Electroencephalography and clinical Neurophysiology, 1991, 78: 378–388
© 1991 Elsevier Scientific Publishers Ireland, Ltd. 0013-4649/91/$03.50
ADONIS 001346499100094R

EEG 89161

A new statistic for steady-state evoked potentials

Jonathan D. Victor [a,b] and Joelle Mast [a]

[a] *Department of Neurology and Neuroscience, Cornell University Medical College, New York, NY 10021 (U.S.A.),*
and [b] *Laboratory of Biophysics, The Rockefeller University, New York, NY 10021 (U.S.A.)*

(Accepted for publication: 6 July 1990)

Summary Steady-state evoked potentials are often characterized by the amplitude and phase of the Fourier component at one or more frequencies of interest. We introduce a new statistic for the evaluation of these Fourier components. This statistic, denoted T^2_{circ}, is based on the same physiologic assumptions concerning the sources of variability of a Fourier component that are made in the use of the Rayleigh phase-coherence statistic as well as the standard T^2 statistic (Hotelling 1931) for multivariate data. However, the T^2_{circ} statistic also exploits the relationship between the real and imaginary components of Fourier estimates, which is not exploited by T^2, and utilizes amplitude information, which is ignored by the Rayleigh criterion. For these reasons, the T^2_{circ} statistic is more efficient than previously used criteria for detection and quantitation of steady-state responses, both in principle and in practice.

Key words: Fourier analysis; Signal detection; Visual evoked potentials

Steady-state sinusoidal analysis is of widespread and established use in neurophysiology, both for single-unit studies (Pringle and Wilson 1952; Hughes and Maffei 1966) and for visual evoked potentials (VEPs) (Van der Tweel and Verduyn Lunel 1965; Regan 1966). In principle (Regan 1989), a steady-state evoked potential is a repetitive evoked potential 'whose constituent discrete frequency components remain constant in amplitude and phase over an infinitely long time period.' Despite the obvious need to have efficient yet rigorous means to assess the variability of Fourier components estimated from experimental data, many previous investigators have simply adapted standard statistics from other contexts (Norcia and Tyler 1985; Picton et al. 1987; Stapells et al. 1987; Strasburger 1987).

These previous approaches ignore one or more important features of estimates of Fourier components. Methods based on the Rayleigh phase criterion (RPC) (Norcia and Tyler 1985; Picton et al. 1987) ignore amplitude information entirely. Methods (Picton et al. 1987) based on the T^2 statistic of Hotelling (1931) for multivariate data ignore relationships between the real and imaginary parts of Fourier components. Full utilization of amplitude information and the relationship between real and imaginary parts of Fourier components leads to a new statistic, T^2_{circ}, which is specifically designed for the analysis of variability of Fourier components.

We present a theoretical analysis which leads to the T^2_{circ} statistic and how it may be used to determine confidence limits and significance of differences between groups. We compare signal detection provided by the new statistic T^2_{circ} with that obtained with other statistical methods (T^2 and the RPC) in numerical simulations. Then, we compare application of T^2_{circ}, T^2, and the RPC to steady-state VEP data.

Methods

The evoked potential data used for practical assessment of the T^2_{circ} statistic consisted of the data of Mast and Victor (1991) collected from 8 adult normal subjects aged 25–35. We summarize the stimulation and recording methods here. The visual stimulus was a checkerboard which underwent sinusoidal counterphase modulation at fundamental frequencies (f) of 5.0 Hz, 7.5 Hz, and 10.0 Hz. Checks subtended 16 min, had a contrast $[(I_{max} - I_{min})/(I_{max} + I_{min})]$ of 0.4, and a mean luminance of 128 cd/m^2. The stimulation frequencies were chosen so that in all cases, an integral number (2, 3, or 4) of reversals were completed within an 0.2 sec period. One episode of data collection consisted of 150 such contiguous periods. Each subject sat for a total of

Correspondence to: Jonathan D. Victor, Cornell University Medical College, 1300 York Avenue, New York, NY 10021 (U.S.A.).

32 episodes (8 at each of the 3 temporal frequencies and 8 episodes of a static checkerboard). Digitized data were saved for later analysis *without* averaging, so that Fourier components could be calculated for arbitrary portions of each episode.

The T^2_{circ} statistic

Here we derive a new statistic, T^2_{circ}, to characterize the variability of a steady-state evoked potential (EP), as measured by its Fourier component. Our strategy is to begin with reasonable hypotheses concerning the sources of variability in the estimate of the Fourier component, and to deduce a statistic that is useful in measuring this variability.

