SIAM-AMS Proceedings
Volume 13
1981

# The Visual System Does a Crude Fourier Analysis of Patterns

NORMA GRAHAM

**Introduction.** About a dozen years ago in the Journal of Physiology, John Robson and Fergus Campbell introduced the notion that the human visual system contains multiple spatial-frequency channels–that is, multiple subsystems working in parallel, each of which is sensitive to a different range of spatial frequencies in visual patterns.

At about the same time, in Psychological Review, Jim Thomas made the closely related point that the existence of visual neurons with different sizes of receptive fields has important implications for pattern vision, and, in Science, Allan Pantle and Bob Sekuler suggested the existence of multiple size-selective channels. Since that time, a tremendous amount of psychophysical and physiological work has been inspired by this theoretical notion that there are multiple channels working in parallel to process visual patterns and that each of these channels is sensitive to a different, narrow band of spatial frequencies. Some people have gone so far as to say that the human visual system does a Fourier analysis of the visual scene.

What I do here is review the history of this multiple-channels model of pattern vision and comment on its current status. Some references will be given here, and the reader can find a more extensive bibliography in Graham [**1981**]. Some of the material here is explained more fully at an intuitive level in Graham [**1980**].

First let me point out that in discussing pattern vision, we ignore many dimensions important to vision–color, time, depth–discussing only monochromatic, unmoving, unchanging, flat patterns. We discuss only the initial visual processing, ignoring the higher-order perceptual or cognitive processes that occur, for example, in reading a pattern of letters on a page. In spite of this

---

1980 *Mathematics Subject Classification*. Primary 92A25.

extreme limitation, there would still be too much to cover, so I will further limit it by concentrating on places where mathematics has entered into the development of the multiple-channels model of pattern vision and on places where more mathematics might be useful. (The formal mathematics, which is not presented here in general, can be found in the references.)

**Early history.** The multiple spatial-frequency channels model developed rather naturally from an earlier model of pattern vision, a single-channel model. In the single-channel model, the important stage of the visual system is a linear system with a two-dimensional input representing the visual stimulus and a two-dimensional output representing the response of the visual system. This model was attractive both because it seemed consistent with known neurophysiology and because it was a very simple model. (People seem implicitly to assume linearity until they have evidence to the contrary.)

*Physiological receptive fields.* As was discovered a few decades ago, vertebrate retinal ganglion cells, which are the neurons in the retina that send their axons up to the brain, do not respond to single points of light. They respond instead to light in a rather broad area of the visual field. This area is called the "receptive field" of the neuron. Further, a neuron responds differently depending on where in its receptive field the light falls. For example, if light falls in the center of the receptive field, the neuron might respond with increased firing–that is, the neuron is excited. If light falls in an annulus surrounding the center of the field, the neuron responds with decreased firing, that is, the neuron is inhibited. (Some neurons have the opposite arrangement, receptive fields with inhibitory centers and excitatory surrounds.) Importantly, the response of these neurons to the light pattern is approximately linear. If two spots of light are shown, the response to the two is approximately equal to the sum of the responses to each alone. If both spots fall in the center of the receptive field, the neuron responds more to the combination than to either alone. If one spot falls in the center and one in the surround, the responses cancel.

Neurons in the visual cortex of vertebrates were later discovered to have even more complicated receptive fields. Some of these cortical cells are still linear systems with a central excitatory area and adjacent inhibitory areas, but now these areas are rectangular with one or two rectangular inhibitory areas adjacent to the long edge of the rectangular excitatory area. For a review of this physiology, see Robson [**1980**].

*Single-channel model.* Thus it seemed reasonable to model the important stage of the visual system as an array of many of these neurons. Their receptive fields were assumed to be heavily overlapping and densely distributed across the visual field. (See, for example, Ratliff's delightful book, *Mach Bands* [**1965**].) This conception is called a *Single-Channel Model.* If (1) each neuron responds linearly, (2) the output of this array is taken to be the two-dimensional function giving the response of each neuron as a function of the central position of the neuron's receptive field, and (3) the input to this array is taken to be the two-dimensional function giving light intensity at each point in the visual field, then this array is a linear system.

Behavioral as well as physiological evidence supported this model. For example, the perceived appearance of edges, that is, the existence of Mach bands, is consistent with such a model (Ratliff [1965]). I will not go further into this evidence, however, as we are about to discard this model, after we have given it credit for inspiring the use of a new stimulus with which to study vision.

*Sinusoidal gratings*. This new stimulus was a *sinusoidal grating*. Since the rather prevalent, rather reasonable model of the visual system was a linear system, people who knew about Fourier or linear systems analysis naturally thought of using sinusoidal inputs to the system (Bryngdahl [1962], DePalma and Lowry [1962], and Patel [1966]). What would the appropriate sinusoid be? One candidate, the candidate that was chosen, is a one-dimensional sinusoidal grating like that shown in Figure 1. A sinusoidal grating is a pattern in which the luminance in one direction varies sinusoidally, while the luminance in the perpendicular direction is constant. The *spatial frequency* of a sinusoidal grating is the number of cycles of the sinusoid per unit distance. The *mean luminance* of a grating is the average luminance across the whole grating and is generally kept constant. (One of the attractions of sinusoidal gratings is that the mean luminance of a grating can easily be held constant, keeping the observer in a relatively constant state of light adaptation and thus avoiding manifest early nonlinearities in the visual system, while the contrast and spatial frequency are varied.) The *contrast* of a grating is a measure of the amplitude of the sinusoid; it is usually taken to be one-half the difference between the maximum and minimum luminance divided by the mean luminance.

