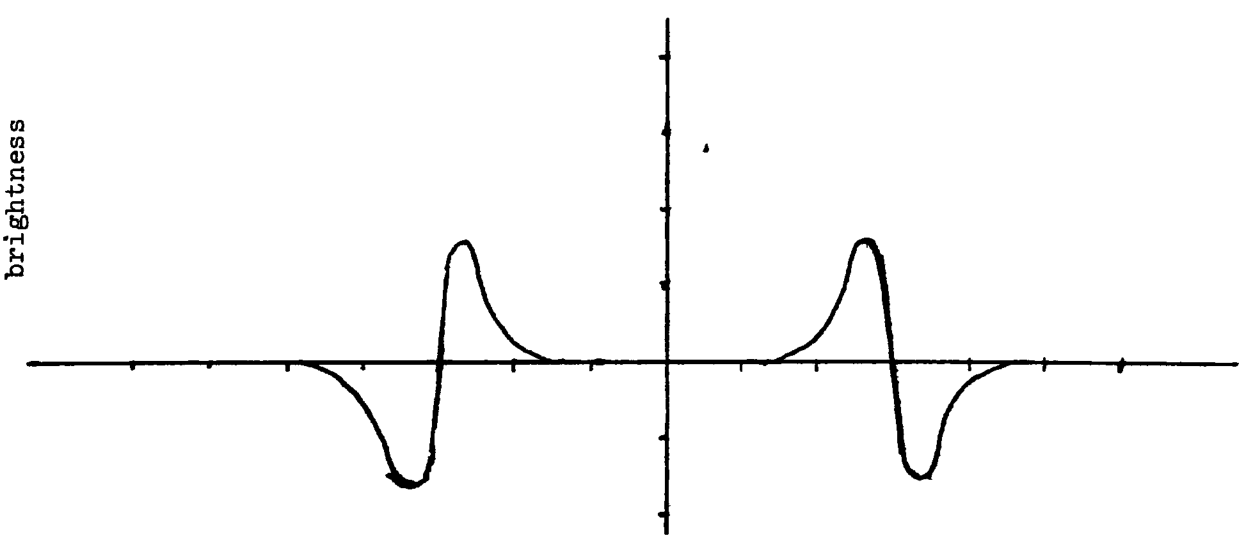


Figure 10 (predicted response-convolution)

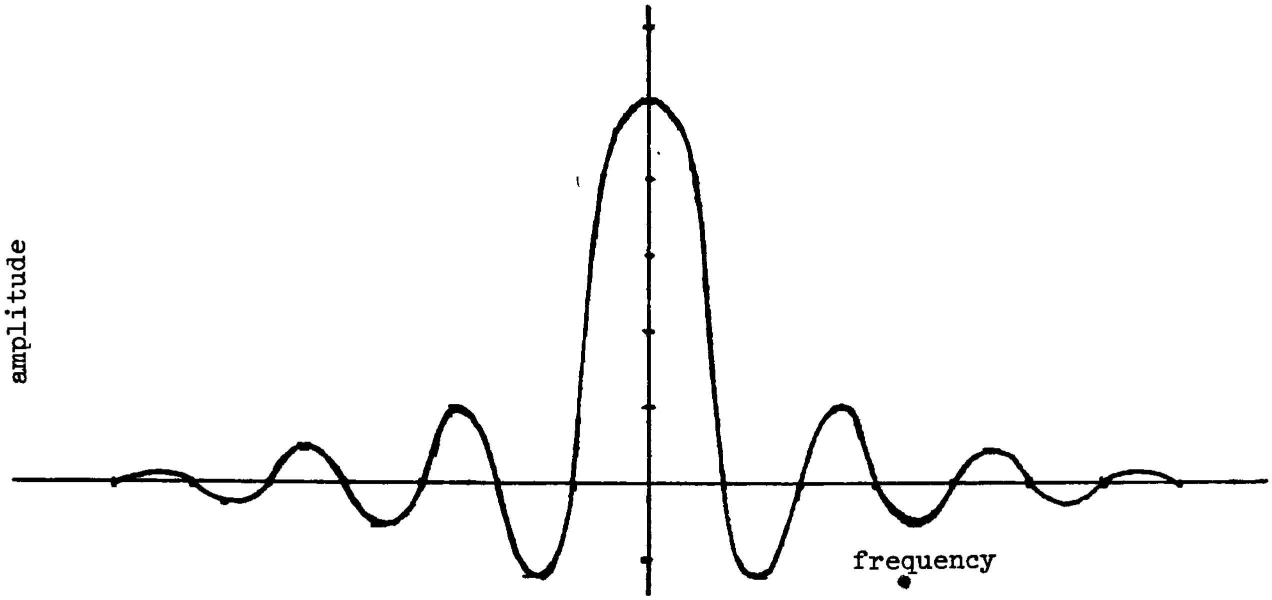


Figure 11 (Fourier transform of figure 9)

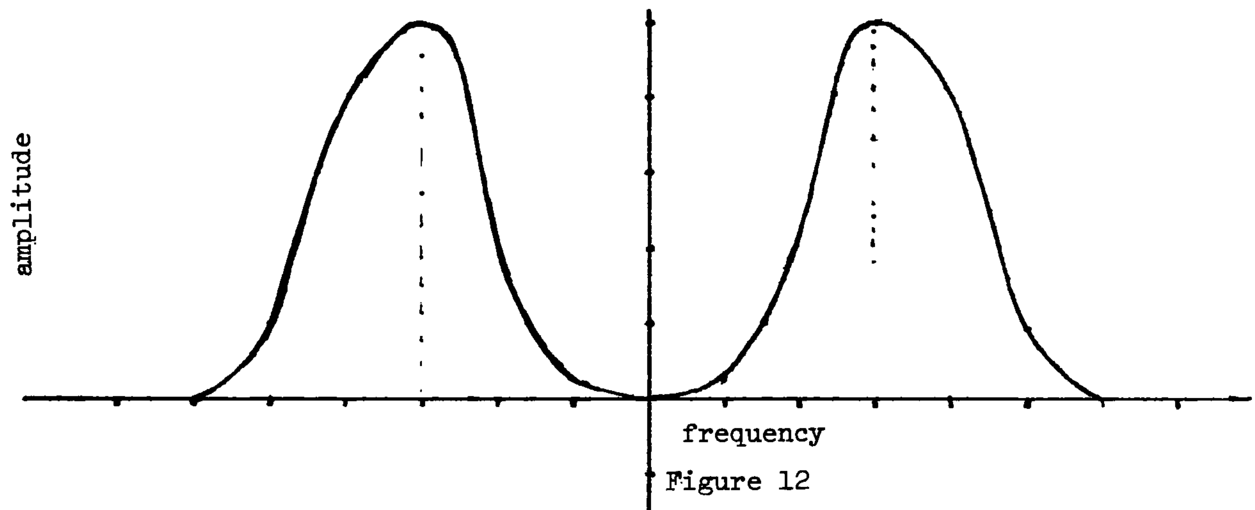


Figure 12
(transform of 8)

