

AD-A077 622

AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OH
SPATIAL FILTERING AND MECHANISMS OF PERCEPTION, (U)
APR 79 A P GINSBURG
AMRL-TR-79-47

F/G 14/5

UNCLASSIFIED

NL

| OF |
ADA
077622



END
DATE
FILMED
| -80
DDC

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER AMRL-TR-79-47	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Spatial Filtering and Mechanisms of Visual Perception		5. TYPE OF REPORT & PERIOD COVERED Proceedings Paper
7. AUTHOR(s) Arthur P. Ginsburg		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, OH 45433		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F/2313AV1-22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE April 1979
		13. NUMBER OF PAGES 9
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
Presented at the Tenth Annual Pittsburg Conference on Modeling and Simulation held April 25-27 1979, Pittsburg, Pennsylvania.
Sponsored by Instrument Society of America, Instrumentation Systems Automation, 400 Stanwix Street, Pittsburg, Pennsylvania 15222.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Attempts have been made to distinguish between visual mechanisms as being space domain "feature" detectors or transform domain spatial frequency detectors. Those descriptions can be considered equivalent in terms of filtering. For example, Fourier Transformation can be viewed as space domain filtering with sinusoidal basis functions whose weighting functions can be described as edge and line detectors.

DDC
R
DEC 5 1979
REGULATED
E

LEVEL 11

ADA 077622

009 850

LB

SPATIAL FILTERING AND MECHANISMS OF PERCEPTION

Arthur P. Ginsburg

Aviation Vision Laboratory
Human Engineering Division
Aerospace Medical Research Laboratories
Wright-Patterson AFB, Ohio 45433

ABSTRACT

Attempts have been made to distinguish between visual mechanisms as being space domain feature detectors or transform domain spatial frequency detectors. Those descriptions can be considered equivalent in terms of producing similar filtered images. However, Fourier-like transformation appears to be a more accurate description of the filtering process when biological data is considered. → b PG 189

INTRODUCTION

There is still no generally accepted conceptual or physical framework with which to understand spatial vision. Few would argue the importance of receptive fields of neurons as the prime mechanisms of vision. Much research has attempted to determine if receptive fields are either feature or spatial frequency detectors. It is suggested that such a dichotomy is largely semantic and masks the true problem, how to quantitatively describe those mechanisms of spatial vision. The answer will most likely come from an understanding of the function of neuronal receptive fields.

The function of neurons in the visual system is to process selective information, that is, filter information along certain stimuli dimensions. For example, there are three mechanisms selective to only one color: red, green, blue. These mechanisms are called color channels, not color feature detectors. If the function of these mechanisms is to filter color, then the relevant stimulus dimensions to relate to these mechanisms is wavelength, hue, and intensity. This same view is suggested for mechanisms in the visual system that process spatial information. Rather than attempt to distinguish particular semantic differences of names for spatial mechanisms, perhaps a more relevant approach to their function is in terms of filtering.

The relevant properties of a filter is its bandwidth (the range of stimulus dimension that the filter is maximally tuned to), center frequency (a reference point, usually the frequency of the maximum response of the filter bandwidth), and the weighting function (the shape of the filter, the way it attenuates the information over the bandwidth).

This paper will describe certain aspects of two different filtering processes. Those processes will be related to known biological mechanisms to help determine the most relevant way of describing spatial filtering in vision. The general filtering properties of certain mechanisms will suggest a common transformation of spatial information that can be used to describe their individual function and their collective process. The relevance of this approach will be discussed in terms of quantifying visual perception.

THE SPATIAL FILTERING PROCESS

The term spatial filtering is commonly used to describe the process of attenuating the frequency spectrum of an image. Here that definition is extended to describe the attenuation of spatial size or frequency using either space or frequency domain techniques. There are two general spatial filter processes: convolution and spectral attenuation. Both processes will be discussed and related to known visual mechanisms in an attempt to determine which process can best provide a model for biological spatial filtering. One dimensional analysis will be used to simplify the discussion with no loss of generality.

Convolution is a space domain filtering process given by

79 22 185 3 138

$$p_f(nx) = \sum_{m=0}^N p_1(nx - mx^{-1}) h(nx^{-1}) \quad (1)$$

In words, pattern p_1 is rotated, that is, folded about pattern h and shifted in increments of mx^{-1} over pattern h . These patterns are multiplied after each shift. The result is a filtered image $p_f(nx)$. If either pattern is symmetric, then the folding process can be ignored. It should be noted that the process of convolution is identical to correlation if there is no folding.