*Fourier analysis*. Although I suspect the following will be familiar to most readers, let me review briefly the relevant facts of Fourier analysis. Any two-dimensional function like that describing the luminance at each point in a visual pattern can be Fourier analyzed, that is, the (two-dimensional) Fourier transform of the function can be computed. Further, the Fourier transform can be inverted to give back the original function. Or, to put it in terms of visual patterns, any visual pattern (remember we are only talking about flat, unmoving, uncolored patterns) can be synthesized by adding together sinusoidal gratings of different frequencies and orientations in appropriate phases and contrasts. And there is only one set of sinusoidal gratings which will synthesize any given pattern.

Further, the response of a linear, translation-invariant system, like the single-channel model described above, to a sinusoidal grating is particularly simple. It is a sinusoid of the same frequency and orientation as the grating. Thus only its amplitude and phase need to be specified. The function specifying amplitude and phase for each frequency-orientation combination is known as the transfer function of the system.

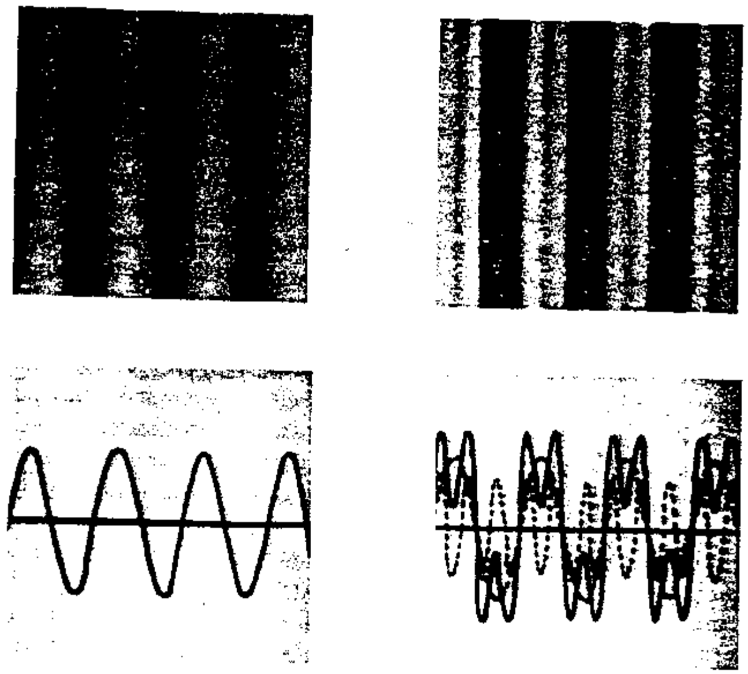

FIGURE 1. A simple sine-wave grating containing one spatial frequency is on the top left, and a compound grating containing two frequencies (one three times the other) is on the top right. The luminance profiles of the patterns are shown underneath each pattern as solid lines. (The dotted lines on the lower right show the luminance profiles of the individual component sine-waves.) From Graham and Nachmias [1971].

By definition, the response of a linear system to any stimulus which is the sum of components is equal to the sum of the responses to the components by themselves. Therefore, according to the linear, translation-invariant single-channel model described above, the response of the visual system to any pattern at all can easily be computed from its response to sinusoids as long as one knows its transfer function.

*Application to single-channel model.* Thus, if the simple single-channel model introduced above were correct, it would be quite easy to characterize all the parameters of pattern vision. One would just have to know the responses to all sinusoidal gratings (which are a small subset of all possible patterns) and one would know the response to all patterns. Since the response to any sine is a sine of same frequency and orientation, to know the responses to all sinusoidal gratings one would just have to know the amplitudes and phases of the responses.

How does an array of neurons with antagonistic excitatory and inhibitory areas in their receptive fields respond to gratings of different frequencies and orientations? It responds to a limited range of spatial frequencies, responding best to a medium frequency. It responds less well to high frequencies because the nonzero width of the excitatory center smears high frequencies. It responds less well to low frequencies because the inhibitory surround, which is wider than the excitatory center, depresses responses to low frequencies. If the array is made up of neurons with rectangular receptive fields like those in the cortex, it will also respond only to limited ranges of orientation. The orientation it will respond to best is the one where the bars of the gratings are parallel to the long dimension of the rectangular segments in the receptive fields. The phase of the response will depend on the symmetry of the receptive field. Even-symmetric receptive fields like those having equal-sized inhibitory flanks on either side of the excitatory center do not introduce any phase shift.