A graphical representation. A steady-state EP experiment records many repeats of the response to a periodic stimulus. Since the steady-state EP is a periodic signal, it may be described by its Fourier components. We consider here only the situation in which a single Fourier component of the response is to be estimated. At any particular frequency of interest, the Fourier component is a complex number z. Written as $z = r\,e^{\iota\theta}$, the amplitude r represents the size of the response and the phase θ represents the timing of the peak of the response relative to the stimulus cycle. Written equivalently as $z = x + \iota y$, the real quantities x and y represent the cosine and sine components of the response. The polar representation $z = r\ e^{\iota\theta}$ is more convenient for interpretation of responses, since it separates response size r and phase θ. However, the cartesian representation $z = x + \iota y$ is more suited for the present analytical purposes, since it makes explicit the notion that Fourier components may be considered as vectors in the (x,y)-plane.

An estimate of a Fourier component of an evoked potential may be obtained from one or more repeats of the stimulus. Individual estimates of the Fourier component, considered as vectors, form a cluster in the complex plane. The center of the cluster is a pooled estimate of the response, and the scatter of the cluster describes the reliability of that estimate.

This is illustrated in Fig. 1. In each part, the heavy vector along the x-axis represents a 'true' response (i.e., the response that would be recorded in the absence of variability). Individual response estimates are represented as thin-line vectors and were constructed by adding random vectors to the 'true' response. The mean of the 4 response estimates is indicated by a vector of intermediate thickness. In general, as is illustrated in this figure, the mean of the response estimates is not equal to the 'true' response. However, the scatter of the individual estimates provides an estimate for the reliability of the sample mean. As is shown in this figure, the 'true' response typically lies within the cluster of response estimates. Thus, estimation of the amount of scatter within the cluster should provide an index of how much the 'true' response may deviate from the mean of the observed responses. This is the basis of the T^2_{circ} statistic.

The critical requirement for the derivation of the T^2_{circ} statistic is that the errors in the real part x and the imaginary part y of the Fourier component $z = x - \iota y$ are *independent* quantities distributed according to gaussian distributions of *equal* variances. We will denote this common variance by V.

This is the key point of departure between the T^2_{circ} statistic and the T^2 statistic as used by Picton et al. (1987). The T^2 statistic assumes no information about the variances and covariances of these quantities, while the T^2_{circ} statistic assumes that the variances are equal and the covariance is zero. Graphically, the T^2 statistic considers the cluster of estimates of Fourier components to form an ellipse, whose axes and orientation are unknown. The hypothesis of equal variances and zero covariance corresponds to the notion that the cluster of estimates of Fourier components is circularly symmetric. As we shall see, this increases the number of degrees of freedom and produces a more efficient statistical test.

In a previous analysis (Mast and Victor 1991), we showed that sufficient conditions for this hypothesis to be true are: (i) the 'noise' (i.e., the background electroencephalogram (EEG)) is a gaussian process independent of the evoked potential (EP), and (ii) that EEG 'noise' and the EP signal combine additively. We also showed that although measurable departures from these assumptions occurred, the real and imaginary parts of Fourier estimates nevertheless were quantities with equal variances and zero covariance.

Derivation of the T^2_{circ} statistic. In a typical physiological application, we have a set of M estimates of a Fourier component from independent data sets and wish to know whether these estimates are consistent with an a priori value. Let us denote the M estimates of this Fourier component by $z_1, z_2, \ldots z_M$, their empirical mean value by $\langle z\rangle_{est} = (\Sigma\ z_j)/M$, and an a priori hypothetical value by ζ. We assume no information concerning the variability of the estimates z_j other than the general considerations above, and their experimental scatter. The quantities z_j, $\langle z\rangle_{est}$, and ζ are complex numbers, whose decompositions into real and imaginary parts are given by $z_j = x_j + \iota y_j$, $\langle z\rangle_{est} = \langle x\rangle_{est} + \iota\langle y\rangle_{est}$, and $\zeta = \xi + \iota\eta$.

If indeed the set of experimental estimates are drawn from a population whose mean is equal to ζ, then there are two independent estimates of the population variance V of real and imaginary parts. The first estimate, V_{indiv}, is derived from the scatter of the individually determined components z_j about their mean. Each of the deviations $x_j - \langle x\rangle_{est}$ and $y_j - \langle y\rangle_{est}$ provide one such estimate. There are $2(M - 1)$ degrees of freedom, rather than 2 M degrees of freedom, since the means of the x_js are constrained to be $\langle x\rangle_{est}$, and the means of the y_js are constrained to be $\langle y\rangle_{est}$. Thus, one estimate for the population variance V is

$$V_{indiv} = \frac{1}{2(M-1)}\Sigma\left[(x_j - \langle x\rangle_{est})^2 + (y_j - \langle y\rangle_{est})^2\right]$$
$$= \frac{1}{2(M-1)}\Sigma|z_j - \langle z\rangle_{est}|^2 \qquad (1)$$

Note that this estimate is *independent* of the assumed population mean ζ.