Filtering using spectral attenuation requires the transformation of the input pattern into the frequency domain. Fourier transformation is the most common process used. (However, there are many other transformations that could be used for similar filtering. Discussion of these other transforms are beyond the scope of this paper and the term Fourier-like indicates the general transform process.) Fourier transformation is given by:

$$P(x) = \sum_{n=0}^N p(nx) e^{-j2\pi \left(\frac{nx}{X}\right)} \quad (2)$$

where $p(nx)$ is the input pattern and $e^{-j2\pi \left(\frac{nx}{X}\right)}$ are complex exponentials. The complex exponentials can be shown to be equivalent to complex sinusoids using Eulers identity. Thus eqn. 2 becomes:

$$p(x) = \sum_{n=1}^N p(nx) \left[\cos 2\pi \left(\frac{xu}{X}\right) - j \sin 2\pi \left(\frac{xu}{X}\right) \right] \quad (3)$$

In words, the Fourier transform requires that a pattern $p(nx)$ be multiplied by sine and cosine patterns of different spatial frequencies, resulting in two pieces of information. The cosine term is an even symmetric function, the real (Re) part of the Fourier transform. The sine term is an odd symmetric function, the imaginary (Im) part of the Fourier transform. The magnitude of the Fourier coefficients used to determine the Fourier spectrum is given by

$$M_n = \sqrt{\text{Re}_n^2 + \text{Im}_n^2} \quad (4)$$

and phase, the relative position between the cosine and sine functions, is determined from

$$\phi = \tan^{-1} \frac{\text{Im}_n}{\text{Re}_n} \quad (5)$$

The Fourier transform process described so far has neither gained nor lost any information about the input pattern. It has transformed the spatial information of the pattern into another domain. The result of filtering the input pattern by these sinusoids can be related to the amount of size or periodicity of spatial information in the original pattern. If X is the spatial period or size of a particular feature, then $x_n = n/X$ is the spatial frequency. Periodicity is the angular spatial frequency given by $k_n = 2\pi x_n$. Thus, x is cycles per unit length and k is periods per unit length. The fundamental spatial frequency and periodicity is given by x_1 and k_1 respectively. x_1, x_2, \dots, x_n and $k_1, k_2, k_3, \dots, k_n$ are the harmonic spatial frequencies and periodicities respectively.

The main point is that the decomposition of the original pattern by the sinusoids is itself a filtering process. Further filtering is accomplished by using only a subset of the transformed data. If a subset of that data is retransformed into the space domain using an inverse Fourier transform, the result will be a filtered image. Passing only the low spatial frequencies will create a smoothed image as shown in Figure 1, $f_c = 4$. If only the high spatial frequencies are passed, an image having only edges and finer details result as shown in Figure 1, $f_c = 64$. Thus, by passing a certain portion of Fourier components, or spatial information resulting from filtering an object with periodic patterns, different spatial information that can be related to the size of features that make up the original object can be extracted for subsequent use. This important point will be discussed further in a later section.

CONVOLUTION OR FOURIER TRANSFORMATION AS A MODEL OF VISUAL FILTERING PROCESSES?

Filtering by convolution or spectral attenuation can produce identical results. Indeed, there is a theorem in Fourier analysis stating that the convolution of two functions is identical to the product of the Fourier transformed functions. Thus, from a functional point of view, neither method of filtering is indistinguishable from the other. There is, however, a difference between these two methods in how they perform the filtering.

The Fourier process, unlike convolution, does not require a shift of the decomposing pattern (complex sinusoids in the case of the Fourier transform) as part of the filtering process. The Fourier process does its filtering using stationary patterns. Unlike convolution, filtering using the Fourier technique is a two-stage process. The input pattern is first decomposed or filtered using limited periodic patterns (limited in the sense of using a discrete Fourier transform having finite sinusoids, more relevant to biological analogues than the infinite sinusoids of the continuous Fourier transform). The second stage of filtering requires that certain ranges of the decomposed pattern be extracted over certain bandwidths (and orientation in the case of the two-dimensional Fourier transform). Note that there is no need to create a spectrum from these data to describe the process. Only the transformation of the original pattern features into band-limited ranges of spatial information related to periodicity is necessary to specify the Fourier process. The result is a filtered image similar to that produced by convolution.

