Application of two-dimensional Fourier transforms to problems of visual perception

F. L. ROYER

Cleveland Veterans Administration Medical Center, Cleveland, Ohio 44141

and

M. S. RZESZOTARSKI and G. C. GILMORE

Case Western Reserve University, Cleveland, Ohio 44106

Neurophysiological and psychophysical evidence indicates that the visual system performs a sort of Fourier analysis of retinal variations in brightness through multiple spatial frequency-tuned channels. Two types of applications of techniques of two-dimensional Fourier analysis are discussed: (1) modeling and simulation of visual functions, and (2) generation of stimuli in which the spatial frequency information is controlled. Examples of both these applications are presented.

The development of the fast Fourier transform (FFT) has made practical the analysis of the very large amounts of information in two-dimensional images (Rzeszutarski, Royer, & Gilmore, 1983). The technology of image processing by computers made possible by FFT can be applied to problems of perception. The rationale for application arises from the overwhelming neurophysiological and psychophysical evidence that the nervous system performs a sort of Fourier analysis of visual information (for a discussion of the issues, see reviews by DeValois & DeValois, 1980; Graham, 1981; Sekuler, 1974; Weisstein & Harris, 1980). Two applications are discussed here: (1) modeling and simulation, and (2) generation of controlled stimuli.

According to contemporary models of vision, there are multiple spatial frequency-tuned neural channels that analyze the variations in brightness on the retina as sinusoidal waves varying in spatial frequency, amplitude, phase, and orientation. Four spatial frequencies of sinusoidal waves are illustrated in Figure 1. The lower the spatial frequency, the larger the spatial extent or width of the bars. Spatial frequency is typically measured in cycles per degree of visual angle. Amplitude is the contrast (C) in brightness of the wave that is typically expressed as the ratio of the difference between the brightness of the peak and trough of the wave to their sum, or $C = (L_{\max} - L_{\min}) / (L_{\max} + L_{\min})$, where L is luminance. Phase gives the position of the

Figure 1. Vertical sinusoidal gratings of four different frequencies: 4, 8, 16, and 32 cycles per picture width; picture width is one-quarter the width of the figure.

information relative to the origin. Orientation is spatial direction of the wave.

Orientation is a property of spatial frequency information not present in unidimensional Fourier analyses and can be seen in the power spectrum that gives the amount of energy in each frequency. Figure 2 shows two alphabetic letters and their power spectra. Lightness in the picture of the spectrum is proportional to the power or energy in the frequency; the lighter the point, the greater is the power.

This work was supported by NIA Grant R01 AG 03178 and Veterans Administration Medical Research Service. Reprint requests should be sent to Fred L. Royer, Psychophysiological Research Laboratory, 151(B), VA Medical Center, Brecksville, Ohio 44141.

Figure 2. Two digitized alphabetic letters and their power spectra.

Figure 2 demonstrates that orientation is an important aspect of the two-dimensional analysis. At the center of the spectrum is the value for the zeroth harmonic, which gives the average brightness of the entire spectrum. Radiating outward are increasingly higher frequencies. At the edges are the highest possible frequencies, given the sampling rate. Because the orientations of sampled waves are perpendicular to spatial axes (horizontal and vertical), the axes of the spectrum carrying the vertically and horizontally oriented information appear rotated 90 deg.

The letter H has information primarily in the horizontal and vertical directions; the lightest part of the picture of the spectrum is along those axes. The letter W, by contrast, has diagonal information. Considerable energy is distributed in spatial frequencies that are oriented at an angle to the horizontal and vertical axes.

VISION AND FOURIER ANALYSIS

As indicated earlier, there is strong evidence that the visual system performs some approximation of a Fourier analysis. There are multiple spatial frequency-tuned channels in the visual system of two different types, "transient" and "sustained" (Breitmeyer & Ganz, 1976). Transient channels are sensitive to low spatial frequencies, have short latencies, and are active for a brief time. Sustained channels are sensitive to intermediate and high spatial frequencies, have long latencies, and are active for sustained periods when stimulated. The latencies of the channels are directly related to their spatial frequency sensitivity. The lowest spatial frequency information arrives earliest for processing, and progressively higher frequencies arrive later.