The second estimate depends on the assumed population mean ζ. Since $\langle z\rangle_{est}$, the sample mean, is the mean of M independent estimates, both its real and imaginary parts have variance V/M about the population mean ζ. Real and imaginary parts of $\langle z\rangle_{est}$ are independently distributed, and each is *unconstrained* by the estimate of V_{indiv}, because the estimate V_{indiv} is independent of this mean. Thus, we have a second estimate of the variance V:

$$V_{group} = \frac{M}{2}\left[(\langle x\rangle_{est} - \xi)^2 + (\langle y\rangle_{est} - \eta)^2\right]$$
$$= \frac{M}{2}|\langle z\rangle_{est} - \zeta|^2 \qquad (2)$$

Under the hypothesis that the experimental data z_j are samples of a population whose mean is ζ, each of the quantities V_{group} and V_{indiv} are estimates of the variance V derived from independent quantities. Therefore, the ratio V_{group}/V_{indiv} is distributed according to the F distribution (Sokal and Rohlf 1969), with 2(M − 1) degrees of freedom for the denominator, and 2 degrees of freedom for the numerator.

To maintain a close analogy with the T^2 statistic (Anderson 1958), we define a statistic T^2_{circ}, which is equal to the variance ratio V_{group}/V_{indiv} normalized by the number of observations M:

$$T^2_{circ} = \frac{1}{M}\frac{V_{group}}{V_{indiv}}$$
$$= (M-1)\frac{|\langle z\rangle_{est} - \zeta|^2}{\Sigma|z_j - \langle z\rangle_{est}|^2} \qquad (3)$$

The above argument shows that for M independent estimates of Fourier components z_j drawn from a sample of assumed mean of ζ, $M\cdot T^2_{circ}$ is distributed according to $F_{[2,2M-2]}$.

When is a signal present? The basic application of the T^2_{circ} statistic is the determination of whether an observed set of Fourier components z_j is consistent with random fluctuations alone, or, conversely, whether this set of observations implies (to within a given confidence level) that a signal component is present. The null hypothesis that no signal is present is that $\zeta = 0$ in equation (3). The null hypothesis should be rejected (at a pre-specified confidence level) if the calculated value of T^2_{circ} corresponds to a value of F sufficiently far into its distribution's tail. Thus, a criterion that a set of Fourier estimates is *inconsistent* with random scatter about zero at confidence level 1 − p is:

$$|\langle z\rangle_{est}|^2 > \left[\frac{F_{(1-p)[2,2M-2]}}{M}\cdot\frac{\Sigma|z_j - \langle z\rangle_{est}|^2}{(M-1)}\right] \qquad (4)$$

Calculation of confidence regions. For this application, we consider the null hypothesis that the observed data z_j of sample mean $\langle z\rangle_{est}$ and unknown variance are drawn from a population of mean ζ. Since this null hypothesis implies that $M\cdot T^2_{circ}$ is distributed according to $F_{[2,2M-2]}$, a value of $M\cdot T^2_{circ}$ which lies on the tail (by some predefined criterion) of this F distribution allows rejection of the null hypothesis that the population mean is ζ.

Since according to (3), T^2_{circ} depends on the hypothetical population mean ζ through the difference

$$|\langle z\rangle_{est} - \zeta|,$$

we may calculate a confidence region for the population mean as follows: given a confidence level 1 − p, first calculate (from the $F_{[2,2M-2]}$ distribution) the critical value $T^2_{circ(1-p)}$. All values of ζ that lead to a larger calculated value for T^2_{circ} would be less likely to result in the observed sample mean. Thus, the confidence region for the population mean ζ is a circle centered at the sample mean $\langle z\rangle_{est}$, whose radius is determined by $T^2_{circ(1-p)} = \frac{1}{M}F_{(1-p)[2,2M-2]}$:

$$|\langle z\rangle_{est} - \zeta|^2 = \frac{F_{(1-p)[2,2M-2]}}{M}\cdot\frac{\Sigma|z_j - \langle z\rangle_{est}|^2}{(M-1)} \qquad (5)$$

