Principles and Practice of Clinical Electrophysiology of Vision

Editors

John R. Heckenlively, M.D.
Professor of Ophthalmology
Jules Stein Eye Institute
Los Angeles, California

Geoffrey B. Arden, M.D., Ph.D.
Professor of Ophthalmology and
Neurophysiology
Institute of Ophthalmology
Moorfields Eye Hospital
London, England

Associate Editors

Emiko Adachi-Usami, M.D.
Professor of Ophthalmology
Chiba University School of Medicine
Chiba, Japan

G.F.A. Harding, Ph.D.
Professor of Neurosciences
Department of Vision Sciences
Aston University
Birmingham, England

Sven Erik Nilsson, M.D., Ph.D.
Professor of Ophthalmology
University of Linköping
Linköping, Sweden

Richard G. Weleber, M.D.
Professor of Ophthalmology
University of Oregon Health Science Center
Portland, Oregon

Mosby Year Book

St. Louis Baltimore Boston Chicago London Philadelphia Sydney Toronto

Mosby Year Book

Dedicated to Publishing Excellence

Sponsoring Editor: David K. Marshall
Assistant Director, Manuscript Services: Frances M. Perveiler
Production Project Coordinator: Karen E. Halm
Proofroom Manager: Barbara Kelly

Mosby-Year Book, Inc.
11830 Westline Industrial Drive
St. Louis, MO 63146

1 2 3 4 5 6 7 8 9 0 CL CL MV 95 94 93 92 91

Library of Congress Cataloging-in-Publication Data
Principles and practice of visual electrophysiology / [edited by]
John R. Heckenlively, Geoffrey B. Arden.
p. cm.
Includes bibliographical references.
Includes index.
ISBN 0-8151-4290-0
1. Electroretinography. 2. Electrooculography. 3. Visual evoked response. I. Heckenlively, John R. II. Arden, Geoffrey B. (Geoffrey Bernard)
[DNLM: 1. Electrooculography. 2. Electrophysiology. 3. Electroretinography. 4. Evoked Potentials, Visual. 5. Vision Disorders—physiopathology. WW 270 P957]
RE79.E4P75 1991 91-13378
617.7 1547—dc20 CIP
DNLM/DLC
for Library of Congress

PART VI

Data Acquisition and Analysis

29

Analytical Techniques

L. Henk van der Tweel
Oscar Estévez

Visual electrophysiologists have at present a wide choice of instruments at their disposal, from direct recorders to sophisticated computers. This enables them to record ever smaller responses and to improve their quality. "Thresholds," defined as the weakest stimuli evoking recognizable responses, are continuously dropping, and the range and type of electroretinogram (ERG) and visual evoked potential (VEP) stimuli have been regularly extended. In addition, computer-based analytical methods are increasingly being used for the characterization of responses.

A proper selection from these modern methods requires knowledge about the principles of signal analysis on which they are based. These same principles apply to many quantitative aspects of visual function. The present chapter is meant to help the researcher and the clinician to find their way among the multitude of published methods. Emphasis will be laid less on mathematical rigor than on the understanding of fundamental concepts. The topics to be covered are (1) the recording and processing of electrical responses and (2) analytical questions concerning the stimulus and response characterization.

The unavoidable presence of noise (e.g., the background electroencephalogram [EEG]) demands procedures for noise reduction, especially if weak stimuli are presented. This can be done in a variety of ways, but in clinical practice it is often most important to reduce the recording time, which has encouraged the use of more "efficient" methods such as, for instance, the so-called steady-state stimulation. It is intended that the present chapter will enable the evaluation of such techniques.

For a long time nearly all ERGs and VEPs were recorded with flashes. The responses obtained in this way are often complex and more prone to the effects of strong nonlinearities (for a definition of this term, see below). However, by employing stimuli with other waveforms, among which sinusoidal modulation is the most frequent, it is easier to recognize deviations from linearity and to identify significant nonlinear properties of the system under study. An added advantage of sinusoidal stimuli is that linearity can often be approximated to a satisfactory degree, which facilitates analysis and description.

More recently homogeneous field stimulation has been superseded by the use of spatially structured fields such as checkerboards and sine wave gratings. This type of research has developed in two main directions:

1. Characteristics such as amplitude, wave shape, and latency are used to discriminate between normal and pathological responses. In this case only rather elementary methods of signal improvement need be employed.
2. The responses are used as a criterion for the "effectiveness" of a changing stimulus, e.g., when the size of checks in a checkerboard or the periodicity and the contrast of a grating are manipulated to obtain a constant response. The results of such studies are often expressed as a contrast sensitivity function (transfer function). The theoretical background, however, is complicated and requires among other things an analytical characterization of the stimulus, e.g., that of a checkerboard.

BASIC CONCEPTS OF SIGNAL ANALYSIS

Linearity

One approach to analyzing the results of clinical electrophysiology is to treat all stimuli as if they were transduced and processed by a "black box" between the stimulus and recorded response. The simplest assumption that can be made about the black box is that of linearity, even though no biological system strictly fulfills such a condition.

The definition of a linear system is that it obeys the "superposition principle." Assume that A and B are (quantifiable) input signals (e.g., stimuli) that result respectively in outputs C and D of the system under study. Let us use $\rightarrow$ to signify "produces the response" so that we have $A \rightarrow C$ and $B \rightarrow D$. In this case we say that a system is linear if $A + B \rightarrow C + D$. A consequence of linearity is that the amplitude of the response is strictly proportional to that of the input. Input and output do not need to belong to the same physical category; they may, for instance, represent light values and voltage respectively as in Figure 29–1, where the upper traces represent the voltage of the ERG of an anesthetized cat and the lower ones the modulation of the intense light source that was employed.[17]

A cardinal property of linear systems concerns the harmonic function $A \sin 2\pi ft$ (Fig 29–2). It is the only function that retains its identity (its sinusoidal shape and its period) when submitted to a linear transformation; a light input $A \sin 2\pi ft$ will, for instance, evoke a response $B \sin (2\pi ft + \Phi)$ with, in general, B different from A and Φ an added phase shift. A direct consequence of this property is that, at the output of a linear system, no new frequencies are generated from any input, no matter how complex that input is.

In physics, basic (passive) elements like capacities, inductances, resistances, and their combinations are the simplest linear elements; however, complex (active) devices like electronic amplifiers can also approach linearity to a large degree.

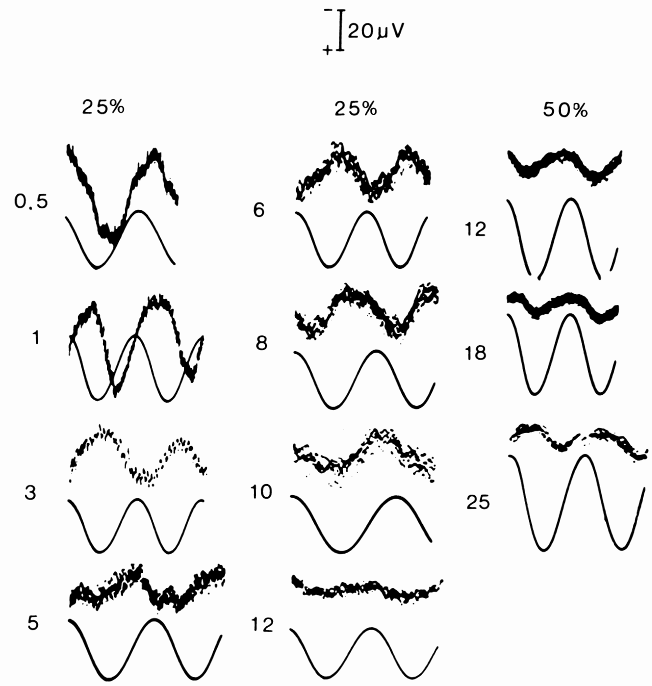

FIG 29–1.
An early example of the responses of a reasonably linear "black box" in which light is translated into voltage. The ERG of an anesthetized cat is recorded with sinusoidally modulated light (retinal illumination, 50,000 trolands; field, 56 degrees; 25% respectively; 50% modulation depth; range, 0.5 to 25 Hz; *upper* traces, ERG; *lower* traces, photocell signal of illumination). Responses approach the sinusoidal shape, a property of linear systems. (From Van der Tweel LH, Visser P: Electrical responses of the retina to sinusoidally modulated light, in *Electroretinographia.* Acta Facultatis Medicae Universitatis Brunensis, 1959, pp 185–196, Lekarska Fakulta University Brne. Used by permission.)

Frequency Dependence

The ratio of output to input amplitudes of harmonic functions will in general depend on frequency; this is expressed in the "amplitude characteristic," i.e., the normalized ratio of output to input amplitude as a function of frequency. Similarly the "phase characteristic" is defined as the function representing the phase shift between the output and the input for all frequencies. Together they fully define the input-output relation of any linear "black box" for arbitrary signals. In filter theory often the two characteristics are combined into the "transfer function."

As an example of the usefulness of this representation, we show in Figure 29–3 the amplitude and phase characteristics of the occipital VEPs of two subjects to sinusoidally modulated light.[16] The responses of both subjects A and B are approximately sinusoidal and show preference for frequencies around 10 Hz. The sharpness (selectivity) of the amplitude plot, however, is much more prominent in B than in A, which is also reflected (as should be expected in a quasilinear system) in the steep course of the phase characteristics around 10 Hz of subject B. At the same time, B exhibits a strong monorhythmic persistent alpha rhythm. The responses have been proved to add to the spontaneous activity without any indication of entrainment phenomena.

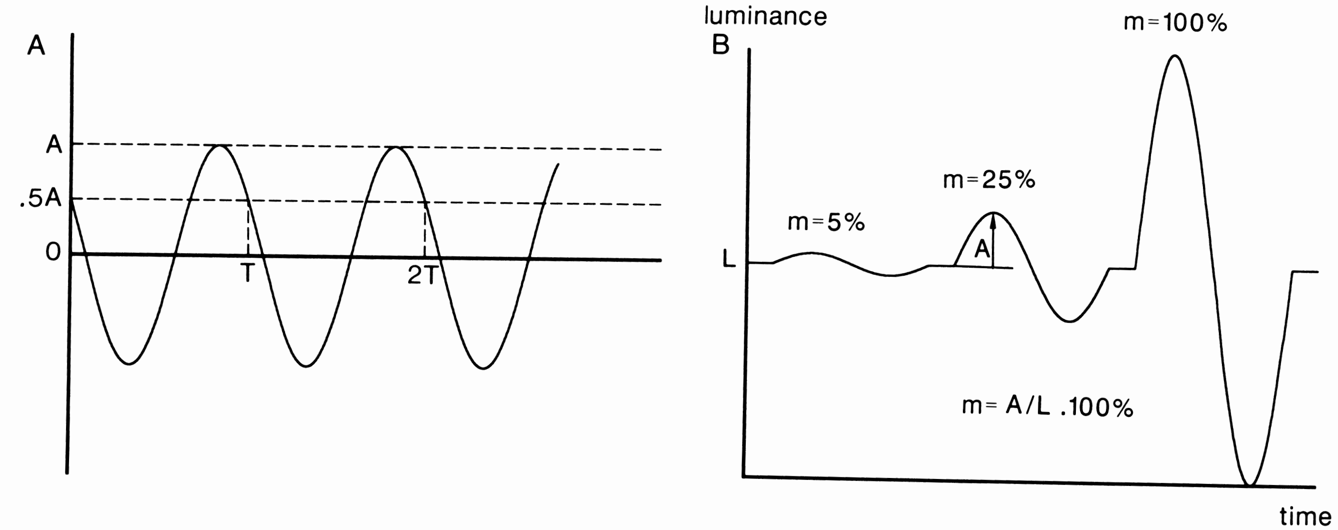

FIG 29–2.
A, an elementary sine wave: $A \sin(2\pi ft + \Phi)$ with the period $T = 1/f$. In the example $\Phi = 5\pi/6$ (150 degrees). An equivalent mathematical representation of the same curve is $A \cos(2\pi ft - \Phi)$ with $\Phi = \pi/3$ (60 degrees). **B,** in case light is modulated, amplitude is given in percent modulation depth (percentage of the average light level).