Recent evidence (Wilson & Bergen, 1979) suggests that there are two types of transient and two types of

sustained channels, each having its own particular frequency response characteristics.

The sensitivity or frequency response characteristics of the channels vary greatly. Figure 3 shows threshold sensitivity of the eye for contrast of vertically oriented sinusoidal gratings. Humans are insensitive to very low frequencies, most sensitive to frequencies from 2 to 4 cycles/deg, and, again, increasingly less sensitive to higher frequencies. Sensitivity of channels varies in another important way not shown; it varies with the orientation of the frequency. In low spatial frequencies, sensitivity tends to be fairly equal for different orientations, but as frequencies increase, especially in the high range, sensitivity varies considerably with the orientation of the frequency. Sensitivity is greatest for vertical gratings, less for horizontal, and least for diagonal gratings. Ginsburg (Note 1) has modeled these anisotropies in his two-dimensional transfer function.

Another important characteristic of the channels is that they vary in their sensitivity to temporal changes in contrast. Current evidence indicates that transient channels, which are sensitive to low spatial frequencies, are also sensitive to high temporal frequencies (e.g., flicker); sustained channels, which are sensitive to high spatial frequencies, are also sensitive to low temporal frequencies. High temporal frequency sensitivity makes the transients sensitive to motion and to the changes in brightness distribution on the retina after a saccadic eye movement.

Finally, it should be noted that these channels interact. They not only carry differential spatial and temporal information, but they interact to preserve the continuity of visual experience and the integrity of the input despite the rapid changes in information resulting from the organism's information-seeking activities through eye movements.

The multiple-channel model has demonstrated significant explanatory and predictive power for such perceptual phenomena as apparent movement (von Grunau,

Figure 3. Typical human contrast sensitivity function for vertical sinusoidal gratings. Sensitivity is the reciprocal of contrast at threshold (after Ginsburg, Note 1).

1978, 1981) and critical flicker fusion (Levinson, 1968; Matin, 1962). It has produced a comprehensive theory of visual masking (Breitmeyer, 1980; Breitmeyer & Ganz, 1976). One can anticipate that our understanding of other perceptual and cognitive phenomena will be increased by application of the model. Image processing technology and the FFT provide tools for making those applications.

In the discussion that follows, the primary tools of the application are the forward FFT, convolution, and the inverse FFT. An image in the spatial domain is sampled and digitized. A forward FFT performs the transform of the image information into the frequency domain. Through convolution, a transfer function that describes the weighting to be given to the frequency information is applied. This produces a filtered function that is then back-transformed into the spatial domain by the inverse FFT. This newly synthesized image is a filtered-image displaying the spatial information appropriately modified by the transfer function of the system.

MODELING AND SIMULATION

Extensive modeling using the FFT has been done by Ginsburg (Note 1), who developed a model of face perception by applying various filters and cutoff frequencies to photographs. He found that low-pass filtered information is adequate for facial recognition.

Ginsburg (Note 1) also developed a model of certain gestalt principles. In conformity with neurophysiological evidence about the latencies of channels, he developed a model in which the perceived image is developed progressively from the lowest to highest frequencies. Figures 4 and 5 show what information might be available at different times in the formation of the letter G composed of discrete dots. Presumably, the information is continuously accessible for decision making at any time during the development of the image. Thus the earliest low-frequency information would provide evidence of an object's presence. Later, intermediate-frequency information would provide evidence for identifiability, and still later, the high-frequency information would provide evidence for discrimination.

Ginsburg (Note 1) stresses the consequences of low-pass filtering. Notice in Figure 4 that although the image of a G is only a collection of dots in space, low-pass filtering at very low cutoff frequencies closes the spaces and generates a coherent object. According to his model, the qualities of the image after low-pass filtering are preserved in the final percept along with the high-frequency information, which tells us that there are sharp-edged dots separated in space. Marr (1982) postulates a contour-generating process that operates on the filtered information that could account for closure.