Which process is more meaningful in terms of our current knowledge of the function of certain receptive fields? For this argument, the filtering process of the retina and LGN will not be discussed. The major decomposition of the visual scene appears to take place at the visual cortex. It is here that perhaps a distinction can be made between convolution and transformation as a key filtering process of spatial vision.

Convolution, if used by biological systems, would require two processes. One process is needed to shift the image of the input pattern over the filtering mechanism, the receptive field, or, like a shift register, record the outputs of a group of similar sized receptive fields, located along the extent of the pattern features. There is no evidence for biological mechanisms having a shift process. A counter argument for convolution could be the necessity of eye movements to mediate spatial vision: it's the function of eye movements to sweep the retinal image across space and convolve it with visual receptive fields. However, the magnitude of involuntary saccadic eye movement is about 12 minutes of arc. Since this would mean that objects whose size is greater than 12 minutes of arc would not be convolved, this hypothesis seems untenable. Voluntary eye movement can be ruled out for two reasons: they are generally random and objects can be correctly processed under tachistoscopic presentation even though there is insufficient time for voluntary eye movements.

There is another way to perform filtering using convolution. If convolution is implemented by rows of evenly spaced, adjacent, possibly overlapping, receptive fields of similar spatial properties, then convolution can be achieved by reading out the summed product for each group of similar receptive fields. But this is the process earlier defined for Fourier transformation. Therefore, implementing convolution using in-place filters is given by the mathematics of Fourier transformation.

A Fourier processes requires certain additional necessary and sufficient conditions be satisfied before it can be considered to be biologically relevant. First, there must be evidence for receptive fields having periodicity over a range of periods consistent with the size of objects that are seen. Second, these receptive fields must have even and odd symmetry, i.e., have cosine and sine weighting functions. Third, the receptive fields must be collected over certain ranges of periodicity for subsequent use. Filtering with one extended sinusoid would result in only the first stage of a Fourier process, not particularly useful by itself. Finally, even if these conditions are satisfied, further criteria for use is simply utility. Is this method of describing certain visual processes useful? Does it provide insight? Each point will be discussed in turn.

Hubel and Wiesel (1) found receptive fields in visual cortex maximally tuned to two general spatial dimensions: size and orientation. The general weighting function of these rectangular receptive fields are periodic, having alternating increasing and decreasing weighting considered to be regions of "excitation" and "inhibition." Although Hubel and Wiesel cells are typically represented by only 1 or 1.5 periods, recently Albrecht (2) has shown simple cells having over 4 periods. Thus, the first condition of a Fourier-like transform, periodicity is satisfied.

The next condition of even (cosine) and odd (sine) symmetric receptive fields is easily satisfied. Receptive fields having these kinds of weighting, shown by Hubel and Wiesel, have been characterized as line and edge detectors. Acknowledging these receptive fields as cosine and sine filters would explicitly imply a Fourier-like process. The idealization of these periodic receptive fields having three cycles is shown in two-dimensions in Figure 2a (cosine on top, sine on bottom). Figure 2b shows a three-dimensional view of these receptive fields. The typical schematic of the Hubel and Wiesel cells is shown in the upper right corner of their respective idealized functions. It is pointed out that these idealized receptive fields are the first three sine and cosine terms of a two-dimensional discrete Fourier transform.

The third point is that these periodic receptive fields be used in clumps collected over space having limited bandwidths. Evidence for simple cells having different bandwidths and different center frequencies was shown by DeValois (3). Bandwidths, ranging from 0.5 to over 2 octaves, had center frequencies that covered a large range of size.

The preceding considerations suggest that Fourier-like transformation is a relevant way to describe certain visual processes. The suggestion of sinusoidal receptive fields is idealized. A more exact description of the receptive fields would cause the periodicity to be tapered consinusoidally. An approximation to these filters would be Hermitian polynomials (Young, 4). However, it may be more useful at this stage of our knowledge to pursue the function of the two general kinds of receptive fields that exist. Filtering done by the author using many filters having different weighting functions suggests that the particular shape and bandwidth of the filters do not critically change filtered images for perhaps 80-90% of perceptual phenomenon. Exact brightness functions and lengths of features will be somewhat dependent on the particular bandwidth and weighting function.