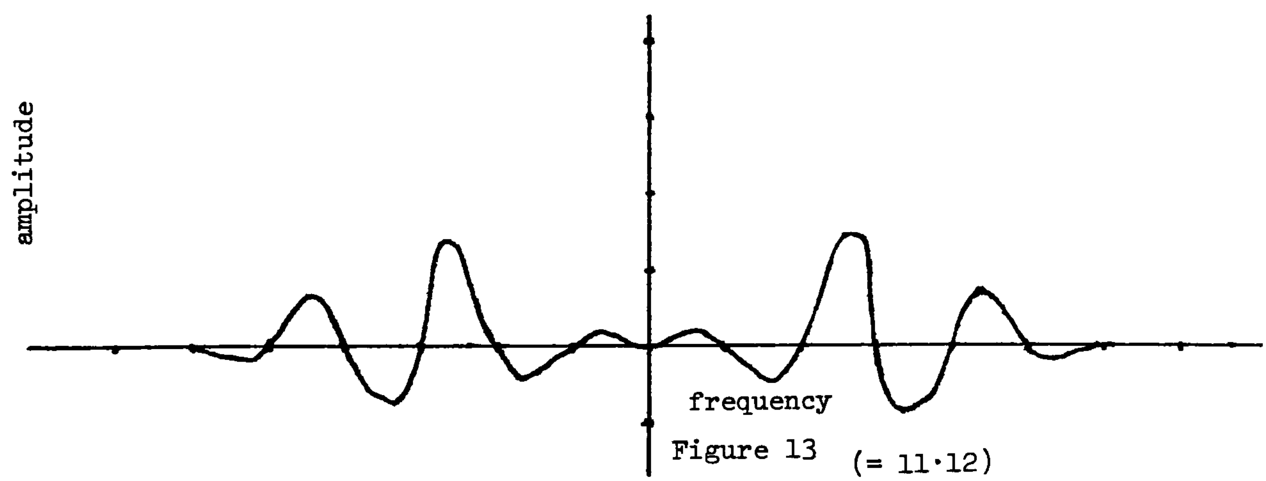


Figure 13 (= 11·12)

The mathematical models being proposed are:

1rst model $\overline{r(x)} = C_a + i(x) * (s(x) * f(x))$

where: $\overline{r(x)}$ = predicted spatial brightness response(degrees)

C_a = constant which changes the overall average value

$i(x)$ = spatial integrator

$s(x)$ = spatial function corresponding to the MTF of the retina

$f(x)$ = spatial brightness distribution(calculated from microdensitometer trace)

$*$ stands for convolution

2nd model $\overline{r(x)} = i(x) * (s(x) * f(x)) + C_b\, i(x) * f(x)$

where: C_b = coefficient for integrator part only

In Fourier transform notation, these two models can be written as:

1rst model $\overline{R(w)} = I(w)MTF(w)F(w)$

where: $\overline{R(w)}$ = predicted response transform in(cycles/degree)

$I(w)$ = integrator model in frequency space

$MTF(w)$ = Modulation transfer function of human retina from Cornsweet

$F(w)$ = Fourier transform of $f(x)$

2nd model $\overline{R(w)} = I(w)(MTF(w) + C_b)F(w)$

Model 1 is the simplest of the two. Since Cornsweet's MTF is 0.0 at 0 frequency, the area under the predicted response is 0.0. Constant C_a is needed to raise the predicted values to their proper level. The integrator $I(w)$ and C_a must be solved for.

The second model says that the integrator works by itself, besides operating on the MTF modified transform. Using the same $I(w)$ that was found for model 1, constant C_b must be solved for. The value of C_b indicates the contribution to the total response, that the integrator function makes by itself.

APPARATUS AND APPROACH

The first experiment performed was to determine a relationship between relative brightness and luminance. In the experiment, the subjects viewed three patches simultaneously on a 16"x20" light box. The patches were 1.5"x1.5". A dark patch and a bright patch were kept constant while the middle patch was varied. The bright patch was clear and the subjects were told it had a brightness value of 100. The dark patch had a diffuse density of 2.44 and the subject was told its relative brightness was 0. The subjects compared patches of different densities to the standard patches. A percent brightness was obtained for each patch. 32 subjects were used, each compared 19 patches, the results were averaged and relative brightness vs. diffuse density curve plotted. The curve was essentially linear with a correlation better than 96 % for the linear model.

Next, came the making of test targets for our models. The method used to create the targets was to photograph spinning discs. The discs used were 72 cm. in diameter. To create the desired target, the disc was painted with white and black. The amount of area of white vs. black on a target determined what the luminance distribution would be. The disc was spun with a small electric motor. Exposure times of 1.5 minutes were used. The spinning of the disc and long exposure times had the effect of integrating the light and causing greys. The disc was photographed with Kodak Plus-X-Pan 8x10 film. The film was processed in Kodak D-76 for 9 min. at 20°C.

The design of the testing apparatus is shown in figure 14. The brightness reference scale consisted of 14 patches equally incremented in diffuse density. Due to the linear relationship found between density and brightness, the scale was effectively linear in brightness also.

PROCEDURE FOR TESTING

The subject to be tested is asked to sit so that his eyes are 40 cm. away from the target. First, the subject is given a test for consistency. The subject is asked to view a group of 12 patches of densities similar to the patches on the grey scale and give the relative brightnesses. The person is shown the first target, is told where the actual edge is and is told what to look for. The subject is then given a piece of graph paper and asked to draw the brightness distribution of the target using the patches on the right as a reference scale. The procedure is repeated for the 2nd target. The subject is then shown a Mondrian like image and asked to give the relative brightnesses of various patches. This test is designed to test the interaction between the patches. Lack of time, has prevented us from thoroughly analyzing the data. The subject is then shown the final target. (note: It was suggested that people look at the target for a maximum of 3 seconds and a minimum of about .5 seconds.)