Ginsburg (Note 1) has proposed that low-pass filtering accounts for gestalt organizing principles other

Figure 4. Digitized display of the letter G composed of dots (lower left) and low-pass filtered images at cutoff frequencies of 1 (upper left), 2 (lower right), and 4 cycles/deg (upper right), assuming a letter size of 1.7 deg of visual angle.

Figure 5. Digitized display of the letter G and low-pass filtered images with cutoff frequencies of 6, 8, and 10 cycles/deg arranged as in Figure 4.

than closure. He has shown that both proximity and similarity principles of grouping can be explained by the proposed processes. In Figure 6, we illustrate the proximity principle.

Ginsburg's (Note 1) low-pass model of global perception may also have some power in explaining some findings of Pomerantz (1981) in his extensive studies of perceptual organization. Pomerantz used displays such

Figure 6. Proximity grouping of squares into columnar organization with low-pass filtering with cutoff frequencies of 1, 2, and 4 cycles/deg if original display is 1.7 deg of visual angle.

(a)

(b)

(c)

(d)

Figure 7. Displays containing one quadrant that differs from the other three. Locating the disparate quadrant may be easy or difficult, depending on certain organizational principles such as local emergent features, according to Pomerantz (1981).

as those in Figure 7. He found that certain effects in discrimination were best accounted for by hypothesizing the presence of emergent features in configuration. In particular, he conjectured that emergent features were caused by intersections and angles. We simulated the effects of Ginsburg's model on such displays. Figures 8 and 9 illustrate two displays similar to those of Pomerantz

that were filtered with Ginsburg's two-dimensional filtered function. Subjects would easily and quickly locate the one quadrant that does not match the others; in the other, subjects would have greater difficulty finding the disparate or odd quadrant. As the figures and other patterns we filtered show, there are indeed features created at intersections and angles that, under Ginsburg's model, "emerge" from a low-pass system in the same sense that closure emerges from it. There is an intensification of contrast at these points in the image. In the Pomerantz paradigm, discrimination should be speeded if the emergent features, or regions of intensified contrast, are markedly different within the display. If the features are similar across elements of the display, discrimination should be slowed. Pomerantz' data support this explanation. This model is different from the feature-detector model of vision because it would not be the individual features (diagonal and horizontal or vertical lines) that make for differences in discriminability, but the presence of features emerging from filters. The low-pass filtered letters in Figure 10 show the intensifications of contrast at angles and intersections and demonstrate "emergent" features that could contribute to confusability and speed of identification of letters.

Simulation allows the experimenter to examine what the hypothesized filtering functions do to information so as to make predictions about how performance should vary as a function of stimulus parameters. As more becomes known about the transfer characteristics of visual channels, simulation should facilitate modeling and empirical study.

As already shown (Figure 3), the amount of contrast needed to just see a sinusoidal grating varies greatly

Figure 8. "Emergent" features in a Pomerantz display filtered with an ideal square filter after prefiltering with Ginsburg's two-dimensional modulation transfer function (Ginsburg, Note 1). Finding the odd quadrant in the original display is easy.

Figure 9. "Emergent" features in a Pomerantz display filtered with an ideal filter after prefiltering with Ginsburg's two-dimensional modulation transfer function (Ginsburg, Note 1). Finding the odd quadrant in the original display is difficult.

Figure 10. Four alphabetic letters low-pass filtered with cutoff frequency = 3 cycles/deg. "Emergent" features are present at this level.

with the spatial frequency of the grating. It is known that the function differs from individual to individual. There are marked age-related differences in the function (Banks & Salapatek, 1981; Sekuler & Hutman, 1980), and these differences can not be attributed to differences in the optics of the eye. If we wish to observe what information is available to different individuals, such as in the presence of pathology, the spatial frequency contrast sensitivity function, or CSF, as it is commonly known, can be used as a weighting function to filter an image (see Ginsburg, Note 1).

It should be mentioned that image processing methods were used extensively by Marr (1982). In an impor-

tant new book, he develops a model of contour generation and stereoscopic and motion perception.