It is suggested that the relevant information for spatial vision is (excluding for the moment color, motion, etc.) is periodicity (spatial frequency), orientation, contrast and position in the visual field. Note that periodicity is substituted for size. If the visual system were interested only in filtering out size information in objects, then the extent of either an even or odd symmetric field would need be only one cycle. Similar shaped receptive fields could be made having different widths to cover a full range of size. But many receptive fields have been found to have over 4 cycles.

Here is perhaps the best distinction that can be made to determine whether receptive fields are encoding size or another stimulus property. Since spatial frequency is by definition a reciprocal of size, that description does not help any argument distinguish between size and spatial frequency as prime stimulus attributes for receptive fields. However, what does distinguish between a size or other measures is an additional property found in certain receptive fields in the visual system, periodicity or the number of cycles, i.e., the number of bands of positive and negative weighting across the receptive field. It would seem, therefore, that the function of the receptive field may be best described by the periodicity or number of cycles per unit distance that occurs across the receptive field rather than size.

It follows that the function of certain neurons in the visual system is to perform a crude periodicity or frequency analysis. This analysis supports the suggestions of many researchers who, from engineering, neurophysiological, and psychophysical considerations, have suggested that the visual system is a crude Fourier analyzer (e.g., see Ginsburg, 5).

Agreement between the mathematics and the biological process still does not mean that the visual system is creating a two-dimensional Fourier transform spectrum. However, the biological process appears to be taking the information from the spatial domain and decomposing it into a form analogous to a two-dimensional Fourier transform domain. The distinction is important because in the former case one expects to see a literal spectrum at the visual cortex. This is not necessary in the latter case. Here the expectation is for the visual system to exhibit certain processing characteristics that would be consistent with a two-dimensional Fourier-like transform. In other words, it would be expected that certain visual phenomenon exhibit spatial information filtered along the two dimensions that are processed by a Fourier-like transform: spatial frequency or periodicity and orientation.

IMPLICATIONS FOR SPATIAL TRANSFORMATIONS AND PERCEPTION

Are there any perceptual processes that suggest processing is constrained to these dimensions? The answer is yes and is typified by the perception of a field of multistable triangles such as those shown in Figure 3a. When these multistable triangles are gazed upon they all perceptually group together in one of three orientations, changing shape and appearing in depth (Attneave, 6). If these triangles are randomly placed, it is very difficult, if not impossible, to make one of these triangles flip independently of the other triangles. Given that the psychophysics and neurophysiology suggests that the spatial information of these triangles are filtered in terms of spatial frequency and orientation typified by the Fourier magnitude spectrum (Figure 3b), what would happen if these objects are filtered along limited ranges of those two dimensions. This is shown in Fig. 3c, c^1 , d, d^1 , e, e^1 . Filtering out one of two channels in each of three possible orientation combinations demonstrates this perceptual experience. Perhaps similar filtering is occurring in the visual system.

The preceding example demonstrated one consequence of being constrained to process spatial information in a Fourier-like way. The next example demonstrates the kind of spatial information that can be extracted from band-limited filter mechanisms. There is much evidence to suggest that the visual system is extracting spatial information using mechanisms whose bandwidths are about 1 to 2 octaves in spatial frequency and ± 15 degrees in

orientation. These mechanisms are called channels, a name coined by Campbell and Robson (7). These channels, if used for object recognition, cannot be single filters such as a receptive field having only one cycle because they would signal similarly for dissimilar objects. Channels filtering complex spatial information need to be created from a range of smaller receptive fields to provide more finely sampled spatial information and phase over the channel bandwidth.

Bandwidths, center frequencies, and weighting functions relevant to object recognition can be deduced from biological data. A complex object such as a portrait requires a bandwidth greater than 1 but less than 2 octaves to provide sufficient spatial information for most visual tasks (Ginsburg, 5). This is consistent with bandwidths obtained from psychophysical and neurophysiological data. Spatial frequencies closer than 1 octave appear to interact with one another, therefore, channel independence requires that they be separated by about 1 octave. The weighting function for channels can be obtained from many psychophysical data (e.g. Blakemore and Campbell, 8; Mostafavi and Sakrison, 9). Further consideration of retinal sampling and the preceding analysis suggests a maximum of 7 channels for central vision (Ginsburg, 5).

These data were used to filter a portrait as shown in Figure 1. The numbers beneath each portrait is the center frequency, f_c , for each channel. The center frequency referenced to the face width is $f_c/2$.