The simplicity introduced by this simple single-channel model is certainly appealing. But this simple single-channel model can easily be shown to be wrong.

**Sine-plus-sine experiments.** I am going to describe an experiment that is similar to the original experiment of Campbell and Robson, but allows more telling comparisons between experimental results and theoretical predictions (Graham and Nachmias [1971]). Four kinds of patterns were used, all of which were patterns varying in one direction only. See Figure 2, left column, for these patterns' luminance profiles. (A luminance profile is a one-dimensional cut in the interesting direction through the two-dimensional function giving light intensity at each spatial position.) The four patterns were: two simple sinusoidal gratings, one of frequency three times the other, and two compound gratings, each containing both frequencies but in different phases. (Photographs of one of these simple and one of these compound gratings are shown in Figure 1.) For each pattern, the detection threshold (that is, the contrast at which an observer can just tell that a pattern is present rather than a blank field of the same mean luminance) was measured.

*Single-channel model predictions.* What would the single-channel model predict for this experiment? The middle column of Figure 2 shows the one-dimensional cuts through the two-dimensional output of the channels. These are called response profiles. The response of the channel to the compound grating will just equal the response to the components (in the appropriate phase). To derive the predictions for an observer's threshold, some assumption must be made linking the responses of the model to the behavior of the observer. Here for ease of explanation I will make the simplest assumption. (Many others have been considered over the years, but none has rescued the single-channel model. If one made a sufficiently complicated linking hypothesis, one could undoubtedly rescue the model, but then the interesting part of the model would be in the linking hypothesis, not in the single channel.) We will here assume that an observer detects a pattern whenever the peak-trough difference in the response of the channel (the difference between the largest and the smallest values) reaches some criterion.

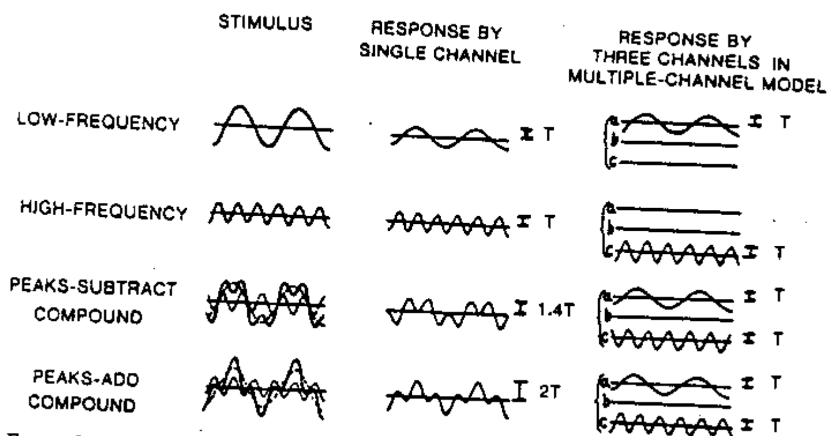


FIGURE 2. Four grating patterns are indicated by their luminance profiles in the left column. (The broken lines show the luminance profiles of the individual components of the compound gratings.) The responses predicted by a single-channel model are shown in the middle column. The responses predicted by three channels of a multiple-channels model are shown in the right column. From Graham and Nachmias [1971].

If each component sine-wave's contrast has been set so that it is just at threshold (as illustrated in the first two rows of Figure 2), the single-channel model predicts that the peak-trough difference in the response to each of the compounds will be far above threshold. For the peaks-subtract phase, the peak-trough difference in the response is 1.4 times threshold and for the peaks-add phase 2.0 times threshold. Thus, if the single-channel model were correct, these two compound gratings should be much more visible than their sinusoidal components and the peaks-add much more visible than the peaks-subtract.

For human observers, however, all four patterns are, to a first approximation, equally detectable. This fact is inconsistent with the single-channel model.

*Nonuniform single-channel model.* A possible variant of the single-channel model that immediately jumps into the mind of many a person is a model in which the size of receptive field changes as one moves from the foveal center to the periphery (although there is still only one size of receptive field at each location). Such a change fits in well with all the phenomena showing that our acuity is better in the middle of the visual field than in the periphery. This modified single-channel model is also wrong, however. A more recent version (Graham, Robson and Nachmias [1978]) of the sine-plus-sine experiment described above was done with small patches of grating. (These patches had slow transitions at their edges to avoid introducing too many other spatial frequencies and other edge effects.) These patches were small enough that, according to a nonuniform single-channel model in which receptive field size changes with position at a rate consistent with other visual data, all the patches would have been detected by the same size of receptive field; therefore, the compound would have been more detectable than the components, and the peaks-add compound would have been more detectable than the peaks-subtract. The experimental results, however, were the same as for the full grating, thus ruling out the nonuniform single-channel model.