GENERATING CONTROLLED STIMULI

Another value of the use of the FFT is to generate stimuli in which the spatial frequency content is controlled. The use of the FFT is illustrated by some stimuli generated for a study of backward masking by pattern in aging and by some data obtained from a small experiment, with two laboratory staff serving as subjects, which used these stimuli.

A VISUAL MASKING EXPERIMENT

Among the major pieces of convincing evidence for the Fourier analysis at a perceptual level is the work of Weisstein, Harris, Berbaum, Tangney, and Williams (1977). They showed in a metacontrast experiment that the amount of masking was determined by the overlap in the frequency content of the mask and target. For example, a dot could provide good contrast reduction of a bullseye, but not of diagonally oriented square-wave grating. On the other hand, a single diagonal bar provided good contrast reduction of the diagonal gratings, but not of the bullseye. The efficiency of masking appears to depend on the disruption or interruption of the processing being done in specifically tuned channels.

Pattern masks are typically large displays containing many edges in different orientations. We would expect the efficiency of masking to depend on the overlap in

Figure 11. Digitized images and power spectra of a letter and mask similar to those used in studies of backward masking by Walsh (1976).

Figure 12. Digitized mask filtered with a low-pass, high-pass, and bandpass filter (see text).

Figure 13. Digitized letter W filtered with a low-pass, high-pass, and bandpass filter (see text).

the spectra of target and mask, such as that illustrated in Figure 11. The spectrum is broad and contains information in many orientations. It should be a relatively good mask.

In this experiment, we wanted to determine the relative contribution of different bands of spatial frequency content of the target and mask on backward masking. For this purpose, control of the frequency content is needed. Various types of filters are available. One can retain only the low spatial frequencies of the mask with a low-pass filter. To retain only the high spatial frequencies, one can use a high-pass filter. To selectively pass or reject certain frequencies, one can use bandpass or notch filters. Although we did not do it in this experiment, orientation of certain frequencies can be controlled by creating filters that attenuate,

pass, or reject only frequencies in certain orientations. Figures 12 and 13 illustrate the effects of three types of filters used in the experiment on a pattern mask and some letters.

Method

Subjects. Two male practiced observers who were laboratory staff served as subjects. Their ages were 25 and 32 years.

Stimuli. Thirteen letters were drawn from the Letraset 439 stencil letters used by Walsh (1976) in a number of experiments on backward masking in aging. Walsh's mask was duplicated. The letters we used were: A, C, E, F, H, L, M, N, O, T, W, and Y. They were selected according to their confusability. Their size was 1.7 deg of visual angle. Letters and mask were digitally sampled and then filtered as shown. The cutoff frequency for the low-pass filter was 2.25 cycles/deg, and that for the high-pass filter was 4.5 cycles/deg. The cutoff frequencies for the bandpass filter were 2.25 and 4.25, which are largely in the region of greatest contrast sensitivity. Filtering was done with a finite impulse response filter, a digital filter using Hamming windowing techniques (Rzeszotarski et al., 1983).

Results and Discussion

We established duration thresholds for correctly reporting all letters in the filter set. A fixation field of 800 msec preceded the presentation of the unmasked letter in the Gerbrands Model GB 300 tachistoscope. For Subject G.C.G., thresholds were: unfiltered, 2; low pass, 2; high pass, 9; and bandpass, 3 msec. For Subject J.J.G., thresholds were: unfiltered, 2; low pass, 2; high pass, 6; and bandpass, 3 msec. Except for high-pass letters, the threshold functions are all the same.

Then we established critical durations for each filtered target and mask combination. Critical duration is the duration of a target that is necessary to enable the subject to escape the effects of masking produced by a mask presented for 20 msec immediately following the end of the target (i.e., with an interstimulus interval of zero). The duration was established by the modified method of ascending limits used by Walsh (1976). The criterion was perfect report of all 13 letters in sequence.

Because of the differences in thresholds for various filtering conditions, critical durations were converted to a maskability ratio by dividing the critical ratio by the threshold duration. The ratio is a factor by which target duration must be increased in the presence of a masking stimulus in order to bring performance to a level equivalent to the visibility threshold.