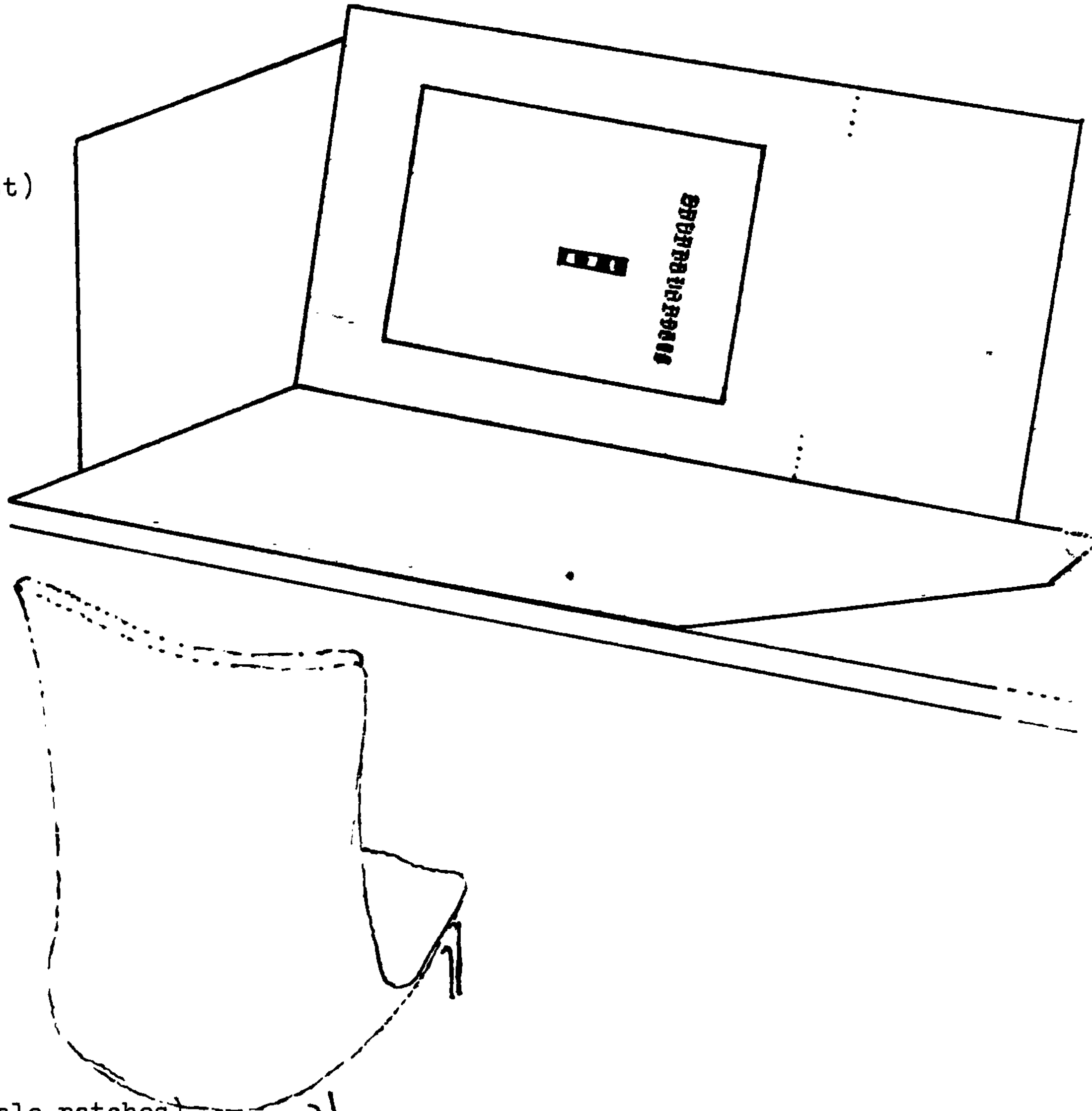
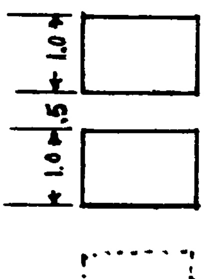


Figure 14

(diagram of test) apparatus

Figure 15



(reference scale patches)

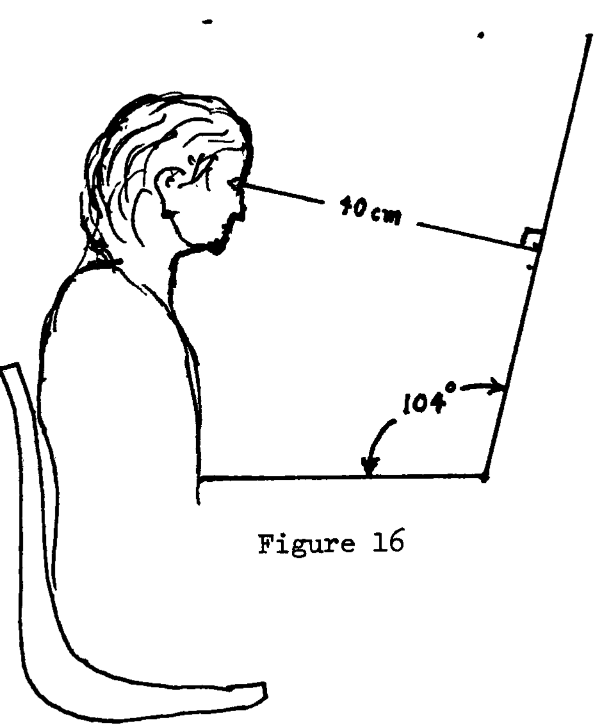


Figure 16

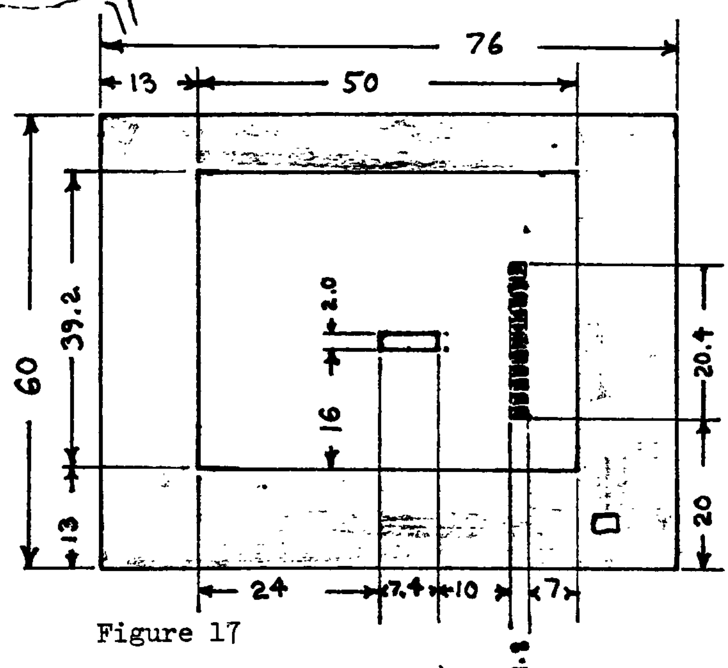


Figure 17

(dimensions of apparatus) in centimeters

ANALYSIS OF DATA AND COMPUTER PROGRAM

Since the particular Fast Fourier Transform that we used required data arrays of 2^n (n integer) in size, the array size was set at 256 points. The target width was 74.1765 mm. The viewing distance was 400 mm. and the focal length of the eye was assumed to be 17 mm.. Since there are 255 spaces between 256 sample points, the various sampling intervals and frequency ranges can be calculated as follows.

$$dx = \frac{74.1765 \text{ mm.}}{255} = .29089 \text{ mm. (sampling interval in target plane)}$$

$$\frac{dx}{dx'} = \frac{\text{increment on target}}{\text{increment in retina}} = \frac{\text{distance to target}}{\text{focal length of retina}}$$

$$\frac{.29089 \text{ mm.}}{dx'} = \frac{400 \text{ mm.}}{17 \text{ mm.}}$$

$$\therefore \quad dx' = \frac{.29089(17)}{400} \text{ mm.} = .01236 \text{ mm.}$$

$$d\theta = \tan^{-1}\left(\frac{.01236 \text{ mm.}}{17 \text{ mm.}}\right) = .041666666^\circ \text{ (angular increment in retina)}$$

The relationship between spatial sampling interval($d\theta$), the frequency sampling interval(dw), and the number of array elements is:

$$d\theta \, dw = \frac{1}{N}$$

$$dw = \frac{1}{d\theta \, N} = \frac{1}{(.04167)(256)} = .09375 \text{ cycles/degree}$$

The transform thus has a window from w = -12.000 to w = 11.906 .

In order to process the data with the ability to try many different mathematical models while at a teletype console, a FortranIV program was written. The program (NUMALG., written by William D. Harris) had the capability of spline fitting data, inputting and outputting to permanent files, Fast Fourier Transforming arrays up to 4096 in size, multiplying, dividing, adding, subtracting, editing, and plotting of these arrays. Thus, each array represented a mathematical function sampled at 256 points(all arrays were 256 elements in size) and could be treated as algebraic variables with the restriction that when division by zeroes occurred, or division by very small numbers, run-time errors or extreme calculation errors were incurred.