*Multiple-channels model predictions*. The experimental results are, however, consistent with a multiple spatial-frequency channels model. Each of the multiple channels might itself be an array of receptive fields all of the same size. Different channels have different sizes of receptive field, a low spatial-frequency channel having larger receptive fields than a high spatial-frequency channel. As far as any individual channel is concerned, the compound grating behaves just like one of its components. For example, in the right column of Figure 2, the channel which is sensitive only to the low-frequency component (channel a) responds to the compound just as if it contained only the low-frequency component. If we assume that a pattern is detectable whenever the response of at least one channel is great enough, and if we continue to ignore, as we have been, the possibility of variability from trial to trial, then this multiple-channels model predicts that all four of the gratings in Figure 2 are just at threshold. In general, a compound grating is exactly as detectable as its most detectable component and phase does not matter. To a first approximation, this is what is found.

*Probability summation*. But it is only a first approximation. The compound gratings are both, in fact, just a little bit more detectable than their components–not nearly as much more as a single-channel model predicts, but a little. This added detectability came as no surprise to a psychologist, because a favorite notion of psychologists is "probability summation". In all psychophysical experiments, there is a great deal of variability. An observer's response to a particular stimulus is not always the same. A pattern near threshold just does not look the same from trial to trial. Sometimes an observer will see the pattern very clearly; sometimes he will not see it at all. Suppose this variability is due to variability in the responses of the channels, and the variability in different channels is independent. Then, the two channels that respond to a compound grating each have independent chances of detecting the compound grating, but only one of the two channels has a chance to detect a component by itself. Therefore, the compound gratings should be slightly more detectable than either component. A multiple-channels model with probability summation among the channels quantitatively accounts for the exact thresholds of the compound gratings (Sachs, Nachmias, and Robson [1971]; Graham, Robson, and Nachmias [1978]).

The existence of this probability summation, which is a form of nonlinear summation between channels, greatly complicates calculations from a multiple-channels model.

To further complicate the situation, recent evidence carefully varying the number of bars in a grating (Robson and Graham [1981]) strongly suggests there is also probability summation across the spatial extent of each channel. Or, to put it in terms of possible physiology, if the channels are conceived of as arrays of receptive fields, then each receptive field has its own independent variability. Thus, to make predictions from the multiple-channels model, one also needs to take into account this nonlinear probability summation across spatial extent.

**Quick pooling model.** In order to calculate the predictions of a multiple-channels model with probability summation, the variability in individual receptive fields or channels was originally assumed to be described by a Gaussian distribution. With this assumption, it was extremely tedious to do the calculations of a pattern's threshold from the responses of the multiple receptive fields or multiple channels. Fortunately, a few years ago, Frank Quick **[1974]** pointed out the existence of a function, known in some contexts as the Weibull function, which is a good approximation to the cumulative Gaussian but is much easier to work with as long as it is embedded in a certain psychophysical model.

Let $P_{ij}$ be the probability that the $i$th receptive field in the $j$th channel detects the stimulus (that is, the probability that the response of that receptive field is greater than some criterion). Let $S_{ij}$ be the sensitivity of that receptive field to the stimulus (that is, the reciprocal of the contrast necessary to produce detection by the receptive field on half of the trials). Let $K$ be a parameter determining the steepness of the function. Let $c$ be the contrast in the stimulus pattern. Then the function suggested by Quick is

$$P_{ij} = 1 - 2^{-(c \cdot S_{ij})^k}.$$

If one assumes that the observer detects a pattern if and only if at least one of the receptive fields in one of the channels does, then, letting $P$ be the probability of the observer's detecting the pattern, letting $M$ be the number of receptive fields in each channel, and letting $N$ be the number of channels,

$$P = 1 - \prod_{i,j}^{M,N} (1 - P_{ij}).$$

We still need to specify how the observer's probability of detecting a pattern determines his response in an experiment. We will make the simple-minded assumption that, when he does detect the pattern, he always responds correctly and when he does not detect the pattern, he simply guesses. This assumption, sometimes called the high-threshold model of response bias, is known to be wrong (see Green and Swets **[1974]** for example), but perhaps it is a good enough approximation to the truth to be used. (In any case, when it is embedded within the whole model, the resulting predictions seem to account for many results, as will be described below.)

By substituting Quick's psychometric function into the expression for $P$, the observer's probability of detection, one finds by easy algebraic manipulations that $P = 1 - 2^{-(c \cdot S)^k}$ where $S$ is the sensitivity of the observer to the stimulus (the reciprocal of the contrast necessary for the observer to detect the pattern on half of the trials), and can easily be expressed in terms of the sensitivities of the underlying units

$$S = \left[ \sum_{i,j}^{N,M} S_{ij}^{k} \right]^{1/k} .$$

Notice that we have made the implicit assumption that $k$ has the same value for each receptive field, that is the variability in response magnitude is the same for every receptive field. Under this assumption, the function specifying the observer's probability of detection (the so-called psychometric function) has the same form for every stimulus and also the same form as the function specifying a single receptive field's probability of detection. Green and Luce [1975] proved that functions of the above form (the base need not be 2) are the only ones having this invariance property.

Empirically, a value for $k$ of about 3.5 seems to describe the psychometric function for a wide variety of stimuli (e.g. Graham, Robson and Nachmias [1978], Legge [1978], Robson and Graham [1981], Watson [1979]).