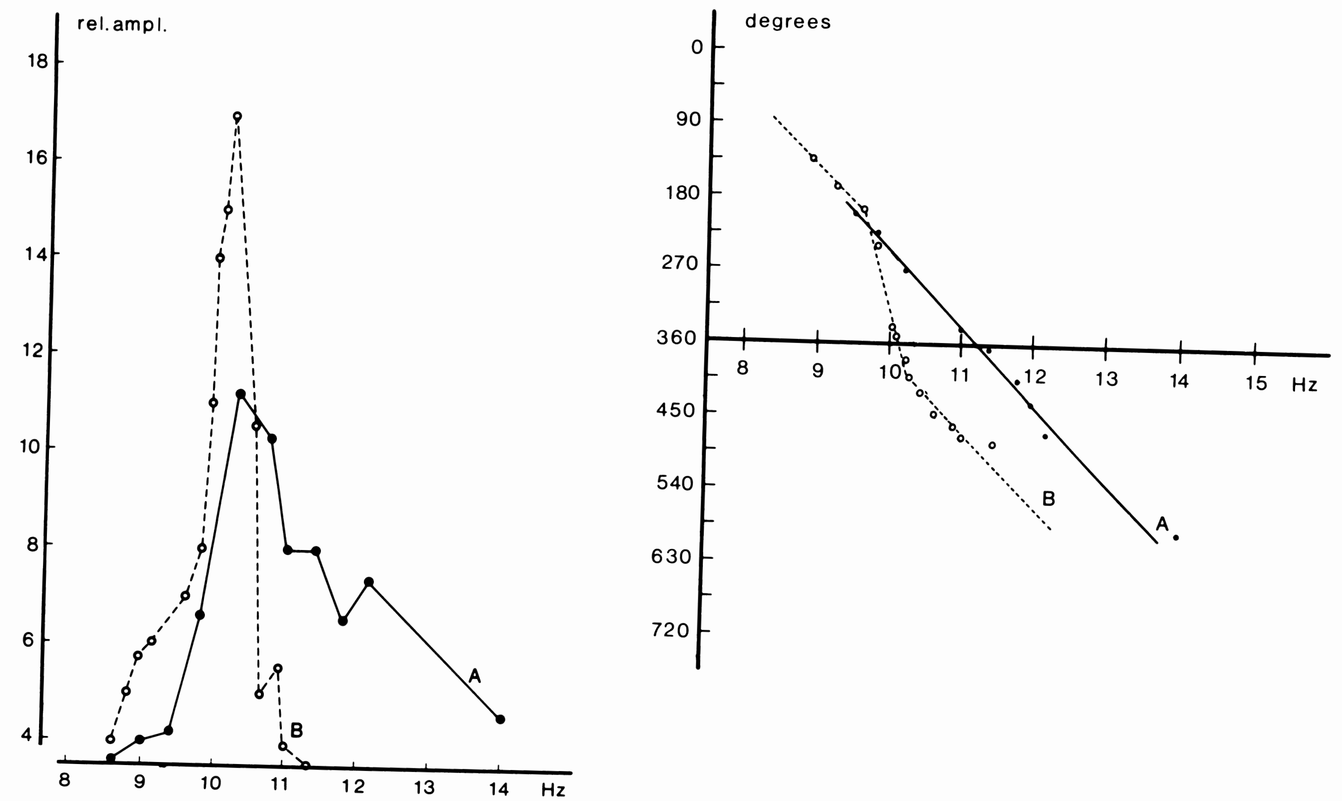

FIG 29–3.
Amplitude and phase characteristics of human occipital VEPs to sinusoidally modulated light. The subjects *A (solid lines)* and *B (dashed lines)* show large differences in the selectivity of the response. This is reflected in the amplitude as well as in the phase plot; the high selectivity in subject B is clearly expressed in an extra phase jump of approximately 180 degrees, while in subject *A* such a jump was not found. Note that the frequency axis is on a linear scale and not on the conventional logarithmic one. The linear phase shift with frequency is equivalent to a delay of approximately 250 ms. (From Van der Tweel LH, Verduyn Lunel HFE: Human visual responses to sinusoidally modulated light. *Electroencephalogr Clin Neurophysiol* 1965; 18:587–598. Used by permission.)

Note that there is no sense in talking about phase relations between sinusoidal signals with different frequencies. If a signal contains more than one frequency, the phase relationship between any two frequencies will vary according to the moment of observation. This applies even to the harmonically related frequencies of a periodic wave shape. For example, Figure 29–4 shows a sine wave and its second harmonic. The phase relationship during one period is different for each point along the x-axis. For certain purposes a characteristic point may be chosen as a reference, for instance, when the voltage changes its sign from negative to positive (positive zero-crossing)

Minimum-Phase Rule, Delay, and Latency

In every physically realizable system a nonconstant amplitude/frequency response is necessarily accompanied by phase shifts that are a function of frequency. In systems without active components (i.e., if only capacities, inductances, and resistances are present) the phase characteristic can be fully calculated from the amplitude characteristic. This is a consequence of what is known as the "minimum-phase rule." If active components are included, phase shifts will often be larger but can never be smaller than the calculated minimum phase for a passive system. The attenuated high-frequency signal at the output of a single low-pass filter, for instance, must be accompanied by a phase shift up to 90 degrees for frequencies approaching infinity. Different amplitude characteristics are necessarily accompanied by different phase characteristics. However, the reverse is *not* true: different phase characteristics can be associated with the same amplitude plot.

An important example of phase shifting without amplitude changes is when a pure delay or latency is involved. In that case the phase changes proportionally to the frequency. The relation is as follows:

$$\tau = (\Phi_2 - \Phi_1)/2\pi(f_2 - f_1) \qquad (1)$$

with τ the delay and Φ the phase shift in radians. In Figure 29–5, the "responses" are subject to a delay of 25 ms. For sinusoidal stimulus "*A*" at 30 Hz the phase delay is 450 degrees, more than one full cycle; *B* exhibits a phase shift of 270 degrees at 50 Hz. The difference is 180 degrees (π). The formula then reads $\tau = \pi/(2\pi \cdot 20)$, which yields a delay of 25 ms, as expected. Note that, if the delay were to be much longer, the phase angles at both frequencies would include more periods, and these extra periods should also be taken into account. If one wishes to avoid ambiguities, a series of closely spaced frequencies should be used. The ensuing phase-frequency regression line should (minimum-phase corrections applied) within experimental accuracy pass either through the origin ($f = 0$, $\Phi = 0$) or through $\Phi = \pm\pi$.

With due caution, therefore, the formula allows for an efficient way to estimate VEP or ERG latencies: by using a frequency band where the response is reasonably sinusoidal and its amplitude does not change too much with frequency. If the amplitude dependency is strong, there will be extra phase shifts that can be estimated from the minimum-phase rule and used for correction.

From the above, it follows that in principle one should not use the phase shift (or, for instance, the peak of the sinusoidal response) at a given frequency as a criterion for delay. Not only are there

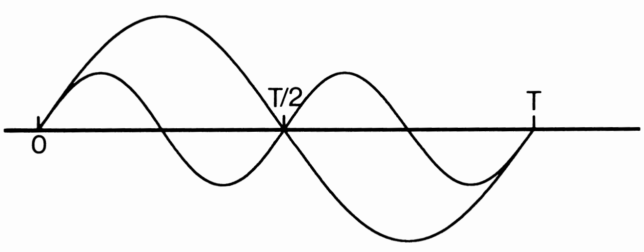

FIG 29–4.
Two harmonically related frequencies. Both phases were chosen to be zero at $t = 0$. At point $T/2$, the phase at frequency $f_1 = 1/T$ is π; at frequency $f_2 = 2/T$ the phase is zero again, i.e., 2π (shifted a whole period).

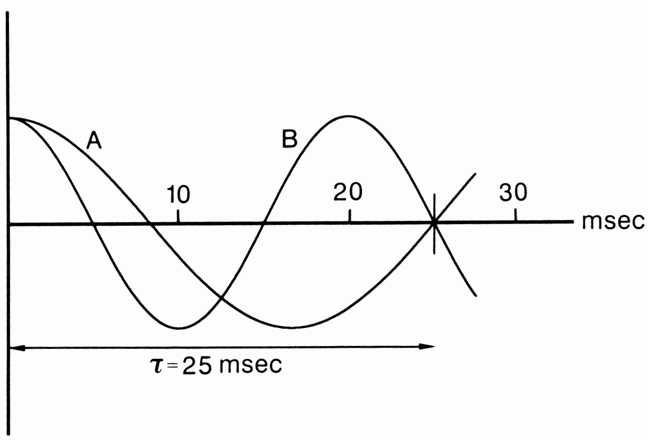

FIG 29–5.
Phase shifts ($5\pi/2$ and $3\pi/2$) at frequencies of 50 Hz **(B)** and 30 Hz **(A)** due to a delay of 25 ms. This delay obeys the formula: $\tau = (\Phi_2 - \Phi_1)/2\pi(f_2 - f_1)$

extra phase shifts possible due to the minimum-phase rule, but there is still another source of error that can trap even the most eminent: due to the periodic character of the stimulus, as was explained above, one or more periods may be missed. This mistake will be mostly evident when unrealistic delays are obtained, but also ambiguities of only half a period may occur. The point to keep in mind is that an increase in light, for instance, has no obligatory relation to the polarity of an electrical response. This is self-evident when recording VEPs because their polarity will also depend on electrode placement. Even when different frequencies are employed, the results must be interpreted with care. For example, Figure 29–1 shows the cat ERG in response to bright, flickering, sinusoidally modulated light of various frequencies. Since the responses are reasonably sinusoidal, linearity is approached. At 0.5 Hz, the phase angle is about 90 degrees and at 3 Hz about 180 degrees. Applying our equation for delay suggests that the latency would be about 100 ms. Now this is quite unlikely, and the probable explanation is that the ERG to this bright modulated light is generated by more than one process, one cornea-positive and the other cornea-negative, and the amplitude characteristics of the two are very different.

There is a special problem with regard to latency determination of (steady-state) VEPs with pattern reversal. Whereas for homogeneous fields sinusoidal modulation is to be preferred, the same is not the case for patterns because abrupt reversal gives a much better defined moment of activation than sinusoidal modulation does. In the latter case the actual moment of excitation will be dependent on the contrast. On the other hand, because abrupt transitions will dominate observation, sine wave modulation is better suited for psychophysical experiments.

Distortion

We have already mentioned that every linear system made of real physical components will exhibit frequency-dependent characteristics. This means that the relation between the original input amplitudes and phases at different frequencies will be subject to alteration, i.e., the output wave shape will not in general resemble the input: it will be distorted. This is called *linear* distortion because the superposition principle applies all the same to the output components. According to Fourier theory, as will be explained below, the shape of an arbitrary input signal is determined by the amplitudes and phases of the frequencies of which it is composed. Since in any real system the relation between input and output amplitudes and phases will depend on frequency, the output shape must differ from that of the input. Shape distortion is only absent in cases where the input is a pure harmonic function, when a sinusoidal input results in a sinusoidal output.

A type of linear distortion that will nearly always be present is attenuation of high frequencies. In mechanics this attenuation is due to inertia, in electricity to (stray) capacitances, and in electrophysiology, diffusion processes (among others) play a similar role. Flicker fusion is an example of high-frequency attenuation, but the ERG and the VEP are also subject to this form of attenuation.

Nonlinear distortion, on the other hand, differs fundamentally from linear distortion. Nonlinearity means that proportionality between input and output does not hold for all signal amplitudes, for instance, often for a large input signal the responses do not grow any more (saturation). Therefore, in the presence of nonlinearities, the superposition principle is *not* obeyed. In principle, all real systems are subject to nonlinear distortion, saturation and thresholds being the most frequent ones (although modern electronic devices can approximate linearity to a large degree). Nonlinearity is an inherent and common property of biological systems. Although the division is not absolute, for our purpose it is useful to distinguish "essential" from "nonessential" nonlinearities. For instance, logarithmic and exponential functions as well as saturation are nonessential: the distortion is strongly dependent on the strength (amplitude) of the phenomenon. If an incremental or decremental signal δ is made small enough, the system may approach linear behavior (because, e.g., $\log(1 + \delta) \approx \delta$). When the input signal is enlarged, quadratic and higher-order terms are playing a role. Rectification, however, belongs to the class of essential nonlinearities; in the ideal case—no matter how small the input signal—the system transmits only one polarity. This means the introduction of significant quadratic and higher-order terms. In reality, rectification will not always be abrupt, i.e., it will not occur at a given break point, but it will be a smooth transition over some small interval and will—for certain small inputs—obey linearity. Nevertheless, in VEP studies rectification can exhibit astonishingly sharp discontinuities, as is shown in Figure 29–6.[12] In this figure it is demonstrated that even for modulations as small as 1.25% of the average light intensity one still obtains a second harmonic response (although with decreasing relative amplitude). Because of the selectivity of the

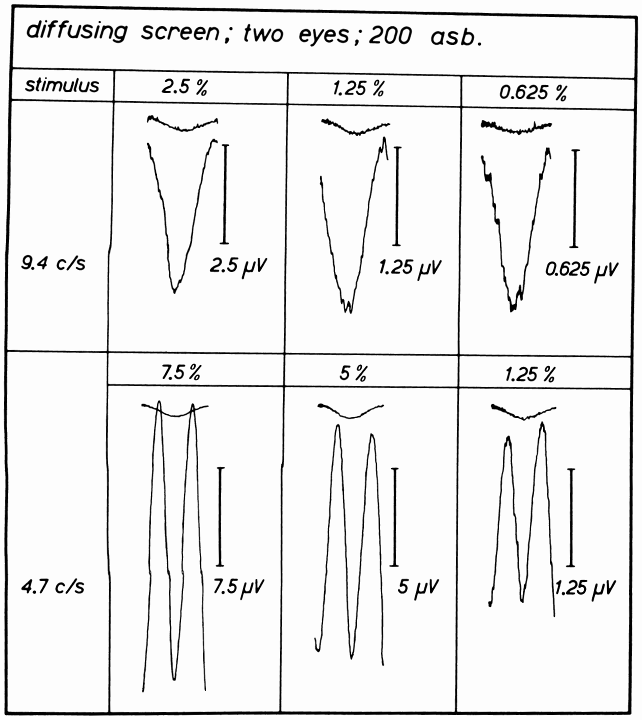

FIG 29–6.
Occipital VEPs to sinusoidally modulated light at 4.7 and 9.4 Hz. The sinusoidal responses at 9.4 Hz are strictly proportional to modulation depth (note the changing amplification factor). At 4.7 Hz sinusoidal second-harmonic responses can be recorded down to 1.25%, although less than proportionally. (From Spekreijse H, van der Tweel LH: *Proc Kon Ned Akad van Wetensch* 1972; 75:77–105. Used by permission.)

responses to sinusoidally modulated light, as shown in Figure 29–3, rectification expresses itself especially at 5-Hz stimulation; the second harmonic is then 10 Hz, just at the maximum of the response characteristic. In fact, at stimulation with 10 Hz the response itself is to a high degree sinusoidal. With respect to rectification, this has been demonstrated in the ganglion cells of the goldfish.[10] It must be noted that distortions caused by rectifiers and by saturation are usually frequency independent. This type of distortion is called static, and such nonlinear elements are called *static* nonlinearities, in contrast to elements whose parameters would change with frequency, which exhibit *dynamic* nonlinearities. Adaptation belongs to the last category.