Note the different information about the portrait that can be found in each of the seven two-octave bandwidth channels. The existence of an object is evident from the large regions of different contrast in the first channel ($f_c=1$). The gross elliptical shape of an object that appears right side up is seen in the second channel ($f_c=2$). The third channel ($f_c=4$) definitely provides enough information to classify the object as a face. The identification of the face needs a little more information. The fourth channel ($f_c=8$) suggests that the face is that of a woman from the hair style, etc. Identification would seem possible given a limited set of similarly filtered portraits to choose from. Clearly, the face in the fifth channel ($f_c=16$) is the same as that of the original portrait, and identification could present no problem given any large set of other portraits. If information about the details of the portrait is needed, e.g., edges and lines, the hair across her forehead, the texture of her hair, size of the pupils, outline of her lips, then that information is found in the two remaining channels ($f_c=32, 64$). The last channel, $f_c=128$, has virtually no usable information. Thus, these seven channels create a hierarchy of filtered images that provide the full spectrum of information needed for any perceptual task.

These results do not imply that all of spatial vision should be described only by Fourier techniques. There are certain analysis that may be best described in the space domain of filtered images. For example, clustering, symmetry, and figure-ground analysis using images filtered from biological data of complex objects have been demonstrated by Ginsburg (5, 10). Recent work in this area using a different approach to texture analysis is given by Marr, Palm, and Poggio (11). Just as the engineer uses convolution and Fourier analysis as dictated by the problem to be solved, similar criteria is suggested for helping select the techniques for analyzing problems in vision.

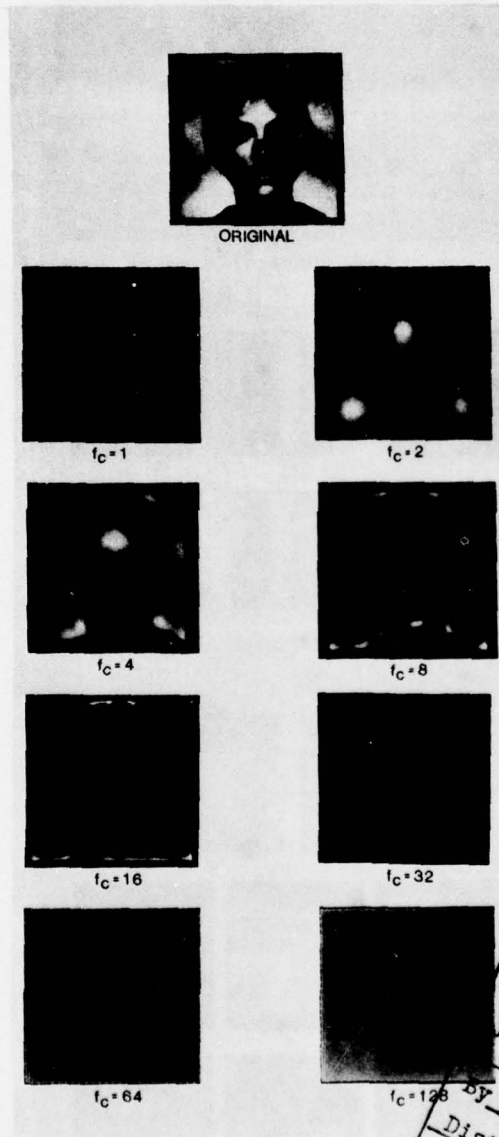
Biological data have been used to create filters that describe the overall filtering characteristics of the visual system. Those results along with other filtering based on channel properties provides a parsimonious, quantitative description of spatial vision (Ginsburg, 5). In general, the overall filtering of the visual system limits the size of objects that can be seen. Further filtering, based on channels, produces filtered images that correspond to forms that are perceived. The activity of all the channels produces the global picture that is seen.

SUMMARY AND CONCLUSIONS

The two-dimensional discrete Fourier transform is a process that requires the input object to be multiplied by a bank of even symmetric cosinusoidal and odd symmetric sinusoidal space functions having periodicity. The periodicity corresponds to the number of cycles per receptive field width. The weighting function of the periodicity is alternating positive and negative values. These properties were related to the properties of cortical receptive fields. Certain receptive fields may functionally perform a generalized periodic analysis for spatial vision. Since these functions are two-dimensional, they are sensitive to orientation, similar to cortical receptive fields. Therefore, it was argued that the visual system contains mechanisms that can process spatial information consistent with the mathematical process of a two-dimensional discrete Fourier-like transform. This analysis of the filtering process of certain visual cortical mechanisms may lead to further understanding into visual perception.