To remind you, the sensitivity of each receptive field $S_{ij}$ is just that of a linear system with the appropriate weighting function (that is, having the size and shape that go with a particular channel's frequency and orientation frequency domain). (Since they are linear systems, the sensitivity of each one to a particular stimulus multiplied by the contrast in the stimulus gives you the response magnitude. Therefore, sensitivity is the same as the response magnitude to a criterion contrast.) In experiments to measure the thresholds of compound gratings containing far-apart frequencies (as in Figure 2) each channel only responds to one frequency so there is no need to actually know the weighting function.

The above expression is easily recognized as a metric in a space where each dimension represents the sensitivity of one receptive field. The sensitivity to a stimulus is just the distance between the origin (which represents the case of no pattern, that is, a blank, homogeneous field) and the point representing that stimulus. (One might extend this approach and let the distance between any two points in the space represent the discriminability of any two stimuli, but almost no work has been done on such an extension so I will not mention it further.) When $k$ is infinity, the observer's sensitivity equals the sensitivity of the most sensitive unit (which makes this version of the above expression equivalent to the model without any probability summation). When $k$ is 1, the sensitivities or responses of different units are just being linearly summed. When $k$ is 2, there is power-summation (which is a model that occurs in many contexts but is false in this context). And when $k$ is about 3.5, this expression quantitatively predicts the thresholds for a wide variety of patterns.

Since this last is an important point, let me elaborate on it a minute. The above expression, with a value for $k$ of about 3.5, produces predictions that agree very closely with the thresholds for a wide variety of patterns including combinations of sines with sines (e.g. Graham, Robson and Nachmias [**1978**], Quick, Mullins and Reichert [**1978**]), patches of sines with various numbers of cycles (Legge [**1978**], Robson and Graham [**1981**]), and some aperiodic stimuli as well as combinations of aperiodic stimuli with sines (e.g. Graham [**1977**]; Bergen, Wilson and Cowan [**1979**]). For the latter predictions, one does need to know the weighting functions. More of that later. Further, one can expand one's perspective and include time as a dimension (varying the length of time a pattern is on, or its temporal frequency content) and this same approach will work (Watson [**1979**], Watson and Nachmias [**1977**]).

In short, a model in which there is probability summation across space (across the receptive fields within each channel) and among channels and in which the variability is described by the function suggested by Quick with a steepness parameter of 3.5 accounts both for the psychometric functions that have been collected for a number of stimuli and for the actual threshold values that have been measured for an even wider collection of stimuli. This is a very impressive feat. However, caution is still in order. The nonlinear pooling across receptive fields and channels exhibited in the equation for $S$ that does such a good job at accounting for thresholds may not actually be due to probability summation (independent noise) but to some other cause, and the agreement between psychometric function steepness parameter and the exponent needed to account for thresholds may be just fortuitious. We know that, in detail, the psychophysical high-threshold theory must be wrong. For discussion of these issues, see Robson and Graham [**1981**].

In any case, the metric in the equation for $S$ has proved its usefulness and has given us insight into the kind of pooling across receptive fields and channels that must be assumed in order to predict thresholds.

Let us now go back to the issue of what the weighting functions for individual receptive fields look like or, to put in another way, what the spatial-frequency and orientation tuning of individual channels is like. Or, to put it still another way, let us go back to the issue of whether the visual system actually does a Fourier analysis of the visual scene.

**Strict Fourier analysis?** What would it mean if we said that the brain performed a Fourier analysis of the visual scene? We might mean, if we were speaking strictly, that there was a set of neurons that computed the Fourier transform of the visual pattern. The magnitude of the response of a particular neuron in the set would be *completely determined* by the amount (or by the phase) of a particular spatial-frequency/orientation component present in the pattern. In other words, each neuron in the set would respond only to an *extremely narrow* range of spatial frequencies and orientations. Different neurons would respond to different spatial-frequency/orientation combinations so the set as a whole would compute an excellent approximation to a Fourier transform. (It is an approximation because there are only a finite number of neurons and the Fourier transform is a continuous function, but this kind of approximation has no practical consequences.)

There is no such set of neurons anywhere in the brain; at least, there is absolutely no evidence, either physiological or psychophysical, that such a set of neurons exists. The brain, therefore, does not perform a *strict* Fourier analysis of the visual scene. Further, I should say as a historical note, that none of the people (Campbell, Robson, Thomas, Pantle, or Sekuler) mentioned above meant to be implying that the visual system did a strict Fourier analysis, although they, particularly Campbell and Robson, have sometimes been blamed (or credited) with so doing.

Although there is no evidence that the brain performs a strict Fourier analysis, there is an accumulating mass of evidence that the brain performs operations with many of the characteristics of Fourier analysis, operations that could be called crude Fourier analyses.