A consequence of nonlinearity in general is that the response to a combination of harmonic functions will contain new nonharmonic frequencies. For example, if two frequencies f_1 and f_2 are presented and there is a quadratic term in the nonlinearity, the frequencies $2f_1$, $2f_2$, $f_1 + f_2$, and $f_1 - f_2$ will also appear in the output. In fact such a property has been used in visual studies to analyze the system, especially if rectification could be expected.

It is interesting to realize, with regard to the above considerations, that we do not appear to be aware of the inherent logarithmicity of the intensity transformations within our own visual system. Neither do we notice the distortions of considerable intensity that usually occur in black and white photographs. In hearing, however, even small nonlinearities in the chain of sound transformations may be intolerable. They can add "extraneous" and disturbing frequencies: hi-fi was not invented for nothing! On the other hand, the visual system is very sensitive to *linear* distortion that introduces phase shifts in an image. Figure 29–7,A shows two of the bars of a medium-contrast square wave grating on

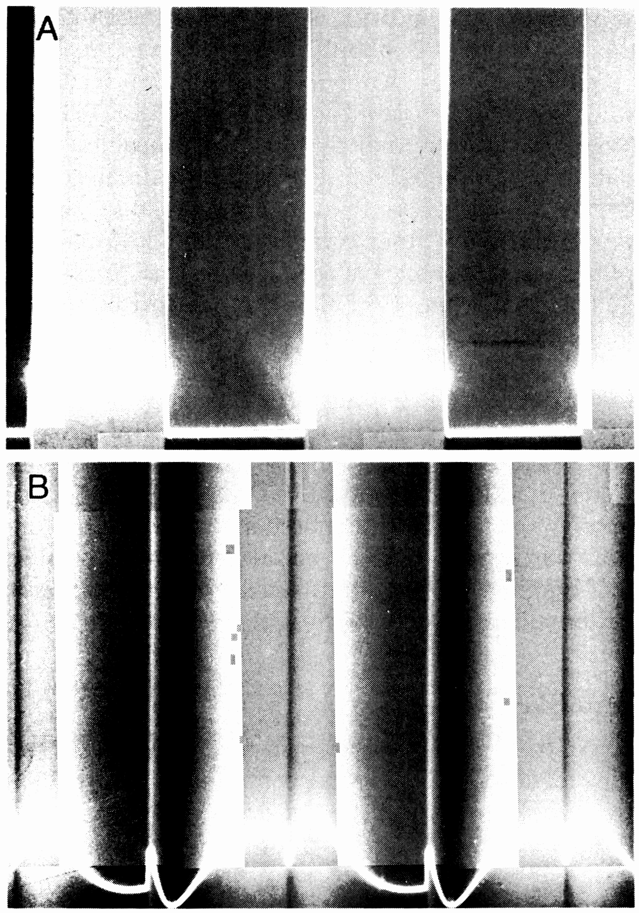

FIG 29–7.
Two cycles of a square wave pattern **(A)** generated on a cathode ray tube. **B,** drastic phase shifting has been performed without affecting the amplitudes of the Fourier components. The generating signals are displayed at the bottom of the patterns. (From Van der Tweel LH: In Spekreijse H, van der Tweel TH (eds). *Spatial Contrast.* Amsterdam, Elsevier Science Publishers, 1977, pp 9–12. Used by permission.)

an oscillograph screen.[13] The white lines at the bottom represent the screen luminance. As will be described in the following chapter, the generating signal of the grating can be resolved into or synthesized from a number of sinusoids, each representing a certain (harmonic) spatial frequency with specific amplitude and phase. The phase relationship of the composing frequencies determines the exact shape of the grating. By electronic means it is possible to shift the relative phases without affecting the amplitudes. As can be seen, the appearance of the grid changes dramatically by such a procedure (Fig 29–7,B).

Again in contrast to the behavior of the visual system, hearing is very tolerant of phase distortion, in any case for periodic signals (Helmholtz's rule). If the two generating wave forms of Figure 29–7 are played into a loudspeaker at, e.g., 300 Hz, A and B sound identical because the same harmonics are present in the same relative amplitudes.

FOURIER ANALYSIS

In 1807 Fourier (whose portrait is shown in Fig 29–8) submitted his epoch-making manuscript on heat conduction to the Institut de France for publication, but it was not until 1822 that his book *Théorie Analytique de la Chaleur* appeared.[9]

In this book one of the most fundamental theorems of physics was developed. According to Fourier's theory, every *periodic* function can be decomposed into or synthesized from an (in principle) infinite number of harmonic functions. The lowest frequency is the inverse of the fundamental period, and all other frequencies are multiples thereof. In addition, there is a term representing the average level of the function. The lowest frequency is also called the *fundamental* or *first harmonic*, and all others form the *higher harmonics*. Their various amplitudes and phases are such that, when added together, they reproduce the shape of the original function.

Fourier's formula for periodic time functions reads as follows:

$$F(t) = a_0/2 + \sum_{n=1}^{\infty}(a_n \cos 2\pi nft + b_n \sin 2\pi nft) \qquad (2)$$

where $F(t)$ is a *periodic* function with the period T and a_0 is a constant that represents twice the mean of the function, also conventionally (and sloppily) called the DC-term or, when light is considered, the average illumination.

The frequency f (in hertz) is the inverse of T in

FIG 29–8.
Portrait of Joseph Fourier, lithograph. (Courtesy of the Museum Boerhaave, Leyden, The Netherlands.) The text below reads: "Membre de la Légion d'honneur, etc. Né à Auxerre, le 21 Mars 1768, élu en 1817 et Sécretaire perpétuel pour les sciences mathématiques en 1822.' Translated: "Member Legion of Honor, born Auxerre, France, March 21, 1768, elected in 1817 and Permanent Secretary for Mathematical Sciences in 1822."

seconds. If $n = 2, 3, 4$, etc., one speaks of the second, third, fourth, etc., harmonic. Often the sine and cosine terms are taken together:

$$F(t) = A_0/2 + \sum_{n=1}^{\infty} A_n \cos(2\pi nft + \Phi_n) \qquad (3)$$

where $A_0 = a_0$, $A^2_n = a^2_n + b^2_n$, and $\tan \Phi = b_n/a_n$.

If distance (x) is the variable, then we use P instead of T: $f = 1/P$. For spatial periodic phenomena in vision, spatial frequency is given in cycles per degree and period in visual angle (degree). The terms a_n and b_n can be respectively computed from the equations

$$a_n = 2/T \cdot \int_{-T/2}^{+T/2} F(t) \cos 2\pi nftdt \qquad (4)$$

and

$$b_n = 2/T \cdot \int_{-T/2}^{+T/2} F(t) \sin 2\pi nftdt \qquad (5)$$

For computational reasons harmonic functions are often displayed in complex notation as $C_n \exp (jn\Omega t)$, with j being the square root of -1 and Ω, $2\pi f$. This is identical with the sin-cos treatment because $\exp(jn\Omega t) = \cos n\Omega t + j \sin n\Omega t$.

Standard Periodic Signals

The scope of Fourier analysis can best be understood by considering the analysis of simple periodic functions like a square wave. Its components can be easily calculated (Fig 29–9). Note that the fundamental has an amplitude ($4/\pi$) that exceeds that of the square wave itself. As the third and a few more of the harmonics are added, the synthesized shape approaches the square pattern more and more. There remains, however, a narrowing overshoot of about 18% that shifts toward the steep flanks of the square wave as more and more harmonics are included; this is called Gibbs' phenomenon.

Another important standard signal is that formed by periodic impulses. Theoretically the impulse (or δ-) function is by no means simple, but for the present purpose, the following explanation will suffice. It is supposed that each impulse is infinitely short and infinitely high. For instance, electrical current impulses can be given with diminishing duration δt and increasing strength i in order to keep the total charge per impulse constant. This total is normalized to *unity*, and the mean charge (dc or zero-frequency component) becomes then $1/T$; light flashes can be treated in a similar way. The Fourier spectrum of these repetitive impulses consists of equally spaced spectral lines of constant amplitude at frequencies f, $2f$, $3f$, etc. These are again impulse functions, but on a frequency axis (Fig 29–10,A), the amplitudes are $2/T$. Since the average of periodic unit impulses is by definition $1/T$, all sinusoidal components extend into the negative part of the amplitude axis (Fig 29–10,B). At present, comparatively simple computer programs are available that easily perform on-line Fourier analysis of any signal, periodic or not, so quickly that the calculations seem instantaneous. Note that the Fourier analysis of a nonperiodic signal is technically performed as if the analyzed interval is part of one period of a repetitive signal! This period should be chosen with due care.

The Fourier Integral

Theoretically, the Fourier series represents or reproduces a *periodic* function that continues *forever*. In practice this can never be achieved, although a sufficiently long interval may be considered to approach the ideal situation. For single events or transients, i.e., signals of the "one-in-a-lifetime" sort, the discrete sum Σ is substituted by an integral. Figure 29–11 gives an example to better understand the meaning of this extrapolation. In Figure 29–11, we start with the presentation of a square wave with its Fourier line spectrum as was treated before in Figure 29–9. Subsequently, one period of this square wave is isolated and repeated with progressively longer periods. The spectral representation for the signals

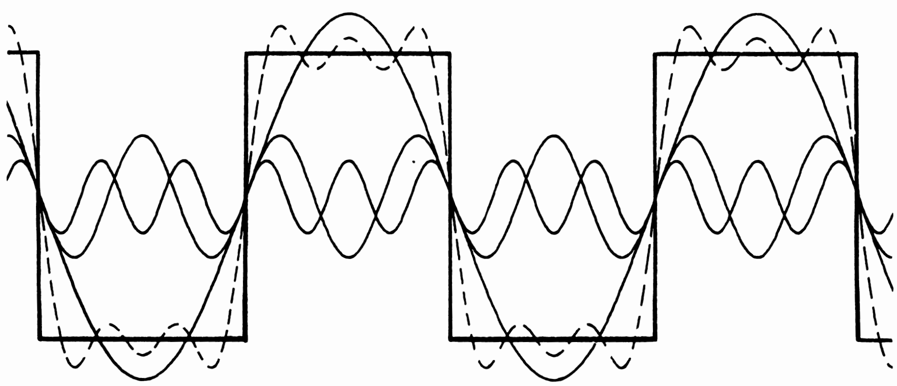

FIG 29–9.
The first three harmonics of a square wave and their sum (dashed line). The amplitude of the fundamental exceeds that of the square wave. The overshoot does not disappear but becomes narrower when more frequencies are included.