Difference between two means. The T^2_{circ} statistic may be used to determine whether two experimental measurements of Fourier components are significantly different. We assume that we have made M_1 measurements $z_{1,j}$ under one set of conditions, and M_2 measurements $z_{2,j}$ under another set of conditions. We assume that all measurements are independent, and that the populations from which the $z_{1,j}$ and the $z_{2,j}$ are drawn have real and imaginary parts of equal, but unknown, variance V. The null hypothesis is that the two means of the two populations, ζ_1 and ζ_2, are equal. As before, under the null hypothesis, there are two independent estimates of the common population variance V. The first estimate is based on the scatter of the individual estimates $z_{k,j}$ about their sample means $\langle z_k\rangle_{est}$:

$$V_{indiv} = \frac{1}{2(M_1 + M_2 - 2)} \times \left[\Sigma|z_{1,j} - \langle z_1\rangle_{est}|^2 + \Sigma|z_{2,j} - \langle z_2\rangle_{est}|^2\right]. \qquad (6)$$

There are $2(M_1 + M_2 - 2)$ degrees of freedom associated with this estimate. The second estimate is based on the difference between the sample means themselves:

$$V_{group} = \frac{M_1 M_2}{2(M_1 + M_2)}|\langle z_1\rangle_{est} - \langle z_2\rangle_{est}|^2. \qquad (7)$$

There are two degrees of freedom associated with this estimate.

As before, under the null hypothesis $\zeta_1 = \zeta_2$, the ratio V_{group}/V_{mean} is distributed according to the F distribution, but now with 2 degrees of freedom for the numerator, and $2(M_1 - M_2 - 2)$ degrees of freedom for the denominator. The corresponding T^2_{circ} statistic is

$$T^2_{circ} = \frac{M_1 + M_2}{M_1 M_2} \cdot \frac{V_{group}}{V_{indiv}}$$
$$= (M_1 + M_2 - 2)$$
$$\times \frac{|\langle z_1 \rangle_{est} - \langle z_2 \rangle_{est}|^2}{\Sigma |z_{1,j} - \langle z_1 \rangle_{est}|^2 + \Sigma |z_{2,j} - \langle z_2 \rangle_{est}|^2} \qquad (8)$$

where $(M_1 M_2 / M_1 + M_2) \cdot T^2_{circ}$ is distributed according to $F_{[2, 2M_1 + 2M_2 - 4]}$. Thus, if the calculated value (8) of T^2_{circ} corresponds to the tail of the F distribution, the null hypothesis that the population means ζ_1 and ζ_2 are equal may be rejected.

Neighboring data segments. The above derivations assumed that the estimates of individual Fourier components are from independent data segments. In principle, this would require that data segments are sufficiently well separated so that fluctuations in the frequency content of one data segment are independent of those in another segment. For data segments that are taken from experimental runs well separated in time, this is a reasonable assumption. However, it is often desirable to estimate Fourier coefficients from neighboring data segments within a run. Provided that the data segments are sufficiently long, the approximation of independence remains valid. The criterion of 'sufficiently long' is that the power spectrum of the underlying noise is approximately flat in a frequency interval of size $2\pi/L$ surrounding the frequency of the Fourier component of interest, where L is the length of the data segment (Mast and Victor 1991, Appendix).

One way to determine whether the power spectrum is flat at a resolution of $2\pi/L$ is to compare spectral estimates obtained from data segments of length L (and frequency resolution $2\pi/L$) with spectral estimates obtained from longer data segments of length $L' > L$ (and finer frequency resolution $2\pi/L'$). Net curvature of the power spectrum would result in a failure of the estimates obtained from the shorter segments to agree with estimates obtained from the longer segments. Con-