METHOD OF SOLVING FOR I(w), C_a, and C_b

The two models in the frequency domain are:

$$\overline{R(w)} = I(w)MTF(w)F(w)$$

and

$$\overline{R(w)} = I(w)F(w)(MTF(w)+C_b)$$

The first model was used to solve for I(w) for both models. Substituting the transform of the average test response,for each target, for $\overline{R(w)}$, I(w) is ideally found by:

$$I(w) = \frac{R(w)}{MTF(w)F(w)}$$

However, since F(w) might be 0.0 for some values of frequency, the following formula was used:

$$I(w) = \frac{R(w)F^*(w)}{(|F(w)|^2 + .003)} \cdot \frac{1}{MTF(w)}$$

note: the 0.0 frequency value for the MTF was replaced by 1.00 to prevent run-time errors. After the division, the 0.0 frequency value for I(w) was set equal to 0.0 .

The three I(w) s were then averaged and a best visual curve fitted to it. At the peak of the smooth integrator, about the 0.0 frequency, the integrator was flattened out at this maximum value. Thus a smooth, real and even function was created. The data and the resultant smooth integrator are plotted in figure 18 for positive frequencies.

In order to solve for the constants C_a and C_b, all functions were inverse transformed to the spatial domain. Using the following formulas, the constants were solved for each target, and the average values taken.

$$C_a = \left[\overline{r(x) - \mathcal{F}^{-1}\{I(w)MTF(w)F(w)\}}\right] = \left[\overline{r(x) - \dot{r}(x)}\right]$$

$$C_b = \left[\overline{\frac{r(x) - \mathcal{F}^{-1}\{I(w)MTF(w)F(w)\}}{\mathcal{F}^{-1}[I(w)F(w)]}}\right] = \left[\overline{\frac{r(x) - \dot{r}(x)}{\mathcal{F}^{-1}[I(w)F(w)]}}\right]$$

(the large bars over the expressions signify that an average value was) taken over all 256 points of the ratio

Figure 18

Target	C_a	C_b
A-2	26.22142	.051164
A-3	51.729675	.110497
A-5	61.903397	.162080
average value	46.61816	.10791
standard deviation	18.38196	.05550

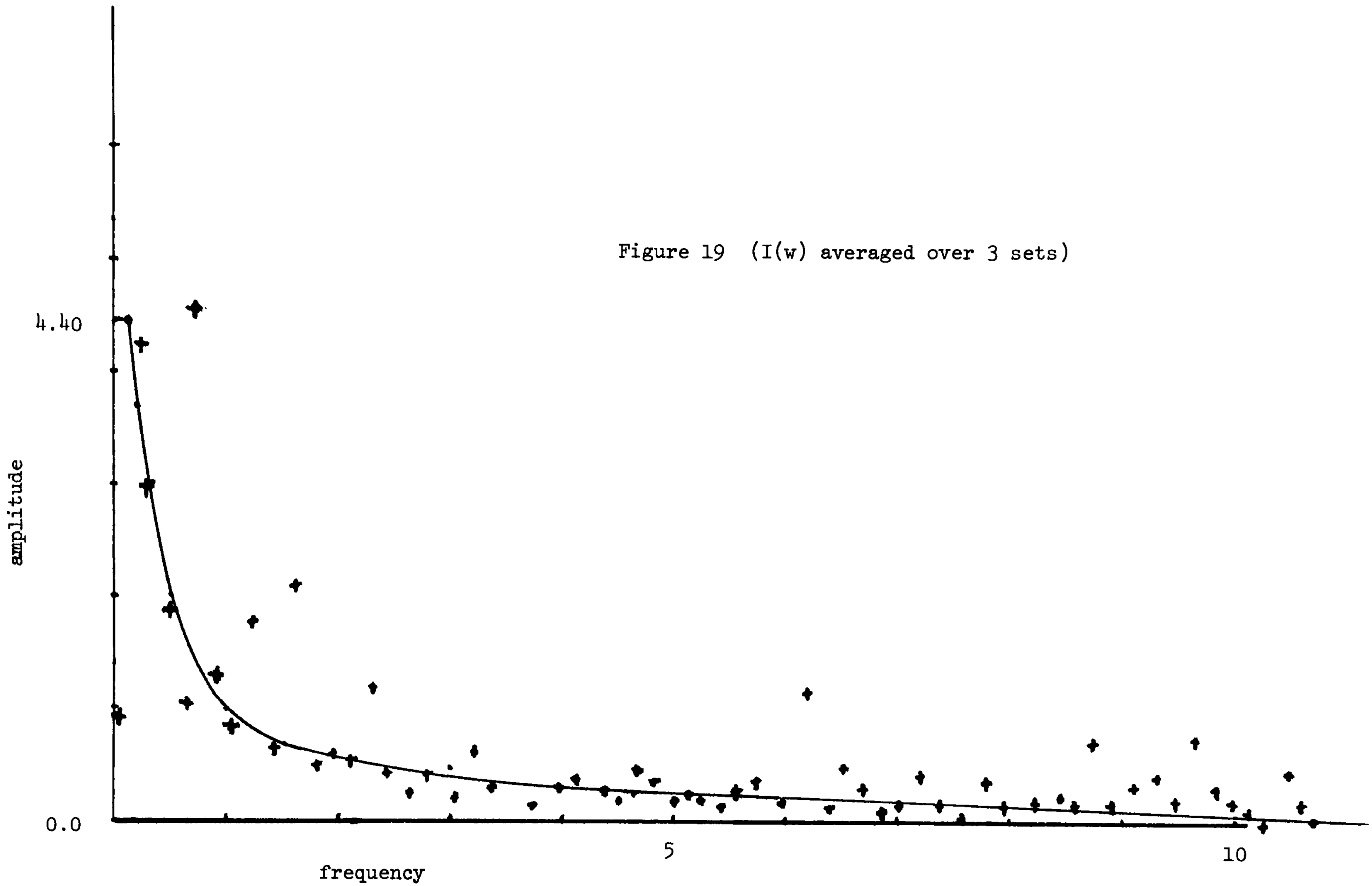


Figure 19 (I(w) averaged over 3 sets)