**Evidence for a crude Fourier analysis.** When the thresholds for compound gratings containing component frequencies that are close together are properly interpreted (that is, probability summation across space is allowed for, see references given above), they suggest a bandwidth for each channel on the order of $1\frac{1}{2}$ or 2 octaves. Channels of this medium bandwidth could not be said to do a strict Fourier analysis, but nonetheless, the response magnitude of each channel does give you a good idea of how much of the channel's preferred spatial frequency is present in the pattern. So one might want to say that a set of channels with this sort of medium bandwidth does a crude Fourier analysis. Sine-plus-sine experiments are only one kind of evidence, however. Let us look now, very briefly, at the other kinds of evidence that have supported the notion of multiple spatial-frequency channels and at the bandwidths suggested by each kind of evidence.

In *Adaptation and Masking Experiments*, the visibility of test patterns is measured either after the observer has inspected a suprathreshold adapting pattern, or while he is inspecting a masking pattern. The visibility of the test pattern should be affected by the adapting (or masking) pattern if the patterns are processed by the same channel, but not if they are processed by different channels. In fact, when the test pattern and adapting (or masking) pattern contain similar spatial frequencies, the test pattern is affected. But when the test pattern and adapting (or masking) pattern contain very different spatial frequencies or orientations, the test pattern is not affected. These effects occur both at threshold and on some aspects of suprathreshold perceived appearance. That is, the threshold for a test grating, the perceived contrast in a suprathreshold test grating, and the perceived frequency and orientation of a suprathreshold test grating are all altered by previous adaptation to or simultaneous masking by a grating of similar spatial-frequency and orientation. How similar? an octave or so.

In addition to the wide variety of adaptation and masking effects that have been studied there are some less well-studied, higher-order perceptual effects that seem to reveal the action of the spatial-frequency channels.

In *Recognition Experiments*, the observer is asked to say not only whether some pattern is present but what that pattern is (out of some small set of possible patterns). Far-apart frequencies are recognized whenever they are detected (Nachmias and Weber [1975]) as if there were spatial-frequency channels and the observer could always tell which of the channels had detected the stimulus. Closer-together frequencies, however, begin to be confused. How close together they need to be to produce confusion may tell something about the bandwidth of channels. The available results suggest medium bandwidths (Hirsch [1977], Thomas and Barker [1977]).

As an aside off the topic of bandwidth, let me tell you two other interesting recognition results from experiments where compound gratings were used. First, compound gratings containing two frequencies in a ratio of three to one can be far enough above threshold that an observer can always detect both components (can always tell you that both are present rather than just one) and yet be unable to tell you which of two phases the components are in (Nachmias and Weber [1975]). This lack of phase discrimination is easy to explain if spatial-frequency channels exist so that the two components are exciting separate channels, and, near threshold, these channels signal nothing about phase so the relative phase of the two components cannot be computed. The second interesting result is that at threshold contrasts, the compound grating is sometimes seen as a compound, sometimes seen as one of its components, sometimes seen as the other component, and sometimes seen as a blank (Hirsch [1977]). To a first approximation anyway, the proportions of times it is seen as each one is what you would expect if there were independent variability in each channel (probability summation).

In *Texture Segregation Experiments* (also called grouping or effortless texture discrimination experiments), the observer is asked whether or not he can immediately see that there are two different areas in the pattern that contain two different textures. Julesz originally conjectured that an observer could only see immediately (without scrutiny) that the textures in the two different areas were different if the second-order statistics of the two textures differed. Different second-order statistics implies different autocorrelation functions and hence different amplitude spectra in the Fourier transforms. If there were extremely narrow-bandwidth channels, i.e., if the visual system did a strict Fourier analysis, throwing out the phase information, Julesz's conjecture would amount to saying that textures could only be discriminated when the outputs of these channels were different. Julesz, however, keeps discovering texture pairs for which his conjecture is wrong. Although they have identical second-order statistics and hence identical Fourier amplitudes, they are discriminable. He explains these by postulating a second class of visual mechanism in addition to the one that computes second-order statistics (the one equivalent to extremely narrowband channels). See Julesz [1981] for a review of this work. I think it might be better if he did not postulate a second class of visual mechanism but just said that the channels were not extremely narrowband but somewhat broader, as all the other evidence indicates.

*Physiology*. Receptive fields come in different sizes even at the retinal ganglion cell level where they are concentric. Thus they respond to different ranges of spatial frequencies, although each responds to quite a broad range. In the cortex, recent work (Movshon, Thompson and Tolhurst [1978], and DeValois, Albrecht and Thorell [1977]) show neurons that respond to very different ranges of spatial frequencies. The bandwidth of these neurons is limited but certainly not extremely narrow. The available physiological evidence, therefore, if you are willing to identify the psychophysical channels with the neurons in the geniculo-cortical pathway, also suggests medium bandwidth channels.

**The role of mathematics.**

*Has mathematics been useful in this work*? Historically, knowledge of the theorems about Fourier transforms and linear systems suggested the use of sinusoidal gratings, and the use of sinusoidal gratings has certainly added to our knowledge of the visual system. In particular, the results of experiments using gratings were a major factor in suggesting the multiple spatial-frequency channels model which is now so popular.