Figure 14 shows the results averaged for the two subjects. First, consider only the data for the unfiltered, low-pass and high-pass sets and masks. Here, the spectra of the filtered sets are clearly separated. In general, maskability of a set of letters is greatest when the mask contains the same frequency information. The unfiltered mask contains all of the information present in the other filtered versions of the mask, and it is an efficient masker in all cases. The weakest masking effect is from a spectrally different mask.

Now consider the data for the bandpass filtered set.

Figure 14. Mean maskability ratios of two subjects for unfiltered (original) low-pass, high-pass, and bandpass filtered letters masked by unfiltered and similarly filtered masks.

They reveal that the bandpass mask is an efficient masker across all of the filtered letter sets and that bandpass letters are, in turn, efficiently masked by all of the filtered masks. In constructing this set, our choice of cutoff frequencies was arbitrary. We wanted to filter out almost everything beyond the most sensitive region of the CSF, but we had to make the decision in the absence of any knowledge of the contribution of various parts of the spectrum to letter identification in masking. The data for the bandpass letters and mask suggest that our cuts were made in or near very critical regions. Because of the nature of impulse filters, there is some spectral overlap of the filters in the region of the cutoff frequencies. The overlap could explain the masking effects if this information is important to performance. It seems possible that setting cutoff frequencies for the high- and low-pass letters at (say) 1 and 6 cycles/deg might alter maskability functions considerably.

The maskability of the bandpass letters may be related to another issue, which is, why is there any maskability ratio greater than 1.0 when there is little or no spectral overlap? Letter recognition is not simply a low-level perceptual process. Remember that the mask duration is 20 msec and is last for processing. The cognitive task for the subject is to extract and make available for decisions the information in the earlier event. Attentional and strategy factors play a role in performance too. Some experimenters (e.g., Broadbent, 1977; Weisstein & Harris, 1980) have indicated that evidence points toward different roles of the low- and high-pass systems (transient and sustained channels) in attentional processes. One of our subjects, an expert psychophysical observer, had a ratio of .9 for high-pass letters masked by a low-pass mask. He felt the low-pass mask helped him a little by creating a ground against which the high-pass letters stood as figures, facilitating

attention. Very different configurations are present in this mask and target. By way of contrast with the observer's experience, Rzeszutarski et al. (1983) present a notch-filtered picture of Albert Einstein in which both low- and high-frequency information are simultaneously present. The effect is disturbing to observers who see the facial configuration of the low-pass image and a sort of "superimposed" outline drawing. There seems to be competition for attention between the two types of information in the same picture.

In the case of bandpass targets, the information lies in a transition zone between the spectral sensitivities of the two types of channels. Attentional strategies may be severely disruptable when the needed information lies only in this zone.

The experiment demonstrates that when the frequency content of stimuli and masks are controlled, masking functions differ. The results do not tell us enough about how masking occurs, but they certainly suggest that the control of available spatial frequency information through image processing is very likely to open the door to unraveling this complex interaction of primary visual processes and higher level cognitive ones.

In attempting to understand visual perception in masking and related phenomena, other manipulations of images are possible by use of the two-dimensional FFT that are not illustrated here. For example, specific types of noise can be added to images to provide degradation of the image when the noise is random within certain broad constraints of orientation and bandwidth. A perceptionist's typical method of stimulus degradation, in order to produce sufficiently high error rates to be able to study a perceptual phenomenon, is to greatly reduce the level of illumination or to greatly shorten the time of exposure. Shortening exposure time may artifactually limit the activity of certain channels because of their variable latencies and durations of response. Also, reducing the illumination level can selectively constrain the activity of certain spatial frequency channels since, as shown by the CSF, there is rapid roll-off in the transfer function at very low and very high spatial frequencies.

It is possible that the technology of image processing will lead to conclusions that our knowledge of visual performance obtained from tachistoscopes is limited only to restricted bands of visual channels and that differences in performance of various groups, such as the aged or schizophrenics, is not attributable to anything other than our artifactually biasing the available information so that it does not match the tuning of their filters. But if tuning is substantially different for some individuals, one prospective application of the technique would be to build compensating filters that selectively boost certain frequencies to contrast levels that would allow subjects to view an image as others do (assuming that the sensitivity for certain frequencies is only depressed, not totally absent).