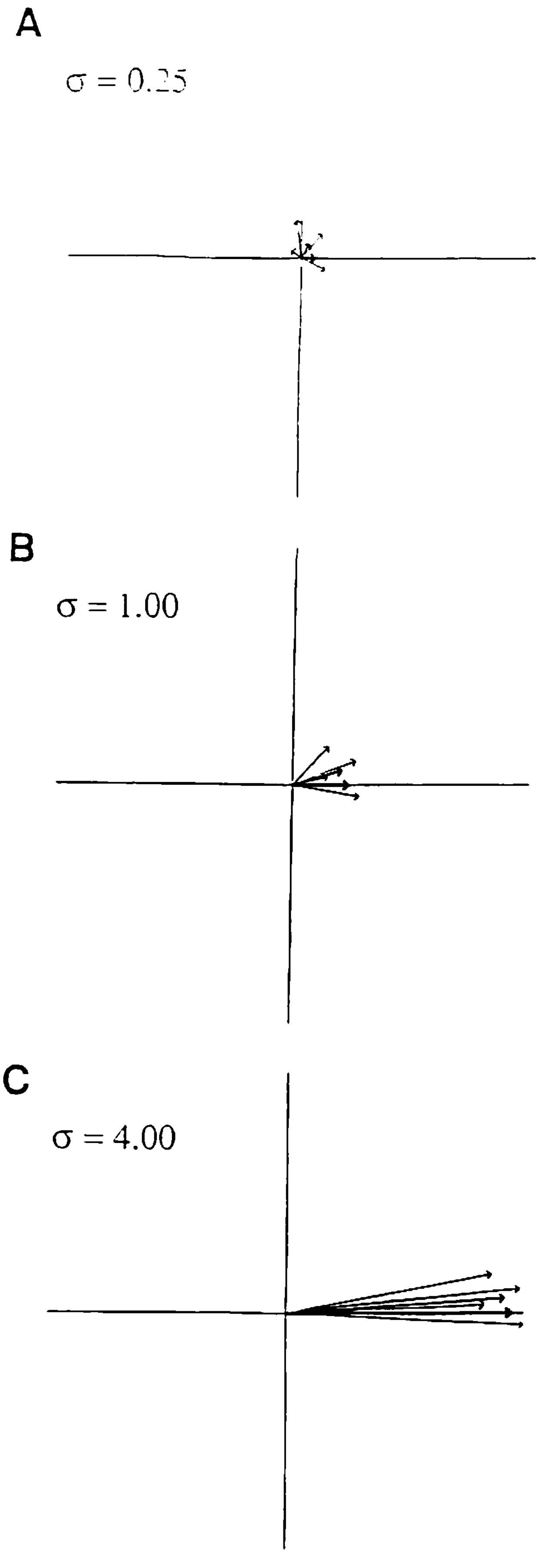


Fig. 1. Graphical representation of the relationship between individual estimates of a Fourier component, their sample mean, and a 'true' value that would be obtained if no variability were present. The heavy vector represents a response that would be recorded if no variability were present. Individual response estimates are represented as thin-line vectors. The mean of the 4 response estimates is indicated by a vector of intermediate thickness. In part A, the signal-to-noise ratio (SNR) σ is 0.25 : 1. In part B, the SNR σ is 1 : 1. In part C, the SNR σ is 4 : 1.

versely, if the spectral estimates obtained at different resolutions are mutually consistent, then it is reasonable to conclude that the condition of approximate independence is met. This is the approach used below, in the section entitled 'Division of a segment into subsegments.'

This problem concerning failure of contiguous data segments to be independent is not specific to the T^2_{circ} statistic, but also applies when the T^2 statistic or the RPC is used for contiguous data segments. If the above frequency-domain criterion is not met, the T^2_{circ} statistic is still useful in a practical sense, but the number of degrees of freedom for comparison with tabulated F values should be adjusted by the effective number of *independent* data segments.

Results

Numerical simulations

In this section, we compare the efficiency of 3 approaches to signal detection: the T^2_{circ} statistic, the T^2 statistic (Anderson 1958), and the RPC (Mardia 1972). We assume that the signal to be detected has an amplitude (in the absence of noise) of σ. For convenience, we may take the phase of this response to be zero (as represented by the heavy vectors in Fig. 1). To simulate estimation of this response from N data segments with additive noise, we added random vectors to this response (as represented by the light vectors in Fig. 1). The random vectors were drawn from a symmetric 2-dimensional gaussian distribution of unity variance, so that the signal-to-noise ratio (SNR) of each estimate was equal to σ. For each statistical test, the criterion for signal detection was that the set of N noisy estimates was *in*consistent with scatter about zero. The criterion was set so that false positives would occur 5% of the time. For the T^2_{circ} statistic, we used the criterion (4); for T^2, we used the method of Anderson (1958); and for the RPC, we used the method of Mardia (1972). This procedure was repeated for 10.000 independent simulated sets of N estimates, for a selection of SNR σ and a range of the number N of estimates.

Results are displayed in Fig. 2. For the low SNR of $\sigma = 0.25$ (Fig. 2A), the T^2_{circ} and T^2 statistics perform similarly: approximately $N = 40$ data segments are needed to detect a response half of the time. The RPC is inferior, requiring approximately $N = 60$ data segments