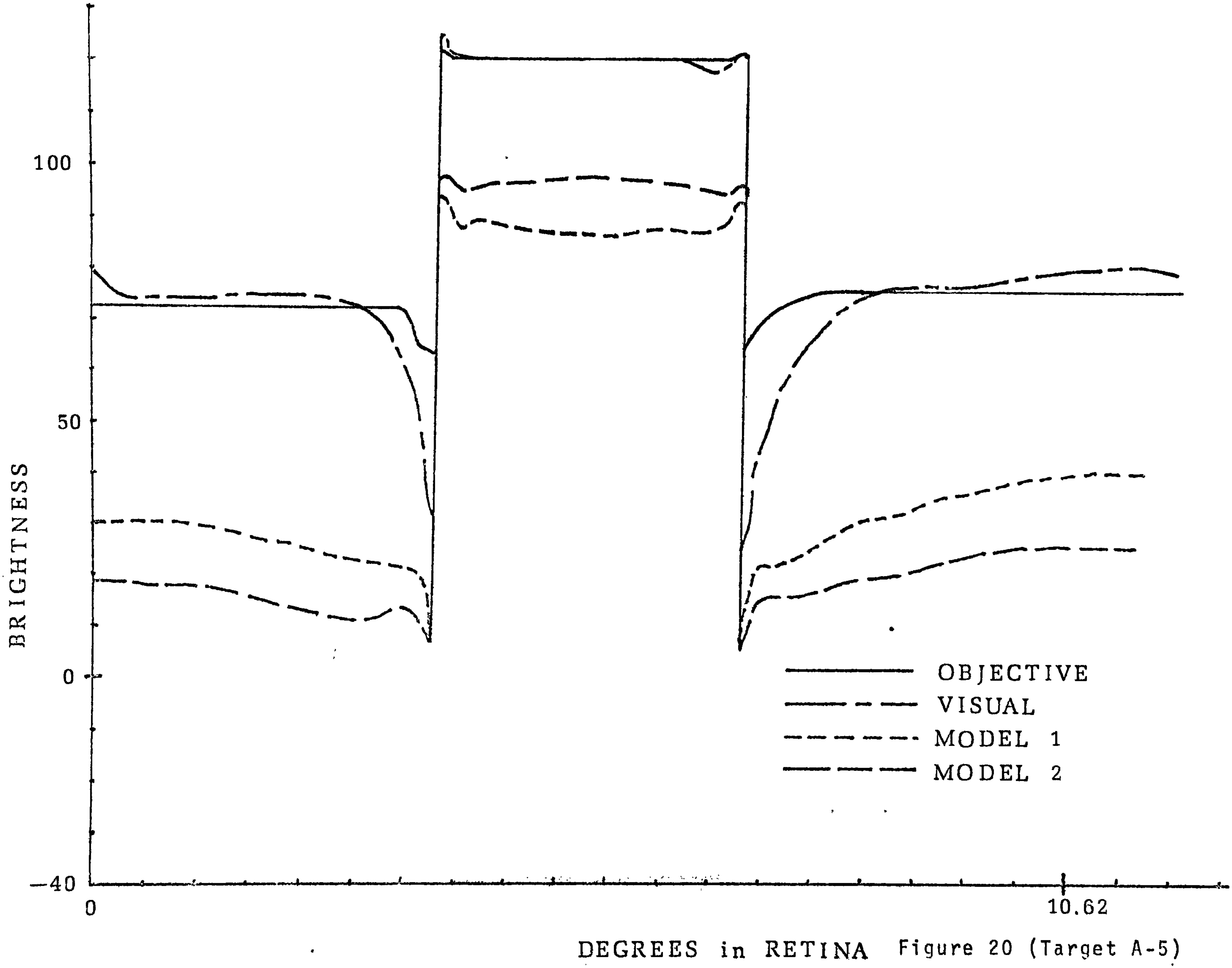


Figure 20 (Target A-5)

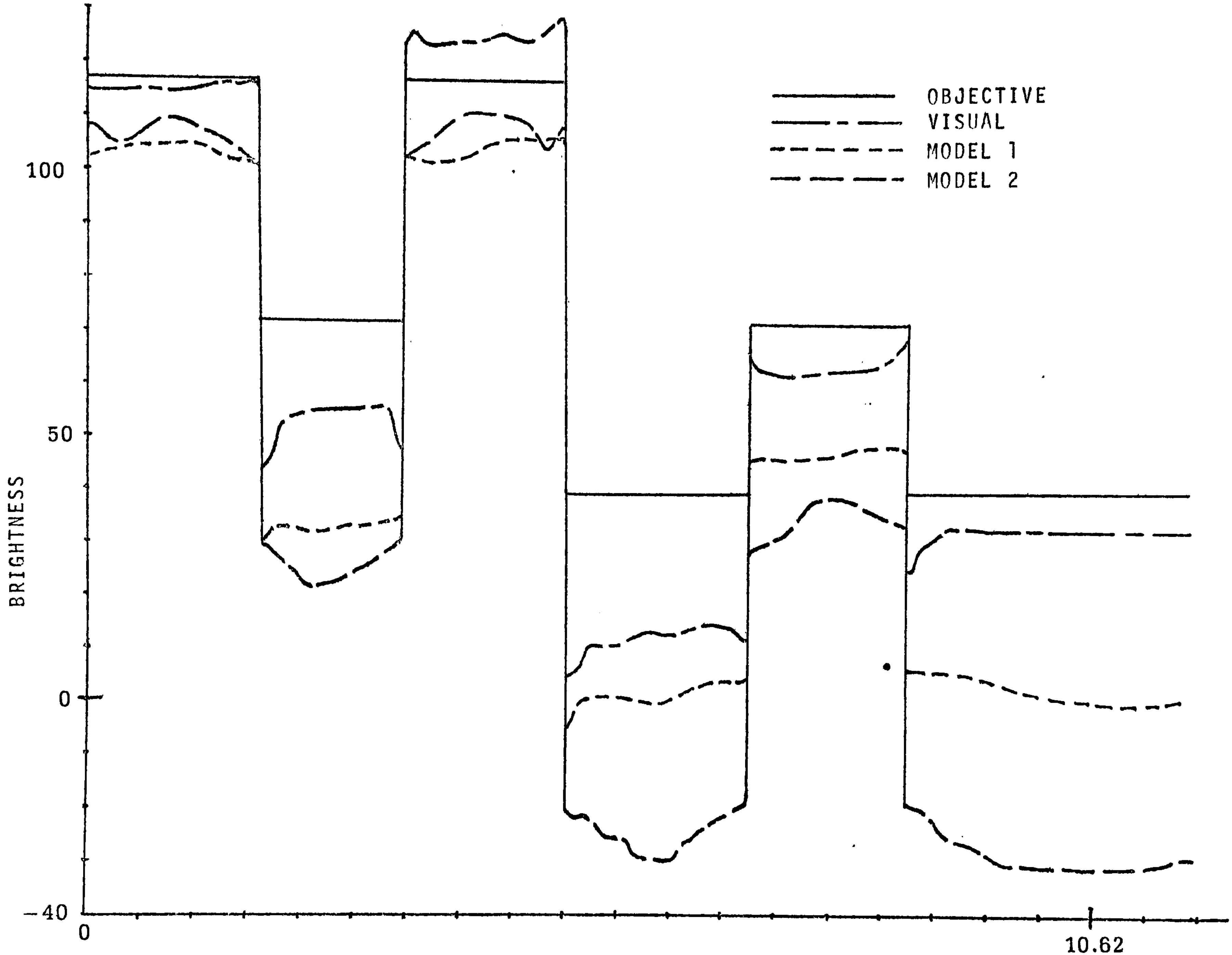


Figure 21 (Target A-2)