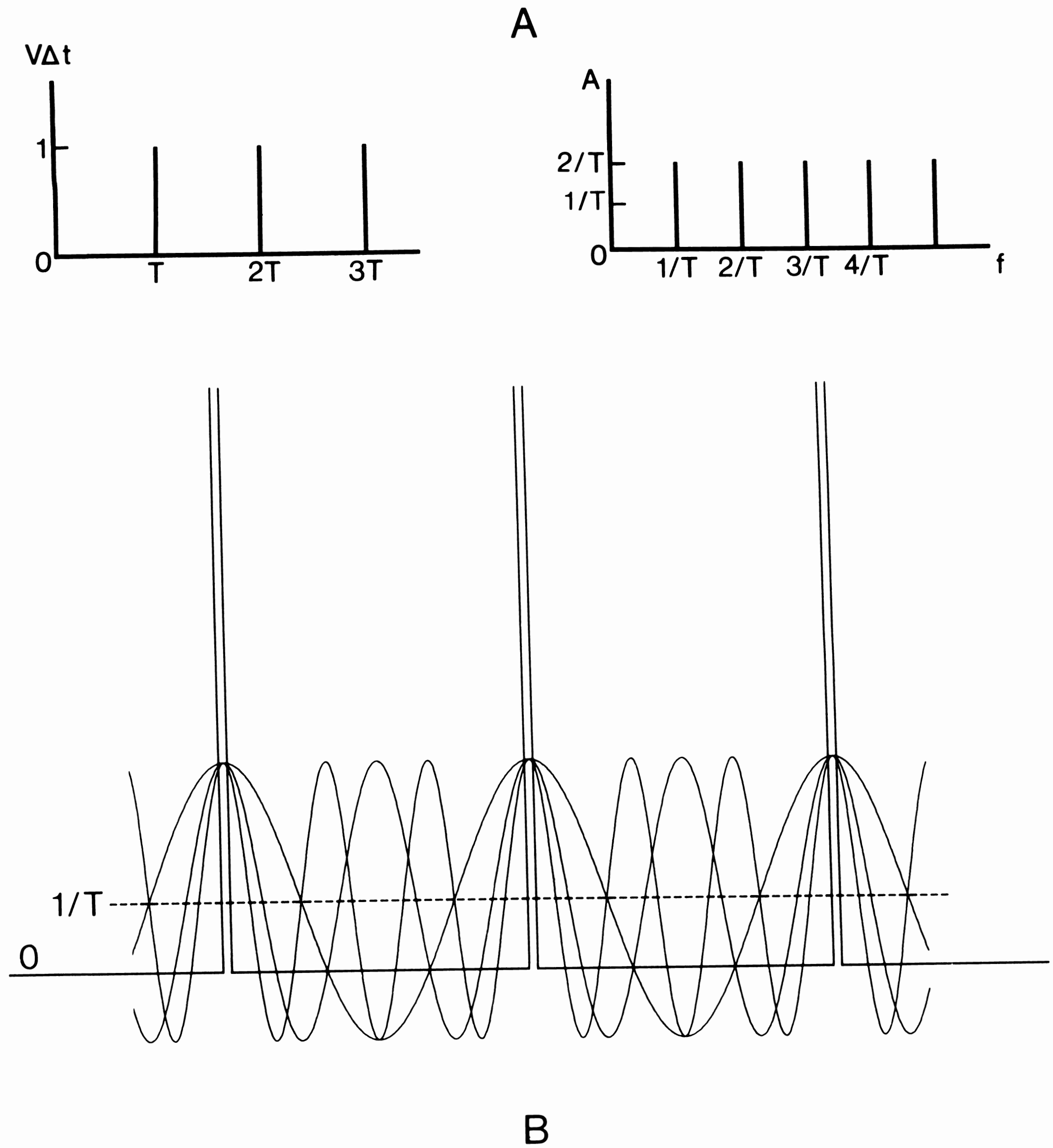

FIG 29–10.
A, Fourier spectrum of periodic impulses normalized to $V \cdot \delta t = 1$. All amplitudes are $2/T$; "DC" level, $1/T$. Note that the impulses in the time domain have the dimension volt · seconds and **not** volt. **B,** periodic impulse functions (period T) as they are built up from cosine waves of frequencies n/T ("DC" level, $1/T$). The virtual modulation depth of all components is 200%.

with lengthened periods is indicated in Figure 29–11. In accordance with the increase in period T, the fundamental frequency $1/T$ decreases, which results in more closely spaced spectral lines.

If this procedure were to be continued (i.e., is extrapolated to an infinitely long interval), the spectral lines would become infinite in number and therefore approach a continuum, at which moment they represent the Fourier transform of a single event; the Fourier *series* (the line spectrum of Equation 3) has become the Fourier *integral:*

$$F(t) = 1/2\pi \cdot \int_{-\infty}^{+\infty} A(\Omega) \exp j\Omega t d\Omega \qquad (6)$$

with $\Omega = 2\pi f$ and f the frequency. At the same time this forms the (nonnormalized) envelope of the former line spectra.

Especially revealing is the case of a single impulse. If the procedure just described is applied to the case of Figure 29–10,A, extrapolating to a single

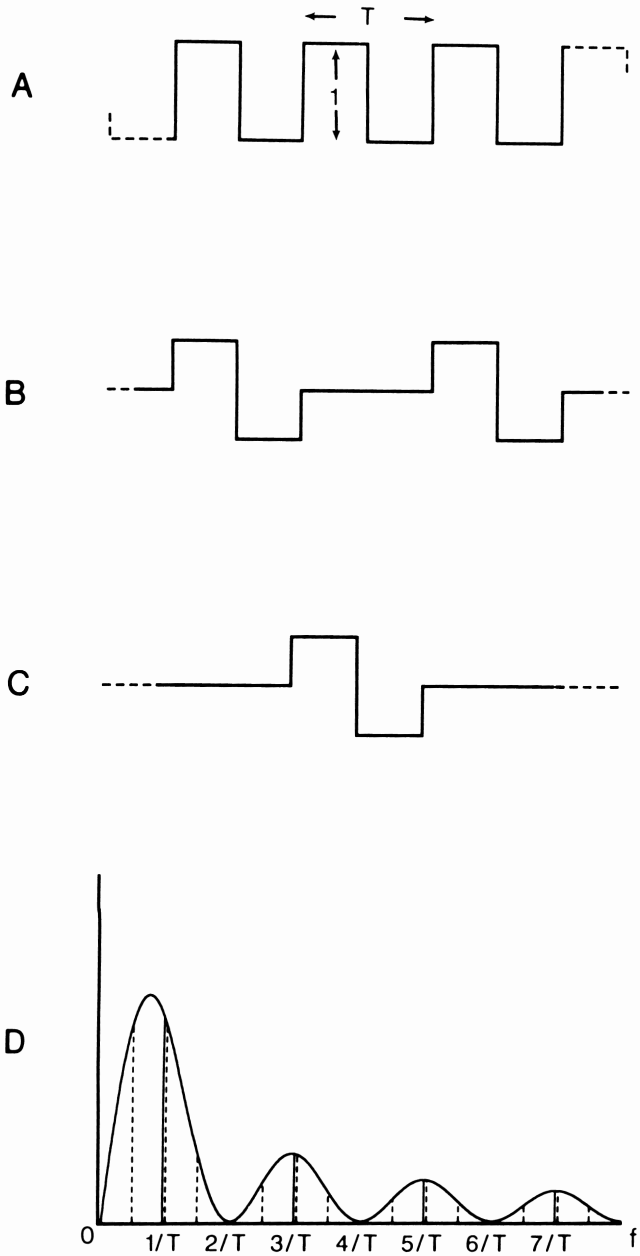

FIG 29–11.
A square wave **(A)** of period *T* is dissected and presented with increasing intervals: 2T **(B)**, . . ., infinity **(C)**. **D,** the original line spectrum at $f = 1/T$, $1/3T$, ... etc., is filled up at increasing periods until the Fourier integral is approached (continuous curve). All spectra are normalized.

impulse, the continuous spectrum that results exhibits a constant amplitude at all frequencies. All composing harmonic functions reach maximal amplitudes at the moment of the occurrence of the impulse (see also Fig 29–10B). Therefore they are represented by cosine functions at $t = 0$. Even though their relative amplitudes will be (infinitely) small, they will add significantly at time zero and nowhere else.

Test Signals

In theory, the characteristics of a linear system can be determined from its response to nearly any transient signal; the transfer function follows from the ratios of the amplitudes of each component, which are obtained by Fourier analysis of the input and output signals. In practice, impulses and step functions are usually preferred as test signals. The impulse function is a natural choice because of its constant amplitude spectrum and well-defined phase spectrum. All constituting frequencies have an equally large, although vanishing, small maximum at $t = 0$ ($A \cos 2\pi ft$). Because of this equal amplitude spectrum the Fourier coefficients of the response directly yield the transfer function.

The response to a *unit step* function, i.e., the integral of the impulse function, can also fully define a system. Step functions are typically useful in slow systems because their low-frequency content is high. Another technique is to directly determine the frequency response with a close enough series of sine waves of different frequencies. In vision research nonlinearities of the saturation type should also be considered; when one wants to avoid them, either weak incremental or decremental flashes (sudden decreases of a certain luminance) with long enough pauses may be employed, or sinusoidal modulation with restricted modulation depth can be used to probe the frequency range of interest. In principle, the two methods will be just as time-consuming to arrive at equally reliable results. If one wants to study nonlinearities, sinusoidal modulation will often be preferable above strong flashes because distortions of the sinusoidal shape are easily detected. Moreover, the adaptation state of the eye is well defined. Which of these techniques is to be preferred depends upon the question at hand and the system to be analyzed.

Domains

Notwithstanding the fact that Fourier analysis is mathematically a straightforward procedure, conceptually it is by no means simple. As previously mentioned, a variable quantity (e.g., luminance) that is a function of time or space can equally well be represented by means of its Fourier transform as a function of frequency. Although probably superfluous, it must be stressed that in the frequency domain "time" or "space" have themselves disappeared as a dimension.

An example from acoustics is probably the most revealing: if we were to perform Fourier analysis on a symphony of Mozart of, say, 30 minutes, we would obtain a fundamental frequency of 1/1800 Hz. If the frequencies present in the music extend to 20 kHz, there will be no less than 3.6×10^7 lines in the amplitude spectrum and as many points in the phase spectrum. *The original representation of the physical phenomenon is replaced by a set of numbers, and time itself is lost.* These numbers, if used to code corresponding physical generators (including phase information), would allow one to reproduce the original phenomenon. (Note: In a recent lecture, Stan Klein has advanced the idea that a musical score can be considered a Fourier representation *avant la lettre*.)

Our auditory system indeed allows a musical person to extract separate pitches from a complex sound, apparently performing some kind of Fourier analysis (Ohm's acoustical law). In agreement with this, a sound wave is in general perceived as an entity, and no individual pressure vibrations are heard. Of course, if a really simple form of Fourier analysis were being performed by the ear, we could only "hear" the symphony after it has been played, and it would be just the same whether played forward or backward. In reality, the ear performs a sort of running frequency analysis with a sliding time window of approximately 50 ms, and in this way there is both frequency analysis *and* flow of time.

Theoretically the duality of the frequency and space domains applies in a similar way to the visual scene as to sound waves. However, in contrast to the perceived uniqueness of a sound wave as described above, each and every element of a grating is always perceived as a distinct entity, even at levels approaching threshold. Therefore the number of periods of a grating plays a much less important role in perception than do the number of oscillations in audition.[5]

There have been suggestions that for certain visual tasks such as recognition of blurred faces the visual system would employ a kind of piecemeal Fourier analysis because this would be technically advantageous. From the above exposition it should be evident that the representations in time or space carry the same information as those in the frequency domain. Both representations will therefore need the same amount of computation, although of course the physical or/and anatomical properties of the brain may confer practical advantages to one or the other modes.

Practical Considerations on Using Fourier Analysis

In practice, a recorded response will always be restricted to a certain duration T or, in space, to a certain restricted region S. In the execution of a Fourier transform, however, the phenomenon is usually interpreted as periodic in T or S, as was mentioned before. Therefore the Fourier transform will not be continuous but display a discrete although dense spectrum with line distances of $1/T$ or $1/S$. In case of repetitive stimulation this is evident: if an ERG or a VEP is presented for 0.5 seconds, the lowest meaningful frequency will be 2 Hz, and harmonics will also occur in multiples of 2 Hz. Although there are programs that will enable a higher resolution by artificially extending the time interval being analyzed, no information is gained by this. If a frequency spectrum is to be computed from an *averaged* signal, it makes no sense to extend the analyzed interval beyond that section of the recording that can be considered to exceed the noise. In practice, such a section will be extended to a longer analysis period to obtain a dense enough spectrum.