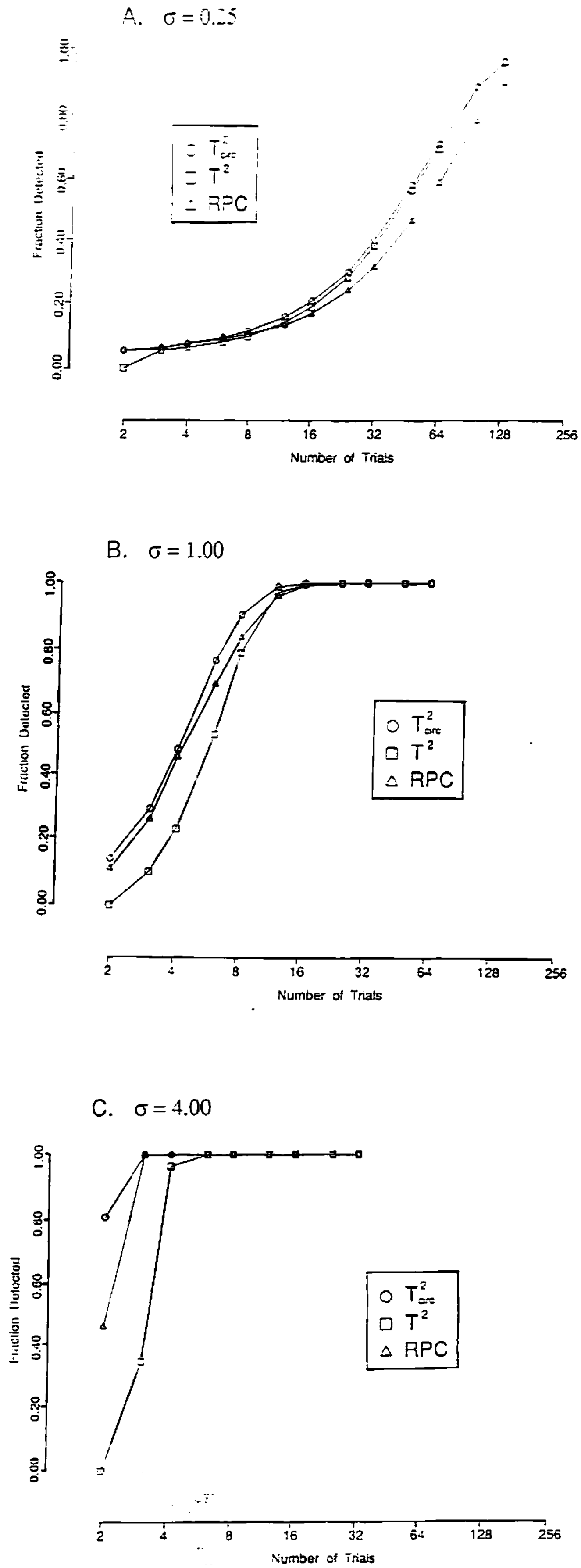


Fig. 2. The frequency with which a steady-state evoked response signal is detected, as a function of the number of segments N and the SNR σ. In A, the SNR σ is 0.25 : 1. In B, the SNR σ is 1 : 1. In C, the SNR σ is 4 : 1. Open circles, T^2_{circ} statistic; open squares, T^2 statistic; open triangles, RPC. These data are obtained from numerical simulations, as described in the text.

to detect a response half of the time. For an intermediate SNR of $\sigma = 1$ (Fig. 2B), there are 3 regimes of interest. With a small number of data segments ($N = 4$ or less), the T^2_{circ} statistic and the RPC are similar in detection rate, and T^2 is inferior. With a large number of data segments ($N = 24$ or more), all 3 methods detect a response essentially all of the time. In the intermediate range, the T^2_{circ} statistic is superior to the other two statistics: with 8 data segments, signal is detected 90% of the time by T^2_{circ}, 83% of the time by the RPC, and 79% of the time by T^2. For the high SNR of $\sigma = 4$ (Fig. 2C), T^2_{circ} and RPC detect a response essentially all of the time with $N = 3$ or more data segments. However, comparable performance requires $N = 4$ data segments if the RPC is used, and $N = 5$ or 6 data segments if the T^2 statistic is used.

In Fig. 2, we examined the performance of these 3 statistics as more and more data segments of a fixed SNR are added. Next, we consider the performance of these 3 statistics as a single data segment is subdivided into more and more subsegments. This is meant to mimic the process of determining whether a Fourier component is present in a particular segment of data by comparing estimates of Fourier components in N equal subsegments of the original segment. In the limit that these subsegments are independent, the SNR σ' from each subsegment is related to the SNR in the entire data segment (σ) by $\sigma' = \sigma/\sqrt{N}$.

Performance of the 3 statistics as a function of σ and N is shown in Fig. 3. In all cases, performance improves as the number of subsegments is increased, and (as expected) reaches an asymptotic value as the number of subsegments N becomes large. In all cases, a similar asymptotic value is achieved for the T^2_{circ} and the T^2 statistics: approximately 9% detection with a SNR of $\sigma = 0.25$, 22% detection with a SNR of $\sigma = 1$, and 72% detection with a SNR of $\sigma = 4$. The asymptotic performance achieved by the RPC is inferior in all cases: approximately 8% detection with a SNR of $\sigma = 0.25$, 18% detection with a SNR of $\sigma = 1$, and 60% detection with a SNR of $\sigma = 4$. For a small number of subsegments, the T^2 statistic becomes inefficient, and the RPC is similar in efficiency to T^2_{circ}.