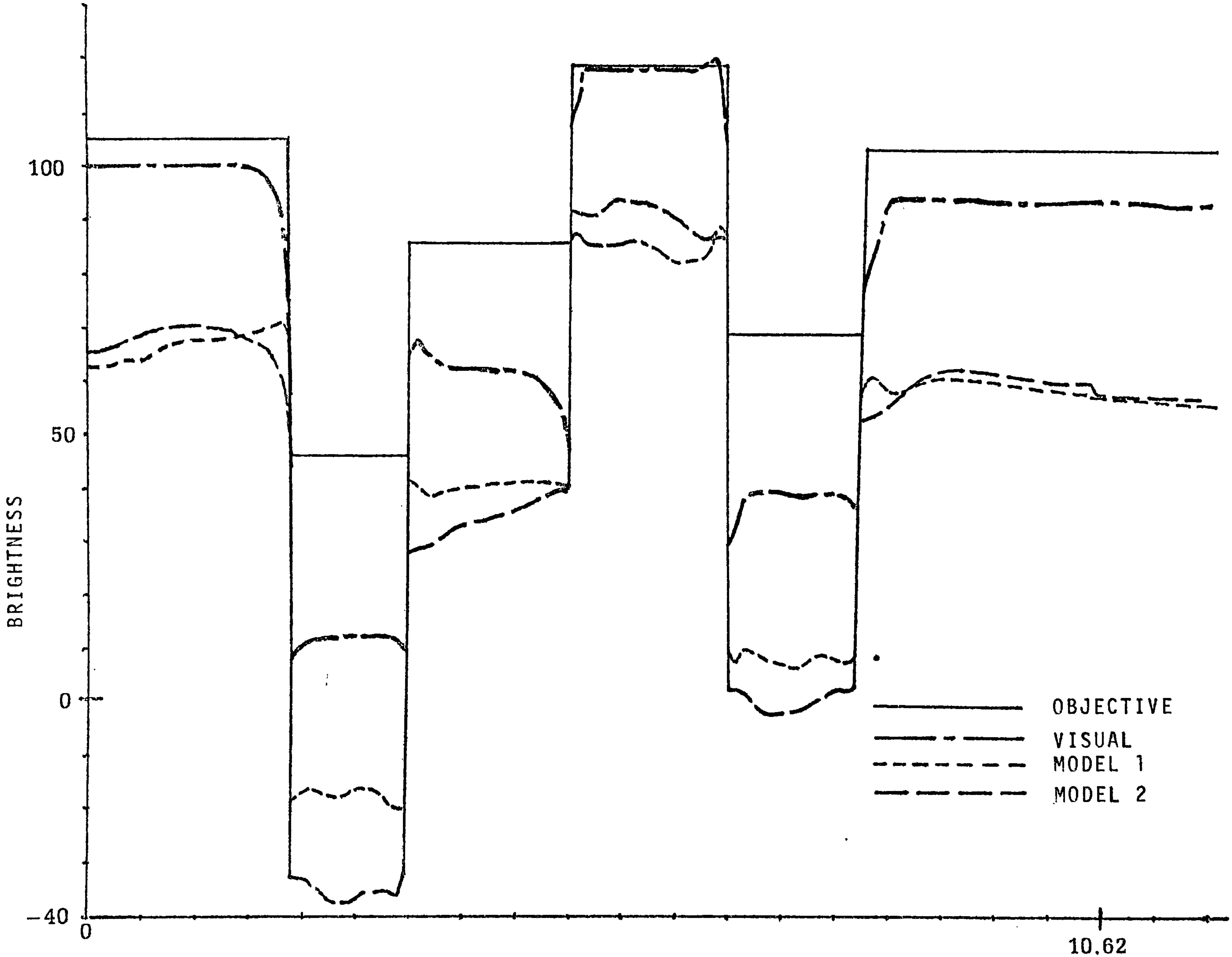


Figure 22 (Target A-3)