An important consideration is that periodicity can only be an idealization in practice. One must realize that in reality a finite number of periods is always present; time or space will be the restricting factor. This means in Fourier terms that the theoretical discrete lines of exact periodicity will broaden: the spectrum is continuous. In Figure 29–12 the effect on the Fourier spectrum of restricting the number of periods of a square wave is demonstrated. An important rule of thumb is that the width of the spectrum δf around the center frequency f multiplied by the number of periods N is approximately constant, thus

$$\delta f \cdot N = K. \qquad (7)$$

The spectral lines of higher harmonics are subject to the same *absolute* widening, i.e., δf. This is of special concern for gratings because even if the number of periods in the optical stimulus is large, the effective number of periods on the retina will be different for various spatial frequencies. As a consequence, the effective bandwidth will be also different for different frequencies. Another complicating factor is due to eccentricity effects: along the bars of a centrally

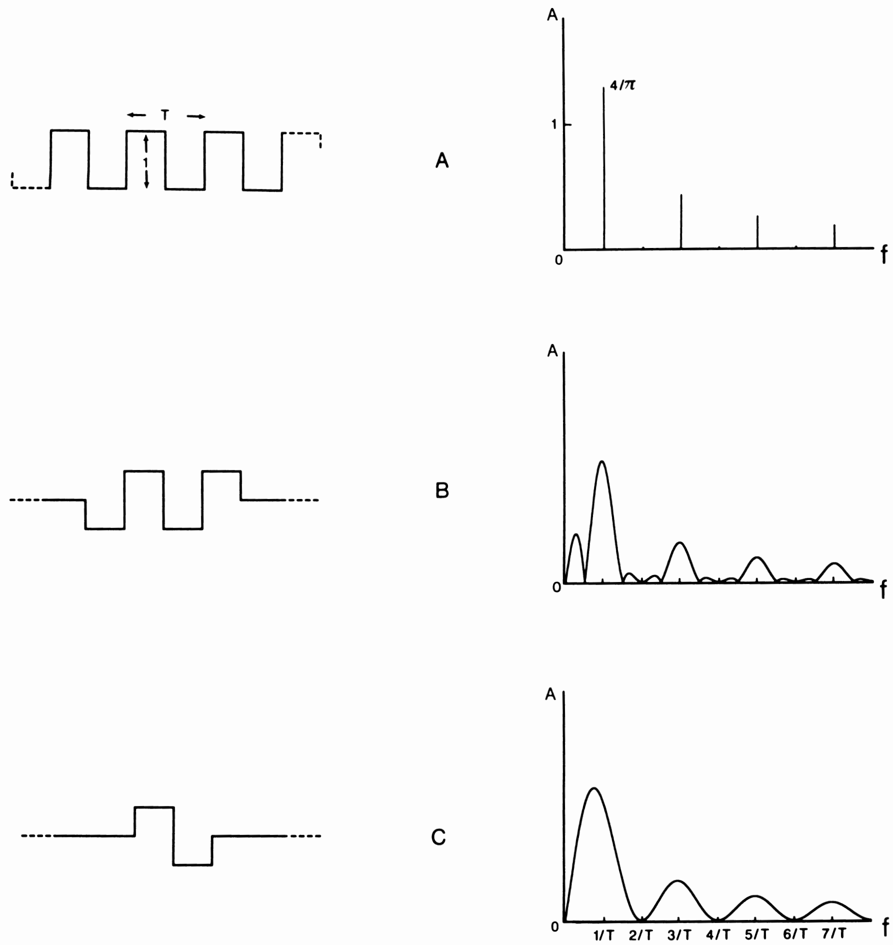

FIG 29–12.
A, the line spectrum of an infinite square wave. **B,** spectrum of two periods of the same. **C,** spectrum of one period, equal to 11 D (Spectra normalized).

presented grating the effective length will be less for fine gratings than for coarse ones. The implications of all this are difficult to oversee and will certainly affect the experiments in different ways. For a system performing Fourier analysis, discrimination of periodicity or frequency will be impaired if only one or a few periods are present due to the broad maxima in the Fourier transform. This is easily demonstrated in hearing: if one or a few cycles of a sine wave are presented to the ear, a transient is heard with very little pitch. However, the fact that this effect has no equivalent in spatial vision suggests that vision does not primarily rely on harmonic analysis. We discuss this question in the next section.

Real signals also deviate from ideal ones in other ways. For instance, neither strict periodicity nor constant amplitude are really achieved in practice, although in our case deviations will be mostly negligible. More important is that in recording VEPs and ERGs people blink, move their eyes, and shift their attention; therefore even for an ideal stimulus, VEPs will exhibit latency jitter and fluctuations of amplitude. Attention to this will be paid when discussing averaging, but here it can be said that in principle fluctuations will also influence the initial Fourier line spectrum of signals that are in origin periodic.

Although the concept of noise will be treated extensively later, one aspect deserves attention within this framework: white noise is defined by its Fourier spectrum as consisting of a continuum of frequencies with constant amplitude. The only difference between this noise spectrum and that of the impulse function is found in their phase spectra. For noise, phases are distributed at random, instead of being equal (cosine functions) at a prescribed time, as is the case for an impulse function. This again demonstrates the importance of phase, which is of such special significance in the visual world.

Just as in the real world the width of an impulse function can not be reduced infinitely, similarly the frequency spectrum of noise will be restricted at the high-frequency end.

SOME ASPECTS OF SPATIAL FOURIER ANALYSIS

Linearity

All optical systems, contrary to physiological ones, are linear in so far as the relative light distribution of an optical image is independent of intensity. However, it is a misunderstanding to think that optical systems are linear in their *reproduction* of images. For example, if a sinusoidal grid is imaged through a simple lens, the image may expand toward the edges and introduce new frequencies in the Fourier spectrum. Only in well-corrected optical systems do harmonic functions retain their harmonic character.

Time vs. Space

There is an essential difference between space and time phenomena: time has an inherent direction, whereas space does not. Therefore, smoothing in time is always accompanied by phase shifts, whereas smoothing in space will generally be performed without phase distortion. Actually the minimum-phase rule is a consequence of causality. So-called phase-free filtering has recently been developed, but it can only be performed by computer or by reversing the playback of a tape.[19] Because the signal is processed artificially, the result is not anymore restricted to the past; it may be looking into the "future" i.e., causality can be violated: the "response" can precede the stimulus. Figure 29–13 gives a schematic explanation of elementary phase-free low-pass filtering.[15] In modern practice filtering is mostly performed digitally but digitally does not necessarily imply "phase-free." In principle time reversal is then an essential condition and therefore physical laws are no longer valid. In the figure the signal is filtered twice, once in the normal direction and once backward. The reversed result of the latter is then added to the first. One can see that the result spreads from $+\infty$ to $-\infty$, which makes estimation of latencies dubious. This in contrast to real-time low-pass filtering when, in principle, the start of a transient can always be found (in the noiseless case), even if more filtering stages are present. On the other hand, peak latencies will generally be much more reliable with phase-free filtering, which is es-

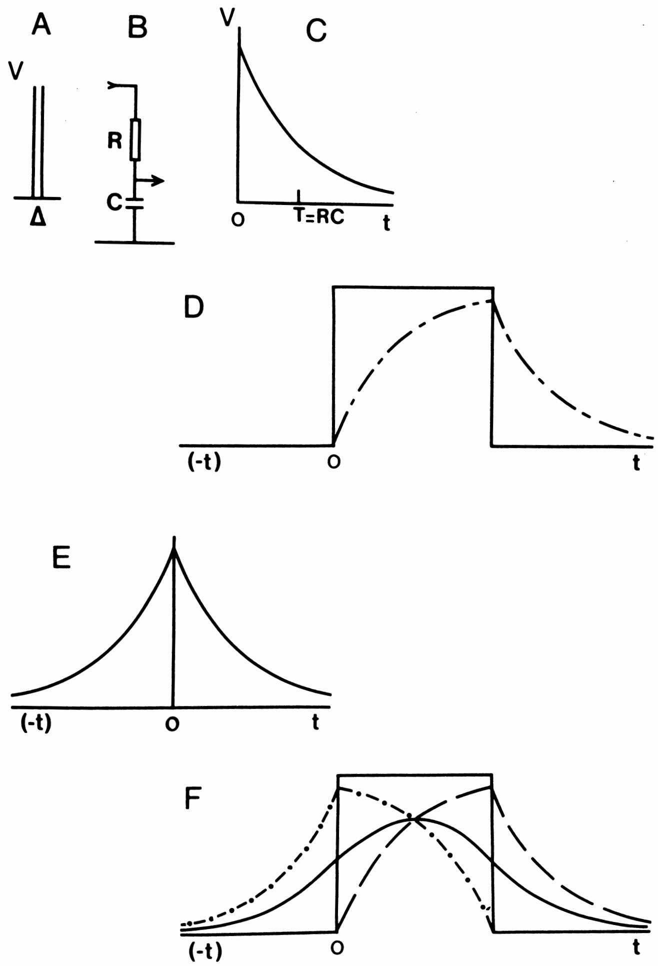

FIG 29–13.
Principle of phase-free filtering. **A,** a (voltage) impulse function (Δ-function). **B,** the circuit—a simple RC network through which it is passed. **C,** the resulting output. **D** shows a square pulse *(full line)*, and *dashed lines* indicate the output after filtering. In a computer or tape the memory can contain all values from $-t$ to $+t$, where t is a large number, so the filter can be made to act upon the signal not only in "real time" but also in the future! The mathematical effect of this is shown in **E** for an impulse response and in **F** for a square pulse. The *dashed lines* in **F** show the operation in the "forward" mode, and the *dash-and-dotted lines* indicate the reverse operation. The resulting (normalized) output is shown by the full line. The result may be pretty, and the peaks may not be "displaced," but the time of origin of the response is lost. (From Van der Tweel LH, Estévez O, Strackee J: Measurement of evoked potentials, in Barber C (ed): *Evoked Potentials.* Lancaster, England MTP Press, Ltd, 1980, pp 19–41. Used by permission.)

pecially advantageous for high-pass filtering of repetitive signals.

In space the situation is different. Often blurring will be more or less circular symmetrical; with modern compound lenses, more complicated defocused patterns may be produced. Most optical systems can be adequately characterized by either their point or their line spread function. The line spread function has much in common with the impulse function in time. Just as the impulse response of a filter gives the frequency response, the Fourier transform of the line spread function directly gives the optical transfer function. The line spread function is very often symmetrical due to the already mentioned equivalency of right and left in space. In optical systems that are not well corrected and also in vision, both point and line spread functions will depend on eccentricity. Moreover, in general, line spread functions should be determined with line elements that are not too long.

A fact that deserves special attention in applying spatial Fourier techniques is that of the axial alignment of object and image. This alignment has to be accurate; if not, linear phase shifts will occur. For instance, a misalignment of 1 degree will give a phase shift of 4×360 degrees for spatial frequencies of 4 cycles per degree (period of 15 minutes) etc., because distance is translated in phase, *including whole periods*.

Spatial Frequency

Fourier theory is obviously also valid in space: any distribution of light can theoretically be obtained by a composition of spatial harmonic functions. However, whereas Fourier analysis in time is straightforward, the two dimensionality of space requires more complicated techniques and raises its own questions when applied to the study of the visual system. The most fundamental problem arises when light distributions are to be expressed in terms of their Fourier spatial harmonic content; it is due to the nonexistence of negative light. Consider the case of a 100% contrast bar pattern (black and white): according to our discussion in the section "Standard Periodic Signals," the fundamental exceeds in amplitude that of the original light distribution, which means that this "component" would go negative!

For an electrical signal there are no such problems because negative and positive are symmetrical. Neural processes also permit encoding of negativity, and therefore in principle Fourier processing of neural signals would be feasible—assuming that the average can also be properly encoded. On the other hand, there being no negative light, a given light distribution cannot in general be synthesized from physical harmonic gratings.

Fourier analysis of one-dimensional patterns, for instance, square or triangular grids, is straightforward; the extension along the elements is in principle infinite, but the resultant normalized Fourier spectrum will not change if shorter elements are employed. In experiments with grids, the length of the grid elements may influence the amplitude of VEPs or as such the threshold. If we take as an example a square grid with a contrast of $A\%$ and a period (twice the bar width) of X degrees, this will be represented by the average level (luminance) and frequencies of $(2n + 1) \cdot 1/X$ c/degree with respective amplitudes of $4A/\pi$, $4A/3\pi$, etc. It is interesting to note that the eccentricity effect will influence the harmonics differently. As far as the authors know, no systematic research has been performed on this matter.

Note that, depending on the sharpness of the bars, many "frequencies" are present in this representation. In this respect, "bar width" is a simpler characterization of a square grid than is spatial frequency. "Frequency" can better be reserved for sinusoidal gratings.