Thus, we see that for some conditions (low σ, high N), T^2 is more efficient than the RPC, and for others (high σ, low N), the RPC is more efficient than T^2. But in all conditions, the new statistic T^2_{circ} is at least as

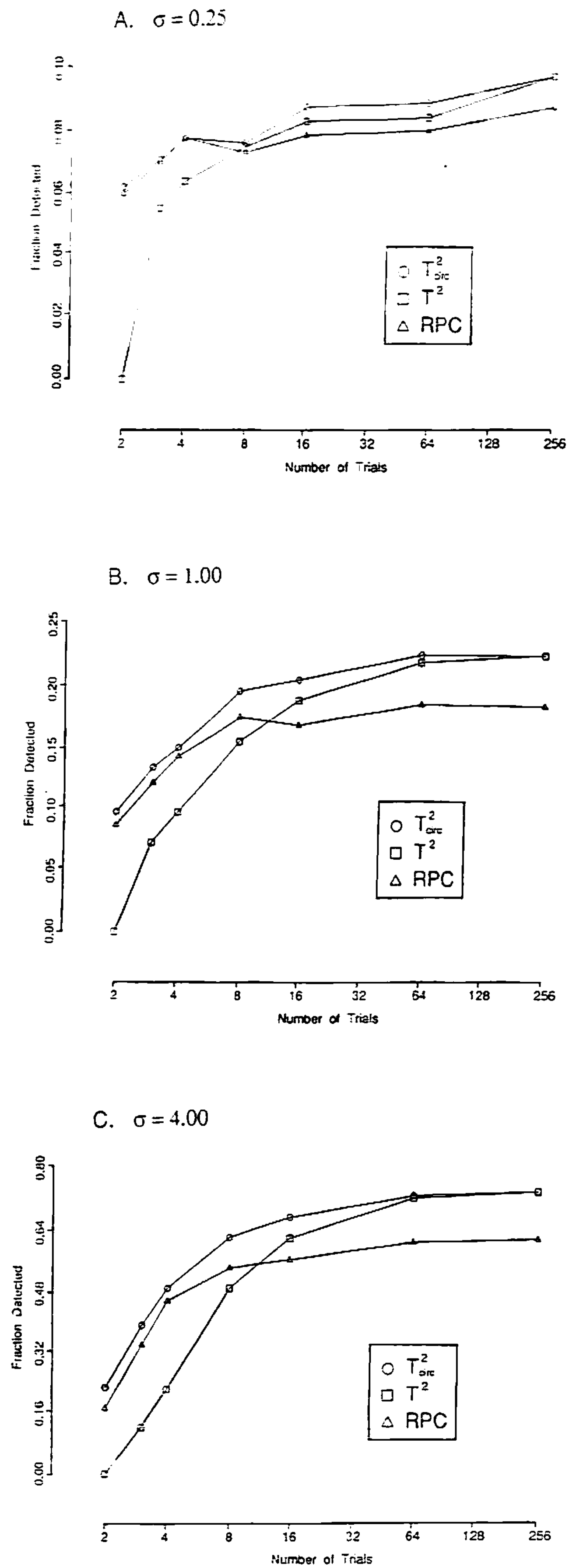


Fig. 3. The frequency with which a steady-state evoked response signal is detected, as a function of the number N of partitions of a data segment into subsegments, and the SNR σ of the entire segment. The SNR σ' from each subsegment is taken to be $\sigma' = \sigma/\sqrt{N}$. In A, the SNR σ is 0.25:1. In B, the SNR σ is 1:1. In C, the SNR σ is 4:1. Open circles, T^2_{circ} statistic; open squares, T^2 statistic; open triangles, RPC.

efficient as either of the two other approaches, and is often better than both. In the Discussion, we will interpret these findings in terms of number of degrees of freedom and optimal use of all of the information in the individual Fourier components.

Comparison of signal detection in VEP data

Next, we compare how these 3 statistics perform when applied to actual visual VEP data. We first consider how many data segments must be averaged together to detect a response, and later we consider the sizes of the confidence regions provided by the T^2_{circ} and T^2 statistics.

The VEP data we examined was elicited by contrast-reversing checkerboards in 8 normal subjects. Each subject viewed 8 presentations of the checkerboards, at fundamental frequencies of f = 5.006, 7.509, and 10.012 Hz. Each data segment was 29.96 sec long; however, for this analysis we used only the first third of each data segment to decrease the SNR. With the false-positive rate set at 5%, the average number of trials needed to achieve significance was 2.7 for T^2_{circ}, 4.3 for T^2, and 3.3 for the RPC. Clearly, the T^2_{circ} statistic is preferable in this situation. Because significance is achieved in a small number of trials, the fewer number of degrees of freedom in T^2 hinders its performance; it cannot even be applied unless there are 3 data segments to compare.