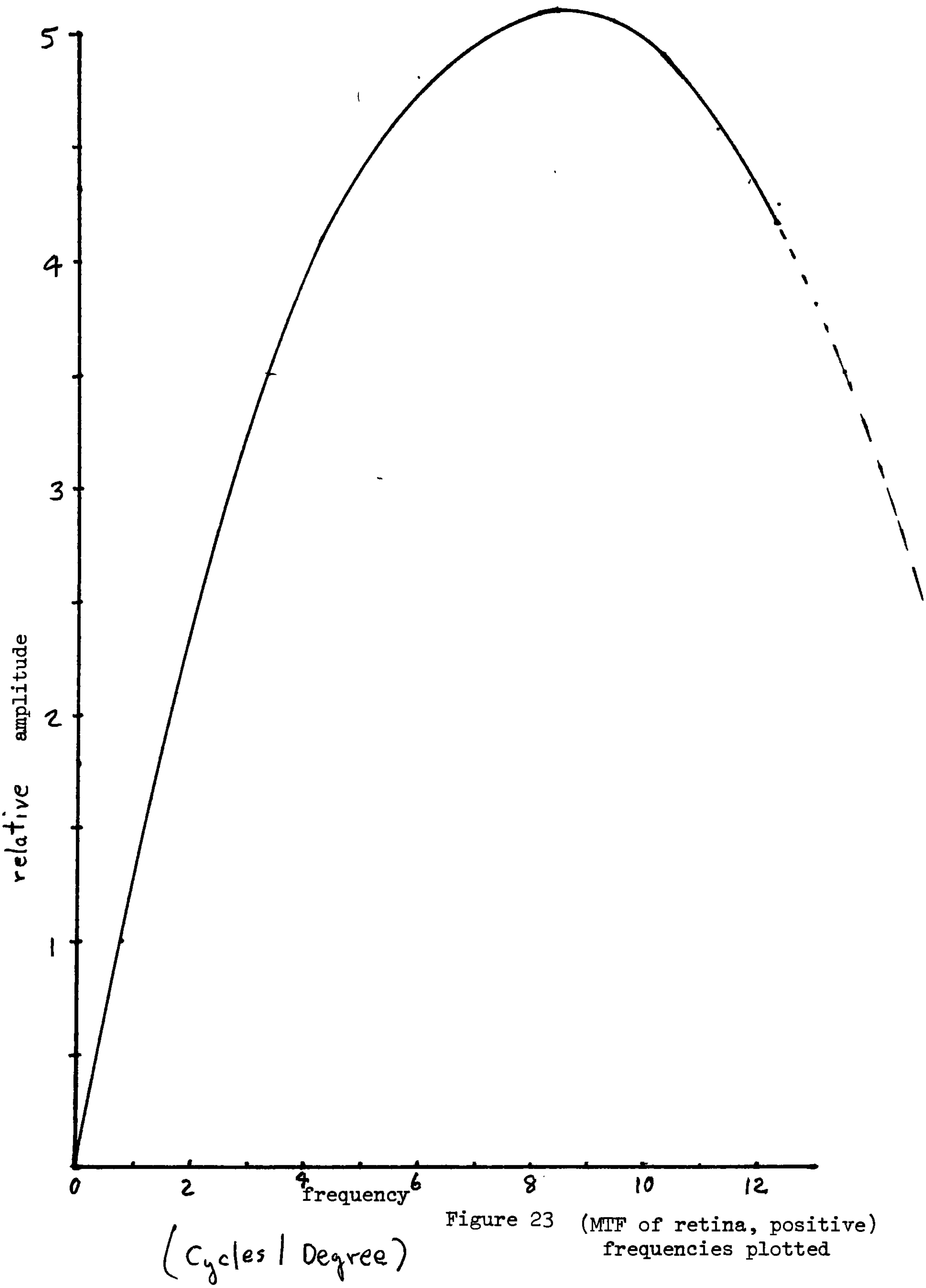


Figure 23 (MTF of retina, positive) frequencies plotted

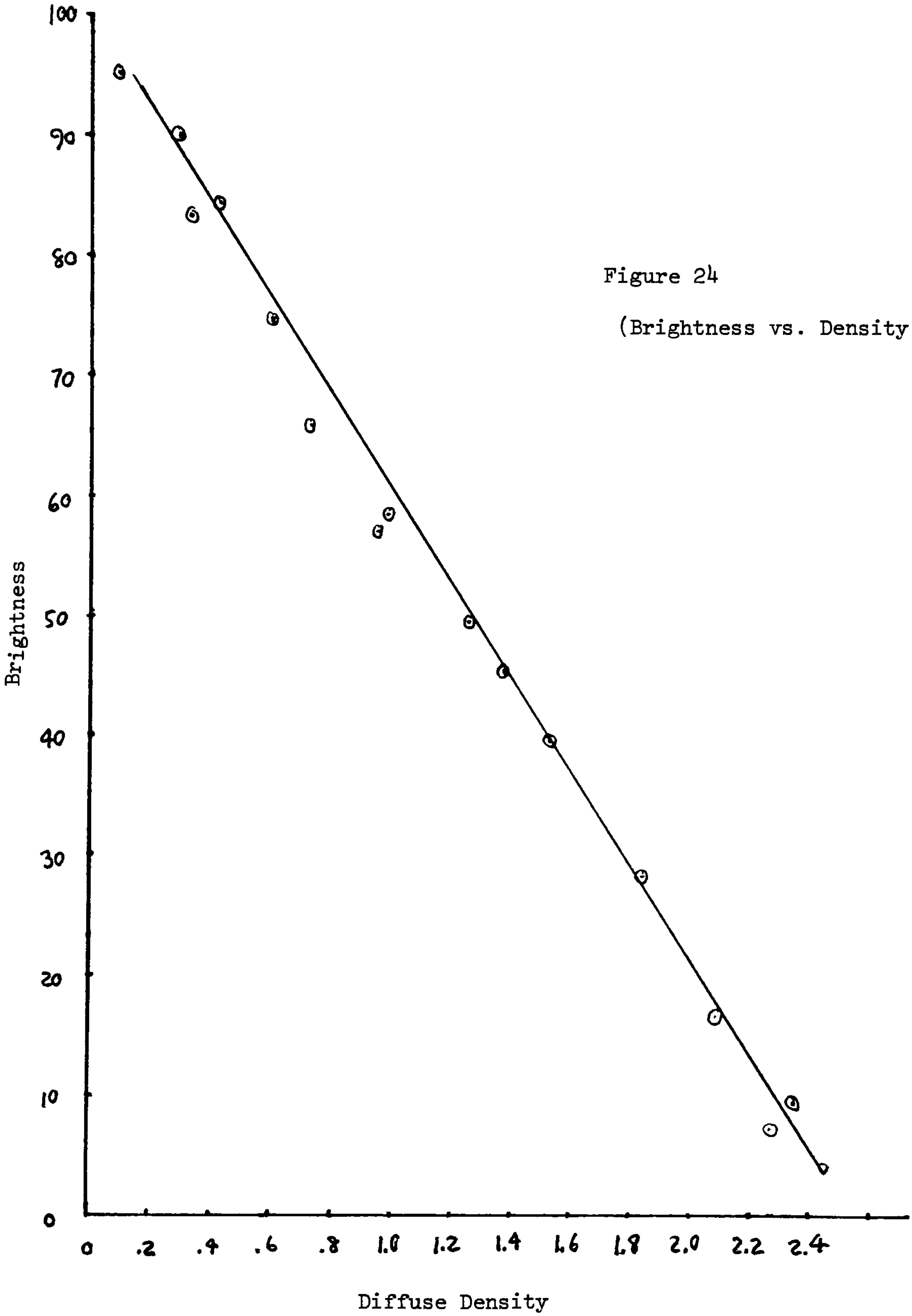


Figure 24

(Brightness vs. Density)

RESULTS AND CONCLUSIONS

The models fit the three curves with varying success. Of the two models, the first seems to come the closest to fitting the actual response. If the average value of the difference between the actual response and the predicted response(Figure 25) it is interesting to note that all of the values are positive.

Figure 25

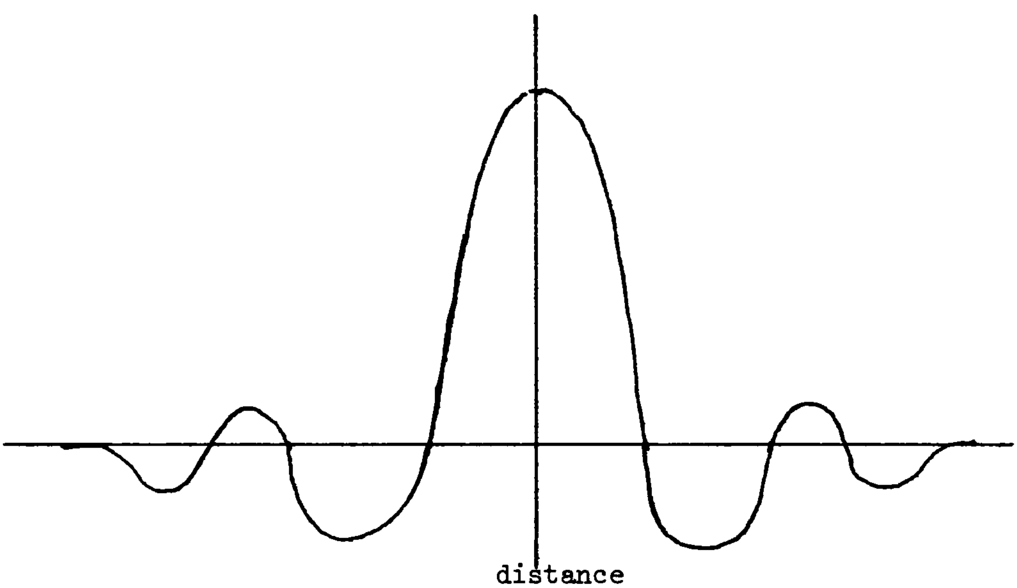
Target	Model	average error	st. dev. of errors
A-3	1	27.25	8.03
	2	32.05	10.5
A-5	1	36.42	9.31
	2	38.39	16.71
A-2	1	17.91	8.31
	2	28.73	15.02

Looking at the predicted responses versus actual responses graphically, it can be seen that the first model would do quite well if a larger constant was added. The first model does predict unequal lobing in reference to Mach Bands. However, the relative magnitudes of these are not predicted well by the model. It would appear that the integrator effectively cuts off too much high frequency information.

The second model has the fault of appearing to buldge in the centers of all the pulses. It is apparent that the smearing effect of the integrator is too strong and also that its ability to raise the predicted response has not worked very well. Another reason for failure of this model is that the I(w) from the first model was used rather than solving for I(w) over again in the context of the new model.

The relative success of the first model in predicting unequal lobes and in producing distinct levels(something the MTF does not do by itself) suggests that it may be a fruitful guide towards future experimentation. The inverse transform of I(w)MTF(w) is given in figure 26. The incidence of two negative lobes on each side is interesting and not totally inconsistent with some of the data.

Figure 26 (inverse transform of I(w)MTF(w))



Some of the greatest sources of error in this form of analysis is due to the extremely discontinuous nature of the discrete Fourier transform. Using the relatively small transform array makes interpretation of transforms difficult since the discrete transform has the property that: the more data one feeds it, the less the output data.