Checkerboards and Fourier Analysis

To understand the meaning of Fourier analysis of two-dimensional figures, namely, that of the checkerboards that are so often used in VEP research, consider in Figure 29–14,A the heavy arrowed lines.[14] The mean luminance along these lines is then half that of the white squares. Therefore there is no periodicity perpendicular to these lines in the sense of Fourier analysis, i.e., no spatial frequency of 1/2a (Figure 29–14,A). For the tilted rays of Figure 29–14,B, alternate diagonals fall in the black or in the white fields, and thus the first true Fourier components are at 45 and 135 degrees. The profile is perpendicular to the diagonals and is triangular with identical maximum contrast as that of the checkerboard elements. The period P in this direction is equal to $a\sqrt{2}$; the spatial frequencies belonging to this triangular profile are therefore the inverse of this and are subsequently $P/3$, $P/5$, etc. The amplitudes will be $8/\pi^2$, and respectively 1/9, 1/25, etc., times this. All the above applies for both diagonal directions. At the position of Figure 29–14,C the integrated contrast will be zero again. The orientation *selectivity* will depend on the number of checks in-

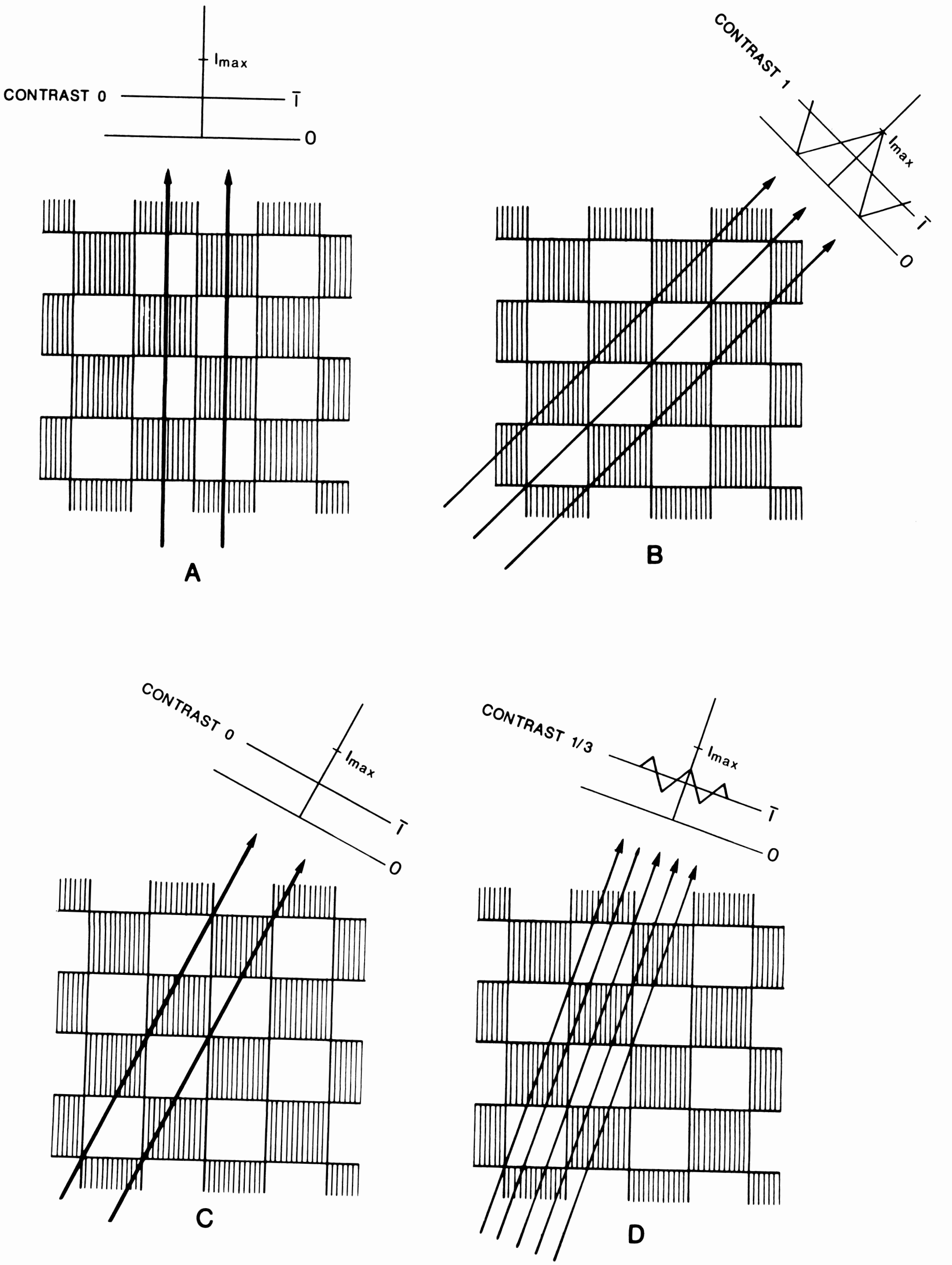

FIG 29–14.
The integrated luminance along the elements of a checkerboard (of 100% contrast, i.e., black and white) in some main directions. **A,** in the horizontal and vertical directions there is no variation. Therefore there is no component with a spatial frequency of 1/2*a*. **B,** diagonally, a triangular distribution with the original contrast is obtained. This forms the first spatial frequency present: $1/2a \cdot \sqrt{2}$. **C,** In this direction and in its counterpart, again, no net result is obtained: **D,** two of three checks cancel, and the contrast is reduced to 1/3. Four equivalent directions exist. (From Van der Tweel LH, Estévez O, Pijn JPM: *Doc Ophthalmol Proc Ser* 1983; 37:439–452. Used by permission.)

cluded; for large numbers orientation selectivity will be very sharp. For the orientation in Figure 29–14,D, by the same reasoning a triangular grid will be produced with a contrast of ⅓ and a spatial frequency $1/2a \cdot \sqrt{10}$. Due to the symmetry of the checkerboard there are now four identical orientations. *Note that there is no harmonic relation anymore between the main components.* Therefore one can never speak of the *fundamental* of a checkerboard, and analysis and synthesis is by no means as simple as for a bar pattern. As in the case of bars, the only unambiguous definition of a checkerboard, and also the simplest, is by the size of its elements.

Receptive Fields

There is a considerable amount of literature about receptive fields and their representation in the frequency domain.[4] As long as one accepts linearity, the two representations are perfectly equivalent. For not too complicated receptive fields, the representation in the frequency domain is necessarily broad because the lowest frequency has to fit more or less the field size and, as we have seen, the Fourier transform of one single period has a very broad representation. For receptive fields with straight edges the transform remains rather elementary with equidistant zeros. Circular fields are much more complicated in this context, however. For instance, no harmonic relations exist anymore between zeros or maxima.

It is of course possible to imagine receptive field structures, including inhibitory zones, that would provide sharper frequency selectivity. A discussion of these problems is beyond the scope of this chapter, but it should be understood that the mathematical background necessary to treat the "receptive field" concept in Fourier terms is much more complicated than is often appreciated.

CORRELATION TECHNIQUES, NOISE, AND POWER SPECTRUM

Autocorrelation

Autocorrelation and cross-correlation functions are of great help in defining and understanding linear (and also certain classes of nonlinear) systems, as well as in performing system analysis. We shall treat functions of time as an example, but our discussion is equally valid for functions of variables with other dimensions.

The definition of the autocorrelation function $R(\tau)$ of $f(t)$ is as follows:

$$R(\tau) = 1/2T \cdot \int_{-T}^{+T} f(t)f(t-\tau)dt \qquad (8)$$

with T going to infinity (Fig 29–15).

If $f(t) = A \cos(2\pi ft + \Phi)$, it is easily calculated that

$$R(\tau) = A^2/2 \cdot \cos 2\pi f\tau \qquad (9)$$

(see Fig 29–16).

Note that τ is a time lag or time difference and must not be confused with time itself: the autocorrelation function is *not* a process in time but a purely mathematical construct. For $\tau = 0$ the autocorrelation function is always maximal because it is then exactly the integrated square of $f(t)$. $R(0)$ is conventionally normalized to 1. Another property of this function is that $R(\tau) = R(-\tau)$, that is to say, $R(\tau)$ is symmetrical around the origin. This symmetry corresponds with the loss of phase information for harmonic functions. In figure 29–16 some examples are given of autocorrelation functions of common signals.

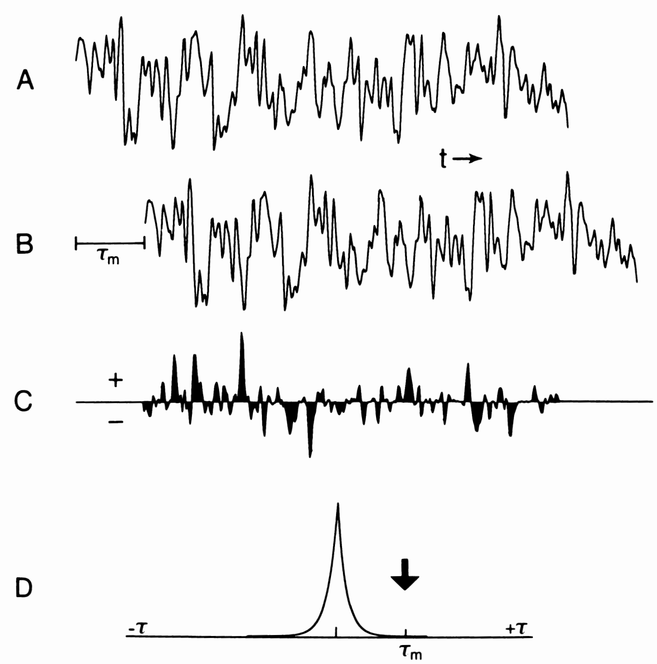

FIG 29–15.
Principle of calculating the autocorrelation function. The signal **(A)**, filtered Gaussian noise, is multiplied with a copy of itself **(B)**, but shifted with a (discrete) lag τ_m. **C** shows the outcome of the multiplication. The values at all sampling points are added. The result is point τ_m of **D**, where the total function with m ranging from $-\infty$ to $+\infty$ is presented.

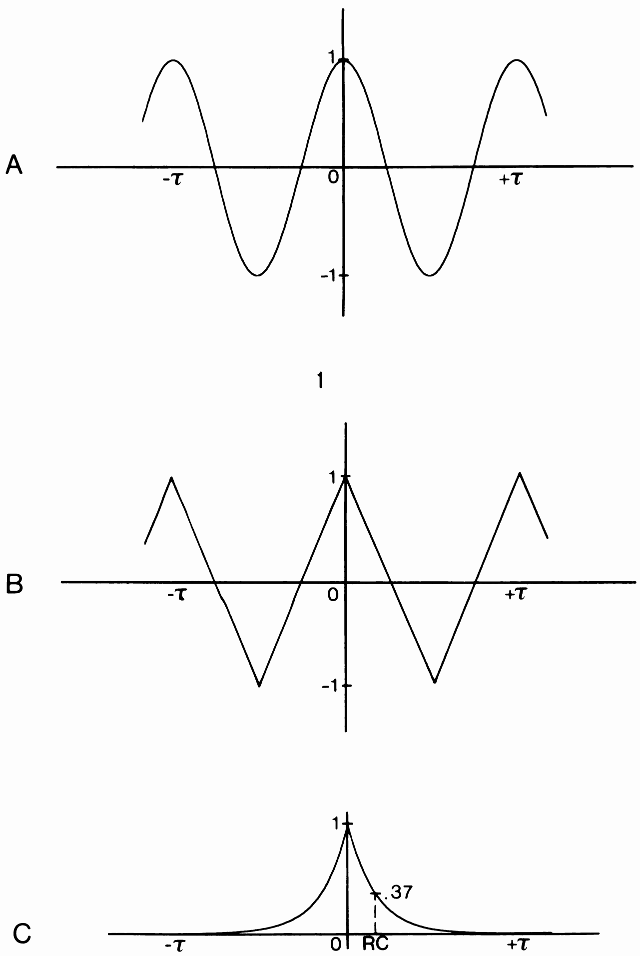

FIG 29–16.
A, autocorrelation functions of A cos $(2\pi ft + \Phi)$. **B,** a square wave. **C,** one-stage low-pass filtered noise with the time constant t_c = RC.

Gaussian Noise

The concept of autocorrelation is especially useful to define noise. An important case is that of white noise; this is a signal whose autocorrelation function is an impulse. Only for the delay $\tau = 0$ is there a net value, normalized to 1; everywhere else it approaches 0. This definition in fact formalizes the inherent unpredictability that characterizes noise. In practice, the frequency band of the noise will be always restricted; therefore the autocorrelation function will be a broadened impulse function, as is exemplified in Figure 29–16,C, where the autocorrelation function of noise filtered by a one–RC stage, low-pass filter is shown. The result is a symmetrical exponential.

We shall consider here only the most elementary form of noise, i.e., noise with an amplitude distribution:

$$P(x) = 1/s\sqrt{2} \cdot \exp(-x^2/2s^2) \qquad (10)$$

with $P(x)$, the probability that amplitude x occurs, obeying a Gaussian distribution function with standard deviation s.

If Gaussian "white" noise (i.e., noise with a flat spectrum) is passed through a linear filter, we obtain "colored" noise. An important property of gaussian noise is that its amplitude distribution remains gaussian after linear filtering, independent of the filter characteristics.

With correlation techniques it could be proved that often the alpha rhythm has the characteristics of selectively filtered Gaussian noise. This is not only of interest theoretically but also for techniques of signal extraction. Moreover, the amplitude histograms of samples of alpha rhythm often show close resemblance to a Gaussian distribution.

Power

The power of an electrical signal is given by the average of V^2/R. In ERG and VEP, however, the resistances R are undefined and thus the powers are also unknown. It is common practice in electrophysiology to ignore R and to express "power" as amplitude squared of the Fourier components (in either μV^2 or V^2).