POSTLOGUE

The auditory system has long been modeled as a sort of Fourier analyzer. The application of the model to vision has produced an enormous volume of research and has greatly advanced our knowledge of vision in recent years. The method of two-dimensional Fourier analysis may also have relevance to problems of tactile perception. Loomis (1981) showed that visual recognition functions for low-pass filtered braille characters can be taken as an approximation of tactile recognition functions for small raised characters. Loomis (1982) draws parallels between the neurophysiological evidence for sustained and transient channels in vision and slowly adapting and rapidly adapting mechanoreceptive afferents in touch.

REFERENCE NOTE

1. Ginsburg, A. P. *Visual information processing based on spatial filters constrained by biological data* (AMRL-TR-78, 129). Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 1978.

REFERENCES

- BANKS, M. S., & SALAPATEK, P. Infant pattern vision: A new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology*, 1981, **31**, 1-45.
- BREITMEYER, B. G. Unmasking visual masking: A look at the "why" behind the veil of "how." *Psychological Review*, 1980, **87**, 52-69.
- BREITMEYER, B. G., & GANZ, L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression and information processing. *Psychological Review*, 1976, **83**, 1-36.
- BROADBENT, D. E. The hidden preattentive process. *American Psychologist*, 1977, **32**, 109-118.
- DEVALOIS, R. L., & DEVALOIS, K. K. Spatial vision. *Annual Review of Psychology*, 1980, **31**, 309-341.
- GRAHAM, N. Psychophysics of spatial-frequency channels. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, 1981.
- LEVINSON, J. Z. Flicker fusion phenomena. *Science*, 1968, **160**, 21-28.
- LOOMIS, J. M. On the tangibility of letters and braille. *Perception & Psychophysics*, 1981, **29**, 37-46.
- LOOMIS, J. M. Analysis of tactile and visual confusion matrices. *Perception & Psychophysics*, 1982, **31** 41-52.
- MARR, D. *Vision*. San Francisco: Freeman, 1982.
- MATIN, L. Fourier treatment of some experiments in visual flicker. *Science*, 1962, **136**, 983-985.
- POMERANTZ, J. R. Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, 1981.
- RZESZOTARSKI, M. S., ROYER, F. L., & GILMORE, G. C. Introduction to two-dimensional Fourier analysis. *Behavior Research Methods & Instrumentation*, 1983, **15**, 308-318.
- SEKULER, R. Spatial vision. *Annual Review of Psychology*, 1974, **25**, 195-232.
- SEKULER, R., & HUTMAN, L. P. Spatial vision and aging. I: Contrast sensitivity. *Journal of Gerontology*, 1980, **35**, 692-699.
- VON GRUNAU, M. Interactions between sustained and transient channels: Form inhibits motion in the human visual system. *Vision Research*, 1978, **18**, 197-201.
- VON GRUNAU, M. The origin of pattern information of an apparently moving object during stroboscopic motion. *Perception & Psychophysics*, 1981, **34**, 357-361.
- WALSH, D. A. Age differences in central perceptual processing: A dichoptic backward masking investigation. *Journal of Gerontology*, 1976, **31**, 178-185.
- WEISSTEIN, N., & HARRIS, C. S. Masking and unmasking of distributed representation in the visual system. In C. S. Harris (Ed.), *Visual coding and adaptability*. Hillsdale, N.J.: Erlbaum, 1980.
- WEISSTEIN, N., HARRIS, C. S., BERBAUM, K., TANGNEY, J., & WILLIAMS, A. Contrast reduction by small localized stimuli: Extensive spatial spread of above-threshold orientation-selective masking. *Vision Research*, 1977, **17**, 341-350.
- WILSON, H. R., & BERGEN, J. R. A four mechanism model for threshold spatial vision. *Vision Research*, 1979, **19**, 19-32.