Not to be forgotten is the extreme variability associated with any psychophysical experiment. When this is compounded with a small test sample and a relatively limited target luminance distribution, (only combinations of rectangular pulses), significant holes or deficiencies exist. However, the size and complexity of a more complete experiment is prohibitive for the time allotted for the Bachelor of Science thesis.

As a suggestion for future research, the use of Fourier analysis should not be ignored or used in a confining fashion. Many interesting and important models can be formulated which place certain operations into a linear form while other subsequent operations are non-linear. If Fourier analysis can be used for the linear portions of the model and brute force or other mathematical techniques drafted to handle the non-linear operations, many interesting and satisfying results may occur.

End Notes

1. Cornsweet,pp. 312-324
2. Ibid. pp. 324-330
3. Ibid. pp. 330
4. Pribram, p. 17
5. Ibid. p. 31
6. Ratliff, p. 79
7. Ibid. p. 120
8. Cornsweet, p. 361

Bibliography

1. Cornsweet, Tom N., "Visual Perception," New York, New York: Academic Press, 1970,474 pp.

2. Davidson, Michael and John A. Whiteside,"Human Brightness Perception near Sharp Contours," in Journal of the Optical Society of America, David L. MacAdam--editor, Lancaster, PA: Lancaster Press Inc.,1971, pp. 530-536.

3. Land, Edwin H. and John J. McCann, "Lightness and Retinex Theory," in Journal of the Optical Society of America, David L. MacAdam--editor, Lancaster, PA: Lancaster Press Inc., 1971, pp. 1-11.

4. McCann, John J. and Jeanne L. Benton,"Interaction of the Long-Wave cones and the Rods to Produce Color Sensations," in Journal of the Optical Society of America, David L. MacAdam--editor, Lancaster, PA: Lancaster Press Inc., 1969, pp.103-107.

5. Pribram, K.H.,"Perception and Action," Universal Litho, 1969,pp.11-82.

6. Ratliff, Floyd, "Mach Bands: Quantitative studies on neural networks in the retina," San Francisco: Holden-Day Inc., 1965, 365 pp.

Appendix

Modulation Transfer Function:

For an imaging system with a given spread function in the spatial domain, the Modulation Transfer Function(MTF) is defined as:

$$\mathrm{MTF}(w) = \frac{\left|\mathcal{F}[S(x)]\right|}{\int_{-\infty}^{\infty} S(x)\,dx} = \frac{|S'(\omega)|}{S'(0)}$$

where $S(x)$ = spatial spread function

Conceptually, the MTF at a given frequency represents the decrease (though it can be increasing) in amplitude of an input signal as it is imaged.

Convolution Theorem:

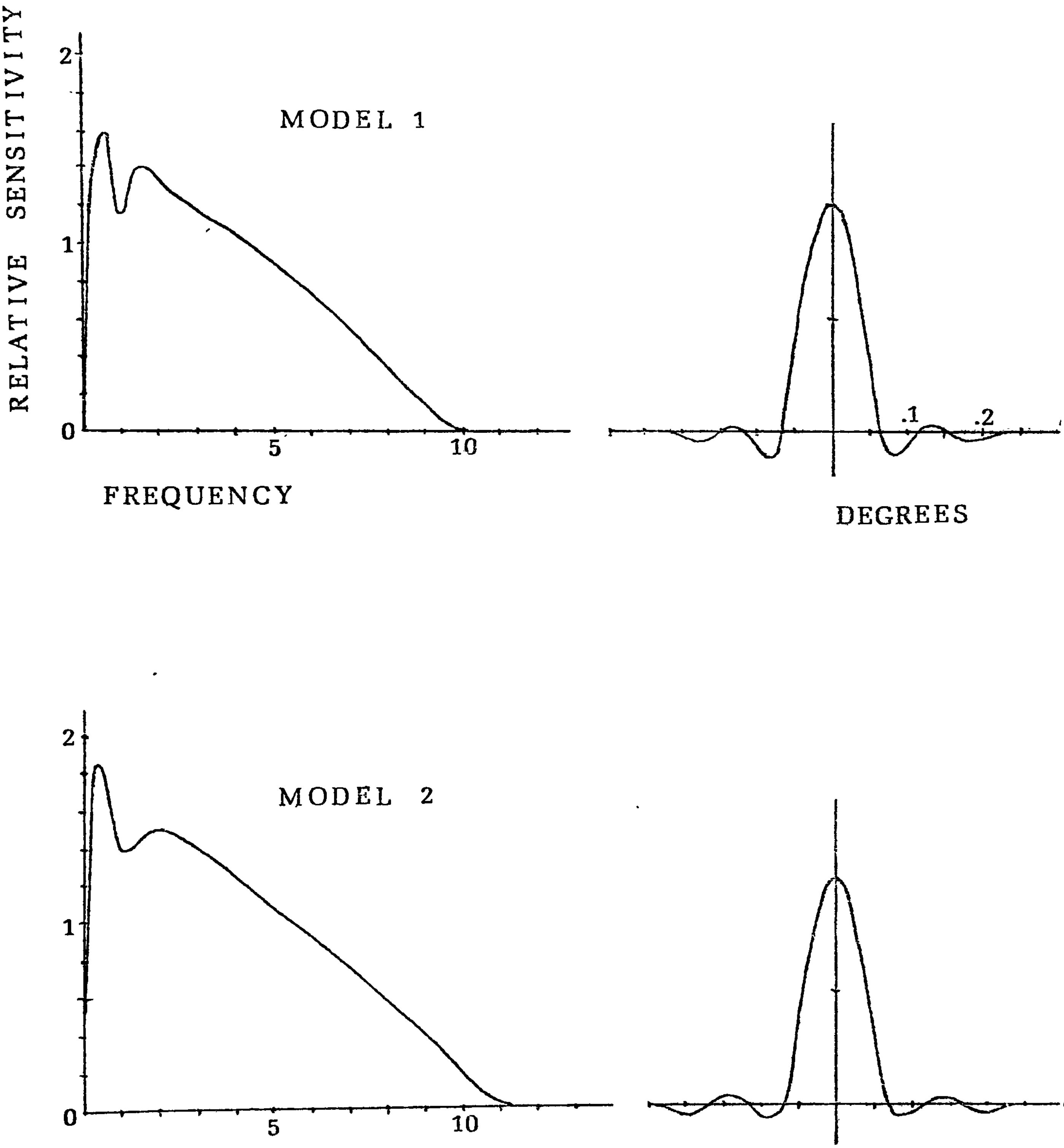
$$i(x)=g(x) * h(x) = \int_{-\infty}^{\infty} g(\varsigma)h(x-\varsigma)d\varsigma \qquad \text{(convolution)}$$

the theorem then says that:

$$I(w) = G(w) \cdot H(w) \qquad \text{therefore, } i(x) = \mathcal{F}^{-1}[G(w)H(w)]$$

$$\mathcal{F}[\ ] \rightarrow \text{Fourier transform} \rightarrow \int_{-\infty}^{\infty} e^{-i\omega x}dx$$

$$\mathcal{F}^{-1}[\ ] \rightarrow \text{inverse Fourier transform} \rightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega x}dx$$


RELATIVE SENSITIVITY
MODEL 1
2
1
0
5
10
FREQUENCY
.1
.2
DEGREES
MODEL 2
2
1
0
5
10