If these channels were extremely narrowband, the mathematics of Fourier analysis would have been a very concise, natural description of the system, expressing something fundamental about the way the visual system processed information. The decomposition of a stimulus into a basis set of sinusoidal components by Fourier analysis would have corresponded very closely to a decomposition done by the visual system. Each sinusoidal component would have excited one and only one channel. But the channels are not extremely narrowband.

Even with medium-bandwidth channels, one can often have better intuitions about how the channels will respond by considering the Fourier transforms of the stimuli rather than the stimuli themselves and by considering the transfer functions in the frequency domain of the theoretical channels rather than the weighting functions in the space domain. Formally, one can often calculate the predictions more easily using Fourier transforms (e.g. the calculations for aperiodic stimuli and combinations of sines and aperiodic stimuli in Graham [1977]). In this case, however, the mathematics of Fourier analysis is primarily serving as a calculating tool rather than as a natural expression of something fundamental about the visual system.

There is another, more subtle, way in which Fourier analysis may be contributing to our understanding of the visual system. Fourier transforms of stimuli may be enlightening simply as a new description of visual stimuli. Independently of the way the visual system actually works, a new description of visual stimuli may stimulate the human investigators of the visual system and provoke them into creative thought, particularly when the new description is as different from the old description as Fourier transforms are from the original function. In a sense, spatial-frequency descriptions (Fourier transforms) and point-wise descriptions (intensity as a function of spatial position) are opposites. A stimulus consisting of a single spatial frequency is completely localized on the frequency dimension, but infinitely extended in space, whereas a stimulus consisting of a single point is completely localized in space but infinitely extended on the frequency dimension. These two descriptions emphasize very different aspects of the stimulus.

In addition to Fourier analysis, another sort of mathematics seems to have been useful in the work of understanding pattern vision, particularly the thresholds for the detection of patterns. That is the metric expression given above for nonlinear pooling (perhaps probability summation) across receptive fields and channels and the Weibull function as a description of psychometric functions. Again, this mathematics seems primarily to have served as a calculating tool rather than an expression of something fundamental. It does seem, however, to have allowed much greater insight into the possible effects of nonlinearities like those involved in pooling across units.

**What would be nice in the way of mathematics?**

*A substitute for sinusoidal gratings and Fourier analysis.* It would be useful to have some mathematics that was as natural a representation of a system involving medium-bandwidth channels (channels with receptive fields that are quite well localized in the visual field and only contain a few excitatory or inhibitory sections) as Fourier analysis is of extremely narrowband channels (channels with a large number of excitatory and inhibitory sections in receptive fields extended across the visual field). There is probably no representation quite as nice as to provide a basis set of stimuli into which any stimulus can be uniquely decomposed and which correspond precisely to the sensitivities of the channels (so that each stimulus in the basis set stimulates one and only one channel). But perhaps some representation exists which is better suited to and thus provides more insight into the case of medium-bandwidth channels. Such a representation might provide a new kind of stimulus (perhaps little patches of something like a sinusoidal grating) which was the optimal kind of stimulus for studying a system of medium-bandwidth channels and might also provide a way of dealing with the responses to those optimal stimuli.

*More ways to investigate and describe nonlinearities.* Even though the channels are not extremely narrowband, they would be moderately easy to deal with as long as the channels were themselves linear systems and one could examine the properties of one channel at a time. Such, however, is not the case, particularly not for behavioral (psychophysical) phenomena but also not always for physiological data. Thus, investigators of pattern vision, like so many other people, need more flexible and insightful and convenient ways of handling nonlinear systems.

*Psychophysics*. We have already discussed above the fact that probability summation is necessary to accurately account for the thresholds of patterns. That is, the threshold for a pattern is not determined by the one channel that is most sensitive to the pattern (the one channel that gives, on the average, the largest response to the pattern), but the threshold is always determined by nonlinear pooling of the sensitivities of all the channels that respond at all to the pattern. (Of course, channels that are very insensitive contribute negligibly, but all channels that are almost as sensitive as the most sensitive channel contribute substantially.) Thus, one is always looking at the action of a group of channels rather than a single channel.

Further, even if each receptive field of the channels operative in the pschophysical phenomena does turn out to be a linear system, as is currently assumed, the output of a channel (the output of a whole array of receptive fields of the same size) seems to be the resultant of probability summation across space or some similar kind of nonlinear pooling. So the response of a channel cannot be taken to be the response of a completely linear system.

Although the Quick pooling model presented above does a very good job at accounting for threshold data, we know it cannot be completely correct.

*Physiology*. If we limit our interest to single neurons, then it is easy to look at only one at a time. Then the only question is whether they are linear or not. Well, many are, or at least linear enough that Fourier analysis works, even in the cortex. But many are not. At all levels in the visual system, a major distinction is being made now between $x$ and $y$, or sustained and transient, or linear and nonlinear cells. The nonlinear cells are, of course, much more difficult to figure out. Some progress is being made using newly-developed versions of Wiener analysis (Shapely and Victor [1979], Victor and Knight [1979]).