An important theorem in this respect is that of Parseval, which states that the time average of the square of a function $F(t)$ with period T equals half the sum of the squared amplitudes of its Fourier components:

$$1/T \cdot \int_{-T/2}^{+T/2} F^2(t)dt = 1/4 \cdot a_0^2 + 1/2 \cdot \sum_{n=1}^{\infty} A_n^2 \qquad (11)$$

This result is obtained when we consider the Fourier series representation of the function $F(t)$ and the *orthogonality* property of harmonic functions, i.e., that only integrals of the form

$$A_n A_m \int \cos 2\pi f_n t \cdot \cos 2\pi f_m t dt \qquad (12)$$

with $m = n$ will contribute to the time average. In this context orthogonality means that the normalized integrated product of any two harmonic functions with different frequencies, if sufficiently long, will equal zero. The only terms that contribute to the autocorrelation are then the squares of the amplitudes at each frequency (see the previous section).

In the case of *nonperiodic* signals, e.g., noise or any transient, the discrete sum at the right side of Equation 11 becomes also an integral (similar to what was described in the section on the Fourier integral). ΣA_n^2 will change into a continuous function of frequency, i.e., $A^2(f)$. This is known as the "power spectrum" of $F(t)$.

Although we shall not go into detail considering the physical and mathematical problems regarding the representation of the (power) spectrum, some remarks may be made: as was discussed before, the natural restrictions to true periodicity mean that in practice no Fourier spectrum will exhibit pure spectral lines; therefore the energy or power will always be distributed over a finite frequency band. As a consequence, the power at an *exact* frequency will be zero. The problem is circumvented by taking not power but the power per infinitesimal frequency range df, "power density". Therefore instead of "power spectrum," the term *power density spectrum* is also often used. Since power is additive, the total power in the frequency domain is obtained by integration over the frequency range.

The advantage of power density (dimension V^2/Hz) as a measure in the frequency domain, especially if noise is involved, is found in the well-defined properties of power such as additivity. However, it must be realized that power seems to have no specific meaning as a measure of properties of electrophysiological phenomena; in any case, ERGs and VEPs are conventionally recorded as an amplitude function of time. This is one of the reasons why, if their frequency spectrum is wanted, often it is not the power spectrum that is given but the amplitude spectrum. In recent publications the term "amplitude density" is used analogously to power density with the "dimension" $V/\sqrt{Hz}$. This may be confusing, however, because "amplitude" is not additive along the frequency axis. Only Gaussian noise obeys strict rules; in other cases there will be ambiguity.

Cross-correlation and System Analysis

An important step in the analysis of systems can be made by using the cross-correlation function R (note we use the same symbol R as above) between the output g and the input f. This function is defined by the formula

$$R(\tau) = 1/2T \cdot \int_{-T}^{+T} f(t)g(t - \tau)dt \qquad (14)$$

for T going to infinity. It expresses the relation between two (time) functions f and g in the same way that the autocorrelation function expresses the relation of a function to itself. If the two functions f and g have nothing in common, the result will be zero. If they are identical, the cross-correlation function becomes the autocorrelation function defined before. A common (hidden) signal embedded in two independent sources of noise will emerge in the cross-correlation function if a long enough sample can be processed.

Cross-correlation is especially effective in identifying transport delays between two signals: the cross-correlation will exhibit a maximum at τ = transport delay. If both signals are subject to noise of different origin, this influences only the normalized value of the maximum and its statistical significance, but not its delay. If, for instance, the EEGs from the two hemispheres have a cross-correlation function differing from zero, it can be concluded that either one hemisphere influences the other or both are acting under a common influence. In the latter case there may be zero delay (i.e., the cross-correlation function will show a maximum at $\tau = 0$). However, such a maximum should in general be interpreted with due caution because in all recordings electrical cross talk can be expected. Even if the common (parasitic) signal is at a very low level, a long enough recording period may eventually produce a significant result at $\tau = 0$. Results with zero delay will have to be specially examined and the power spectra of the initial signals taken into account. Cross-correlation functions with a maximum different from $\tau = 0$ can in general be considered trustworthy. Note that, as observed before, the cross-correlation functions of any two harmonic functions of different frequencies will be zero due to the orthogonality principle.

An important application of the cross-correlation function arises when white noise is used as an input. When the input noise is correlated with the output of the system, the result is identical to the impulse response, and its Fourier transform directly yields the transfer function of the system. The reason is that both the impulse and white noise have the same flat spectrum. The difference between the impulse function and white noise is only found in their respective phase spectra. However, the phase *shifts* after transmission through the system are the same for *each* separate frequency, whether the input is noise or an impulse, and therefore, the results will indeed be identical. Usually band-limited noise is employed with a flat spectrum simulating white noise in the region of interest. In case of noise-modulated light, the effective modulation depth (determined by the standard deviation of the noise signal) is necessarily restricted to about 30%; otherwise noise peaks will (too) often produce prolonged black periods (virtual negativity of light).

The transfer function determined by this method is equal to the product of all transfer functions of the series of linear processes involved. It is fundamentally impossible to separate or to identify, for instance, low-pass and high-pass stages. This means that, for example, in visual physiology distal and proximal frequency-dependent processes (filters) as identified by linear analysis cannot be separated—neither can their sequence be determined.

The equality of the noise and impulse response functions is only generally valid in linear systems. Nonetheless, there is one important class of nonlinearities in which cross-correlation can be usefully applied: cross-correlating white noise input with the output of a system with one static nonlinearity (i.e., a nonlinearity that is frequency independent like a rectifier) yields the shape of the impulse response of the totality of the linear processes involved. This is due to the fundamental property of Gaussian noise, described above, that its Gaussian character is retained after linear filtering. Therefore, the input to the nonlinearity is also Gaussian independent of the linear distal filters mostly present. The only difference with the result in the absence of static nonlinearity is found in the absolute value of the cross-correlation function, which depends on the characteristics of the nonlinearity (Bussgang's theorem[1]). Together with other advanced techniques, this property of Gaussian noise has been successfully employed in VEP studies to determine the transfer functions of various stages in the system.[11]

In contrast, however, to the equivalent use of noise and impulse functions in linear system analysis, in *nonlinear* systems the response to an impulse input, e.g., a flash, will in general *not* represent the impulse response of the linear elements. Already the reversing of the polarity of the impulse will give rise to different results, which makes interpretation difficult if not impossible. For instance, if there is an early nonlinearity like strong saturation, a "positive" unit impulse will yield a different result from a "negative" one; these differences may become crucial when the nonlinearity is interleaved with linear elements.

AVERAGING

Although Fourier analysis has become very popular in visual electrophysiology due to the wide availability of simple programs and personal computers, averaging can still be considered the method of choice in recording weak responses. Averaging is conceptually much simpler than Fourier analysis and is, in principle, maximally effective in noise reduction.

The Stimulus for Averaging

Classic averaging is performed by synchronizing the start of data collection with the stimulus and measuring (digitally) the response at a number of consecutive intervals that are then saved into a buffer in the computer's memory. The process is repeated, and in consecutive stimulus periods the amplitudes of corresponding samples are added by a computer. The stimulus can be periodic or not, as long as the period of interest is kept in synchrony with the stimulus and does not exceed the shortest stimulus interval. Averaging is in fact identical to cross-correlating a signal (ERG or VEP) with a chosen number of impulses of unit size. It is a linear method and as such also subject to treatment by using Fourier theory. (Actually, periodic averaging is equivalent to a filtering procedure that only allows frequencies belonging to the fundamental period and its harmonics to pass through while rejecting all other frequencies; furthermore, the "filter sharpness" of this process increases with the number of responses recorded.)

The stimulus in the case of averaging should have a stable amplitude and, in principle, should also be periodic. Otherwise, dynamic interactions may be a confounding factor. Strong adaptation effects have been described by Jeffreys,[6] from which it also follows that for transient stimulation there is a lower limit to the stimulation period.

Improving the Signal-Noise Ratio With Averaging

Averaging originally was based on the assumption that responses are stable and noise reasonably Gaussian. From this it follows that the signal-to-noise improvement (expressed in power) is proportional to the number of periods added. Because VEPs are characterized by amplitude and shape rather than by power, the rule of thumb is that noise amplitude is reduced relative to the response by the square root of n (the number of intervals added).

In relation to the above, there rises an important question: *What is the signal-to-noise ratio?* In a strict sense the signal-to-noise concept makes only sense for two Gaussian processes, one of which is considered to be the "signal" and the other the "noise" and the other being the superposition of harmonic signals and noise. A requirement is then that figures of signal-to-noise ratio be expressed per frequency band because at some frequencies the signal may be larger than the noise and at other frequencies this may be the other way around. Especially in the case of transient responses, the signal-to-noise concept can be difficult to apply and can easily give rise to ambiguities, for instance, when the frequency spectrum of the transient is very different from that of the noise. Suitable filtering may reduce such problems.

When averaging, both the frequency spectrum of the noise and the stimulus repetition rate may play an important role. Often there is a strong alpha rhythm of, say, 10 Hz with a high selectivity. If the stimulus rate were to be 1, 2, 5, or 10 Hz, the effective noise would be relatively amplified depending on the selectivity of the alpha process and the sweep period chosen: the lower the rate, the lesser the enhancing. A signal-to-noise ratio calculated on the basis of *total* power would then be highly overestimated, although the improvement will still go with $\sqrt{n}$. When the stimulus period is chosen such that it just fits an odd multiple of half a period of the alpha frequency, the recording interval will fall between alternating polarities of the alpha rhythm. Because of this, the result will be much improved: the effective signal-to-noise ratio is increased because the alpha rhythm will tend to cancel in successive sweeps.[11]

Response Fluctuations

VEPs and ERGs are subject not only to noise contamination but also to inherent fluctuations. There is a fundamental difference between the disruptive effects of additive noise and those due to variability of the response itself. If, as sometimes has been described, response and noise interact, then simple rules cannot be given. Theoretically there are two main types of response irregularities: latency jitter and amplitude fluctuations. Because their effects are comparatively small in routine measurements, they will only be treated briefly.

Latency jitter of the response or of parts of it will lead to smoothing during averaging. Actually, time jitter is the most effective low-pass filter that can be physically realized. Techniques have been described to implement adaptive filtering that counteracts these smoothing effects.[19] If the responses occur in clusters with different latencies, interesting methods exist to separate these clusters.[8] Analysis of conventional pattern VEPs in our laboratory by using sophisticated filtering has shown, however, that even near threshold the latency spread was not more than a few milliseconds. Concerning amplitude fluctuations, Dagnelie et al.[3] have found fluctuations on the order of 25% for a 50% modulated, 20 Hz luminous stimulus. In our own experience with recordings between 100 and 200 sweeps, reproducibility of pattern responses is generally very good.

An instrumental artifact is caused by the property that in certain TV stimulators the stimulus is not synchronized with the TV frame rate in order to prevent pickup from the mains or VEPs to TV flicker.[18] Since mostly only the central part of the TV screen is fixated (or as such has a dominant position), this will cause jitter of the stimulus on the order of 20 ms, equivalent to low-pass filtering using a "square" unit impulse response of 20-msec duration (in case of 50 Hz mains). Although this is not unequivocally translatable in a frequency cutoff, it means approximately an attenuation of 3 dB at a frequency of 22 Hz.

In some cases, one might wish to extract individual responses. Those interested in suitable techniques, like *a posteriori* filtering, are referred to the extensive literature covered and discussed by Lopes da Silva.[7]

Practical Considerations on Averaging

In our type of experiments time is often at a premium. Therefore stimulus period and sweep time should be strictly coupled so that there are no loose periods in between. This self-evident facility is often lacking in simple averagers, or no attention has been paid to it in computer programs. In the case of pattern reversal responses, it is advisable to record two responses per sweep. This allows one to check the symmetry of the stimulus while, at the same time, it gives an impression of the stability of the experimental situation. Early artifact rejection can also be strongly recommended.

SYNCHRONOUS AMPLIFICATION

Related to cross-correlation and to Fourier analysis is the technique of synchronous detection. (Lock-in amplifiers are a technically different realization of synchronous detection with the same advantages and disadvantages.) This technique is especially useful when repetitive stimulation at higher frequencies is employed: so-called steady-state experiments. The principle of the method is multiplication of the signal that contains the expected periodic response, with two harmonic functions of the same frequency and 90 degrees phase shift, i.e., a sine and a cosine. The outcome is smoothed (integrated) over a chosen time, e.g., by a resistor-capacitor (RC) filter, but separately for the sine and cosine. After this, the square root of the sum of squares is taken and recorded. For long RC times this approximates performing Fourier analysis on *one* frequency only: that of the fundamental of the stimulus. In other words, one is computing the Fourier series' coefficients a_1 and b_1, and from these the amplitude A is computed in real time. Note that any smoothing should be done before squaring a and b. This optimizes noise reduction. One can also obtain the (tangent of the) phase by using $tg\Phi = b/a$, although this *is not* common practice.