REFERENCES

1. Hubel, D. H. and Wiesel, T. N., "Receptive fields of single neurons in the cat's striate cortex," J. Physiol. 148, 574-591, 1959.
2. Albrecht, D., "Analysis of Visual Form," Ph.D. Dissertation, University of California, Berkley, California, 1978.
3. DeValois, R. L., "Color coding and lateral interactions," abs. Workshop on Spatial Contrast, Amsterdam, the Netherlands (4-9 January), 1976.
4. Young, R., "Orthogonal Basis Functions for Form Vision Derived from Eigenvector Analysis," abs., Association for Research in Vision and Ophthalmology, Sarasota, Florida, 30 April-5 May, 1978.
5. Ginsburg, A., "Visual Information Processing Based on Spatial Filters Constrained by Biological Data," Ph.D. Dissertation, University of Cambridge, 1977. (In press as AMRL-TR-78-129.)
6. Attneave, F., "Triangles as ambiguous figures," Am. J. Psychol., 181, 447-453, 1968.
7. Campbell, F. W. and Robson, J. G., "Application of Fourier analysis to the visibility of gratings," J. Physiol., 197, 551-566, 1968.
8. Blakemore, C. and Campbell, F. W., "On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images," J. Physiol. 203, 237-260, 1969.
9. Mostafavi, H., and Sakrison, D., "Structure and properties of a single channel in the human visual system," Vision Res., 16, 957-968, 1976.
10. Ginsburg, A. P., "Pattern recognition techniques suggested from psychological correlates of a model of the human visual system," Proc. 1973 National Aerospace Electronics Conference (NAECON), Dayton, Ohio: IEEE Trans. on Aerospace and Electronic Systems, 309-316, 1973.
11. Marr, D., Palm, G., and Poggio, T., "Analysis of a Cooperative Stereo Algorithm," Biol. Cybernetics, 28, 223-239, 1978.



ORIGINAL

$f_c=1$

$f_c=2$

$f_c=4$

$f_c=8$

$f_c=16$

$f_c=32$

$f_c=64$

$f_c=128$

FIGURE 1

Accession Number	
NTIS GRA&I	
DDC TAB	
Unannounced	
Justification <input checked="" type="checkbox"/>	
Distribution/Availability Code	
Dist	Avail and/or special
A	

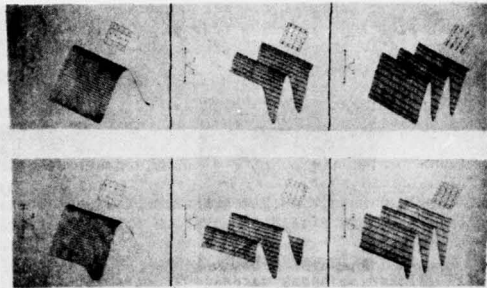


FIGURE 2a

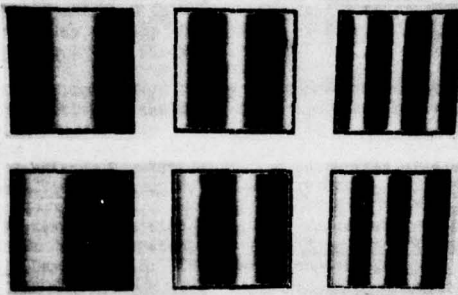


FIGURE 2b

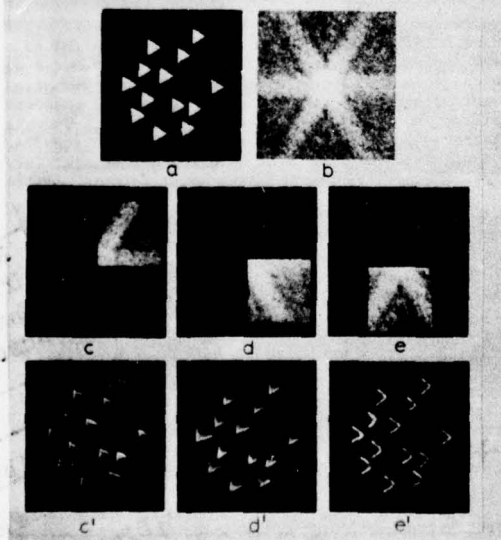


FIGURE 3