## References


Bergen, J. R., H. R. Wilson and J. D. Cowan, 1979. *Further evidence for four mechanisms mediating vision at threshold: Sensitivities to complex gratings and aperiodic stimuli*, J. Opt. Soc. Amer. **69**, 1580–1587.

Bryngdahl, O., 1964. *Characteristics of the visual system: Psychophysical measurements of the response to spatial sine-wave stimuli in the mesopic region*, J. Opt. Soc. Amer. **54**, 1152–1160.

Campbell, F. W. and J. G. Robson, 1968. *Application of Fourier analysis to the visibility of gratings*, J. Physiol. **197**, 551–566.

DePalma, J. J. and E. M. Lowry, 1962. *Sine-wave response of the visual system. II. Sine-wave and square-wave contrast sensitivity*, J. Opt. Soc. Amer. **52**, 328–335.

DeValois, R., D. G. Albrecht and L. Thorell, 1977. *Spatial contrast* (H. Spekreijse and L. H. van der Tweel, eds.), North-Holland, Amsterdam, pp. 60–63.

Graham, N., 1977. *Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels*, Vision Res. 17, 637–652.

———, 1980. *Spatial frequency channels in human vision: Detecting edges without edge detectors*, Visual Coding and Adaptability (C. Harris, ed.), Erlbaum, Hillsdale, N. J.

———, 1981. *Psychophysics of spatial-frequency channels*, Perceptual Organization (M. Kubovy and J. Pomerantz, eds.), Erlbaum, Hillsdale, N. J.

Graham, N. and J. Nachmias, 1971. *Detection of grating patterns containing two spatial frequencies: a test of single-channel and multiple-channels models*, Vision Res. **11**, 251–259.

Graham, N., J. G. Robson and J. Nachmias, 1978. *Grating summation in fovea and periphery*, Vision Res. **18**, 816–825.

Green, D. M. and J. A. Swets, 1974. *Signal detection theory and psychophysics*, Kreiger, Huntington, N. Y.

Green, D. M. and R. D. Luce, 1975. *Parallel psychometric functions from a set of independent detectors*, Psychological Rev. **82**, 483–486.

Hirsch, J., 1977. *Properties of human visual spatial-frequency-selective systems: a two-frequency, two-response recognition paradigm*. Doctoral Dissertation, Columbia University.

Julesz, B., 1981. *Figure and ground perception in briefly presented iso-dipole textures*, Perceptual Organization (M. Kubovy and J. Pomerantz, eds.), Erlbaum, Hillsdale, N. J.

Legge, G., 1978. *Space domain properties of a spatial-frequency channel in human vision*, Vision Res. **18**, 959–969.

Movshon, J. A., I. D. Thompson and D. J. Tolhurst, 1978. *Spatial and temporal tuning properties of cells in areas 17 and 18 of the cat's visual cortex*, J. Physiology **283**, 100–120.

Nachmias, J. and A. Weber, 1975. *Discrimination of simple and complex gratings*, Vision Res. **15**, 217–224.

Quick, R. F., 1974. *A vector-magnitude model of contrast detection*, Kybernetik **16**, 65–67.

Quick, R. F., W. W. Mullins and T. A. Reichert, 1978. *Spatial summation effects on two-component grating thresholds*, J. Opt. Soc. Amer. **68**, 116–121.

Pantle, A. and R. Sekuler, 1968. *Size-detecting mechanisms in human vision*, Science **162**, 1146–1148.

Patel, A. S., 1966. *Spatial resolution by the human visual system: The effect of mean retinal illuminance*, J. Opt. Soc. Amer. **56**, 689–694.

Ratliff, F., 1965. *Mach bands: quantitative studies on neural networks in the retina*, Holden-Day, San Francisco, Calif.

Robson, J. G., 1980. *Neural images: The physiological basis of spatial vision*, Visual Coding and Adaptability (C. S. Harris, ed.), Erlbaum, Hillsdale, N. J.

Robson, J. G. and N. Graham, 1981. *Probability summation and regional variation in contrast sensitivity across the visual field*, Vision Res. (to appear).

Sachs, M. B., J. Nachmias and J. G. Robson, 1971. *Spatial-frequency channels in human vision*, J. Opt. Soc. Amer. **61**, 1176–1186.

Shapley, R. M. and J. D. Victor, (1979). *Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells*, J. Physiology **290**, 141–161.

Thomas, J. P., 1970. *Model of the function of receptive fields in human vision*, Psychological Rev. **77**, 121–134.

Thomas, J. P. and R. A. Barker, 1977. *Bandwidths of visual channels estimated from detection and discrimination data*, J. Opt. Soc. Amer. **67**, 140, Abstract.

Victor, J. D. and B. W. Knight, 1979. *Nonlinear analysis with an arbitrary stimulus ensemble*, Quart. Appl. Math. **37**, 113–136.

Watson, A. B., 1979. *Probability summation over time*, Vision Res. **19**, 515–522.

Watson, A. B. and J. Nachmias, 1977. *Patterns of temporal interaction in the detection of gratings*, Vision Res. **17**, 893–902.

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, NEW YORK 10027