In most commercial synchronous amplifiers the multiplication is in fact replaced by synchronous rectification (Figure 29–17). In this figure it is supposed that the stimulus evokes a sinusoidal response that is passed through a synchronized full-wave rectifier. This is equivalent to alternately multiplying the signal by +1 and −1 during each half-period. In the top half of the figure, the sign of the multiplier changes when the sine wave crosses zero, and the sum of the shaded areas is maximal. When the sign change is delayed by 90 degrees (lower part), the net result is zero. Other phases are usually encountered in practice since there is no telling, a priori, the phase angle (or delay) between stimulus and response. Then the two channels will both record a response of which the sizes will depend on that angle. An estimate of the response amplitude, however, is independent of the phase. The further procedure is the same as described above for true multiplication.

An important consequence of using this technique is that not only will the fundamental component of the response determine the result, but also odd harmonics will because the multiplying function itself contains these harmonics. Therefore, the results of synchronous amplification are only unambiguous when the response itself is nearly a sine wave, as is generally the case at high stimulation frequencies. It should be realized that this type of simple quantification is at the cost of losing all information about possible components in the VEP.

Synchronous techniques are also treacherous at low frequencies when the fundamental is not the main contributor anymore (see the response at four reversals per second in Fig 29–18); low-frequency cutoff may even be suggested simply because the period extends beyond the most prominent part of the response; in other words, the contribution of the fundamental becomes smaller when the frequency is lowered.

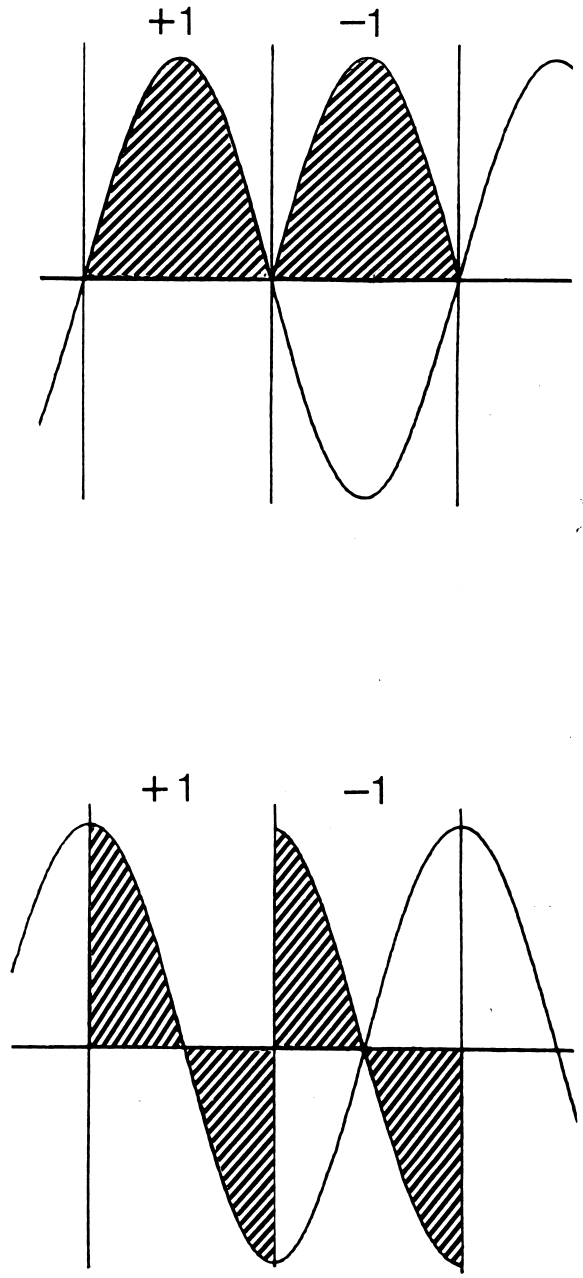

FIG 29–17.
Scheme of synchronous detection.

Note that synchronous detection is governed by the stimulus interval and that with pattern reversal the reversal rate is twice that of the luminance modulation of the elements. Therefore, the detection procedure or the multiplication must be governed by the second harmonic of the modulation frequency of

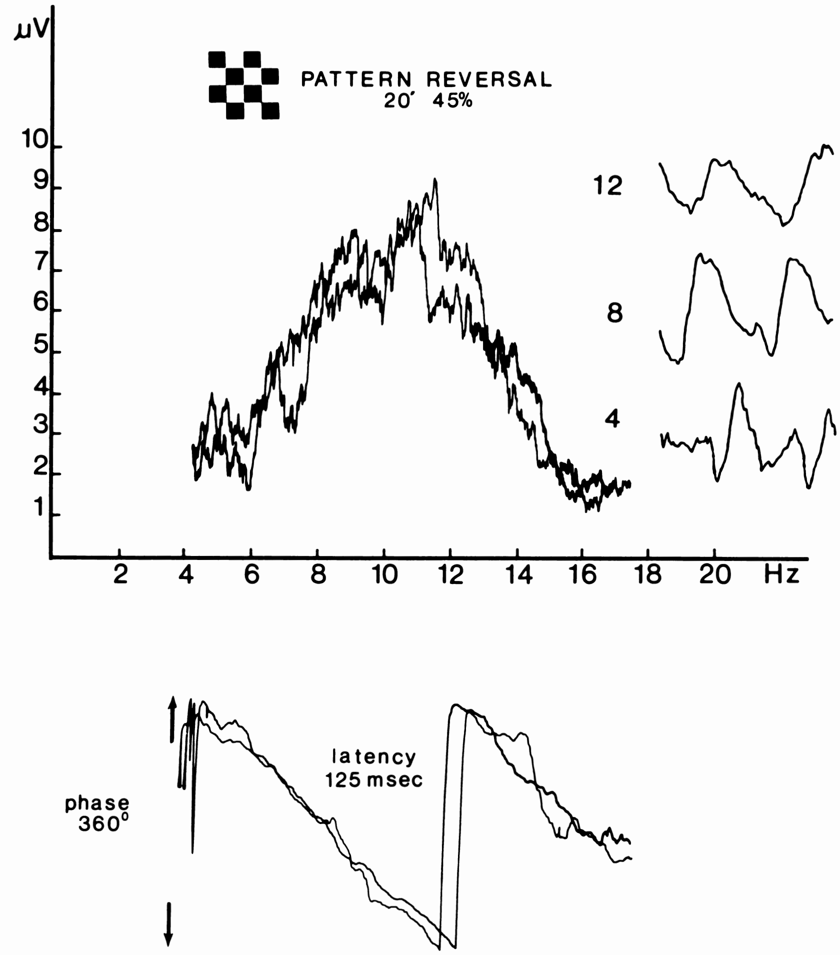

FIG 29–18.
Continuous recording with synchronous detection of the amplitude (*upper* trace) and phase (*lower* trace) of the response to a reversing checkerboard at various frequencies. Two successive averaged evoked potentials at three frequencies are inserted. From approximately 8 reversals per second on, the fundamental dominates. At 4 Hz the higher harmonics become stronger but do not contribute to the synchronously detected response, which contains mainly the fundamental. Therefore, the smaller response at low frequencies is an instrumental artifact. However, latency determination by means of phase shift as a function of frequency is still valid. (From Van der Tweel LH, Estévez O, Strackee J: Measurement of evoked potentials, in Barber C (ed): *Evoked Potentials.* Lancaster, England MTP Press, Ltd, 1980, pp 19–41. Used by permission.)

the two sets of elements. This is the reason why in the literature the response is often characterized as "second harmonic." The term is not only confusing but even wrong in the case of pure contrast responses because the response has evidently the same frequency as the true stimulus, i.e., that of the reversal itself. There is only sense in talking about second harmonics of the modulation frequency if the responses are governed mainly by luminance, as is the case with large elements or probably in part of the pattern ERGs.

The advantages of synchronous detection are those of simplicity and the possibility of continuous recording that they allow. The signal-to-noise improvement is dependent on the smoothing time constant and can be arbitrarily large at the cost of time resolution. This is sometimes expressed as equivalent bandwidth. For a similar noise reduction the consumption of time for averaging and synchronous amplification is comparable. For automatic tracking procedures, however, synchronous amplification is indeed a very appropriate technique.

CONCLUSION

We remarked in the introduction about the need to reconsider concepts such as "threshold." The point is that, even if signal-to-noise improvement may proceed comparatively slowly, the only restriction in improving the sensitivity of a recording is that imposed by the endurance of the subject. In practice, it has proved to be possible under favorable conditions to measure VEPs to modulated light at less than 5% of the psychophysical threshold.[12] The reason is that the computer memory by far exceeds the integration time of flicker perception and also the electrodes cover a much larger visual area than contributes to sensation. In such situations, it will depend on the patience of experimenter and subject what can be considered a nonrecordable response. It is clear that an unambiguous statement will be that at a noise level of x μV no response could be recorded, but this does not give an indication whether recording for a longer period would have yielded a VEP or that a "hard" threshold has indeed been passed, as is common with pattern evoked potentials. Probably the earliest example demonstrating a true electrophysiological threshold is the work of Campbell and Maffei,[2] where the objective threshold and the extrapolated electrophysiological one were shown to coincide very well. It is therefore recommended that the experimental technique used be stated exactly in order to enable judgment by the reader.

Whatever techniques are employed, Fourier theory is as important for the quantification of the responses themselves as it is for the signal analytical description of the stimulus-response relation. As was already stated, linear system analysis is exhaustive, and there remains only the question of choosing the most practical solution. Nonlinear system analysis has no general recipes, but there are groups of systems, e.g., those with one static nonlinearity that are accessible to systematic analysis. But also in those cases Fourier stands central!

REFERENCES

1. Bussgang JJ: Crosscorrelation functions of amplitude distorted Gaussian signals. *MIT Research Laboratory Electronics* Technical Report 216. Cambridge, Mass, MIT Press, 1952.
2. Campbell FW, Maffei L: Evidence for the existence of orientation and size detectors in the human visual system. *J Physiol* 1970; 207:635–652.
3. Dagnelie G, van den Berg TJTP, Reits D: Unfamiliar effects of flicker on the human EEG. *Doc Ophthalmol Proc Ser* 1977; 15:173–178.
4. Enroth-Kugel C, Robson JG: Functional characteristics and diversity of cat retinal ganglion cells. *Invest Ophthalmol Vis Sci* 1984; 25:250–267.
5. Estévez O, Cavonius CR: Low frequency attenuation in the detection of gratings: Sorting out the artifacts. *Vision Res* 1976; 16:497–500.
6. Jeffreys DA: The physiological significance of pattern visual evoked potentials, in Desmedt JE (ed): *Visual Evoked Potentials: New Developments*. Oxford, England, Clarendon Press, 1977, pp 134–167.
7. Lopes da Silva FH: Event related potentials: Methodology and quantification, in Niedermeyer E, Lopes da Silva FH (eds): *Electro-Encephalography*. Baltimore, Urban & Schwarzenberg, 1982, pp 655–664.
8. Pfurtscheller G, Cooper R: Selective averaging of the intercerebral click evoked responses in man: An improved method of measuring latencies and amplitudes. *Electroencephalogr Clin Neurophysiol* 1975; 38:187–190.
9. Regan D: *Human Brain Electrophysiology*. New York, Elsevier Science Publishing Co, Inc, 1989.
10. Spekreijse H: Rectification in the goldfish retina: Analysis by sinusoidal and auxiliary stimulation. *Vision Res* 1969; 9:1461–1472.
11. Spekreijse H, Estévez O, Reits D: The physiological analysis of visual processes in man, in Desmedt JE (ed): *Visual Evoked Potentials: New Developments*. Oxford, England, Clarendon Press, 1977, pp 16–89.
12. Spekreijse H, van der Tweel LH: System analysis of linear and nonlinear processes in electrophysiology of the visual system. *Proc Kon Ned Akad van Wetensch* 1972; 75:77–105.
13. Van der Tweel LH: In Spekreijse H, van der Tweel LH (eds): *Spatial Contrast*. Amsterdam, Elsevier Science Publishers, 1977, pp 9–12.
14. Van der Tweel LH, Estévez O, Pijn JPM: Notes on Fourier methods in vision research. *Doc Ophthalmol Proc Ser* 1985; 37:439–452.
15. Van der Tweel LH, Estévez O, Strackee J: Measurement of evoked potentials, in Barber C (ed): *Evoked Potentials*. Lancaster, England, MTP Press, Ltd, 1980, pp 19–41.
16. Van der Tweel LH, Verduyn Lunel HFE: Human visual responses to sinusoidally modulated light. *Electroencephalogr Clin Neurophysiol* 1965; 18:587–598.
17. Van der Tweel LH, Visser P: Electrical responses of the retina to sinusoidally modulated light. In *Electroretinographia*. Acta Facultatis Medicae Universitatis Brunensis, Lekarska Fakulta University Brne, 1959, pp 185–196.
18. Van Lith GHM, van Marle GW, van Dok-Mak GTM: Variation in latency times of visually evoked cortical potentials. *Br J Ophthalmol* 1978; 62:220–222.
19. Verburg J, Strackee J: Phaseless recursive filtering applied to chestwall displacements and velocities, using accelerometers. *Med Biol Eng* 1974; 12:483–488.
20. Woody CD: Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Med Biol Eng* 1967; 5:539–553.