Division of a segment into subsegments

As seen from Fig. 3, division of a data segment into many subsegments improves the efficiency of any of the statistical procedures. However, all 3 statistics can only be applied if the estimates within individual subsegments are independent.

In Mast and Victor (1991), we presented a theoretical criterion for the minimum length L of independent data subsegments: in the neighborhood of the frequency of interest, the power spectrum of the noise process must be approximately flat over a frequency band width of 1/L. We showed in the Appendix that this minimum subsegment length is the same as the length L required for power spectral estimates to converge to their limiting value.

To apply this criterion to determine the minimum subsegment length required for approximate independence, we obtained spectral estimates from each subject from 8 segments of EEG recorded during fixation of an unmodulated checkerboard, according to equation (5) of Mast and Victor (1991). These spectral estimates will be denoted $\hat{S}(\omega,L_{max})$, where L_{max} is equal to the segment length, 29.96 sec. We also obtained spectral estimates $\hat{S}(\omega,L)$ from the first L seconds of data from each of these 8 segments. If L is sufficiently long so that spectral estimates have approached their limiting value, then the ratio $R(\omega,L) = \hat{S}(\omega,L)/\hat{S}(\omega,L_{max})$ will be distributed according to $F_{[14,14]}$.

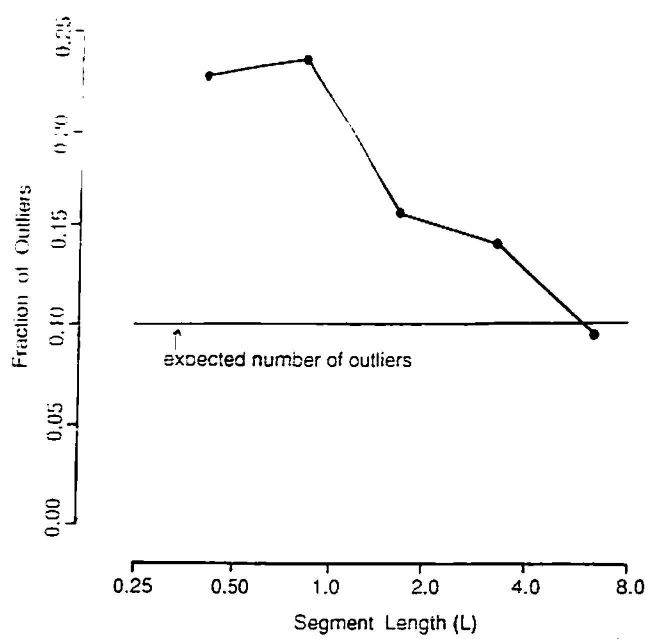


Fig. 4. The frequency with which values of the ratio $R(\omega,L) = \hat{S}(\omega,L)/\hat{S}(\omega,L_{max})$ lie in the two 5% tails of the corresponding F distribution. As L approaches the range 3–6 sec, the expected number of outliers (10%) is found.

The number of degrees of freedom is 2(N − 1), where N is the number of segments. In using the F statistic, we make the approximation that the numerator and denominator are independent. Since the data used for the spectral estimate $\hat{S}(\omega,L)$ is contained in the data used for the spectral estimate $\hat{S}(\omega,L_{max})$, this approximation holds only if $L \ll L_{max}$.

The ratio $R(\omega,L) = \hat{S}(\omega,L)/\hat{S}(\omega,L_{max})$ was calculated for 16 frequencies ω in the range 2.5–40 Hz, for each of the 8 subjects. This provided 128 values of $R(\omega,L)$ for each subsegment length L. In Fig. 4, we have plotted the frequency that the ratio $R(\omega,L)$ lies in the two 5% tails of the corresponding F distribution. For very short values of L, there are between 2 and 3 times as many values in the tails as would be expected. However, when L exceeds 3 sec, then (as expected) only approximately 10% of values lie in the two 5% tails. Since this means that spectral estimates have approached their limiting values, it follows that adjacent subsegments of length 3–6 sec may be regarded as independent.

The above estimate rests on the assumption that the power spectrum of the EEG is independent of whether or not a periodic stimulus is present. This is only approximately true (Mast and Victor 1991). Furthermore, even if this estimate is accepted for this particular paradigm, it is possible that the minimum subsegment length will be different in another EP paradigm. For these reasons, we consider a more practical, though less rigorous, approach.

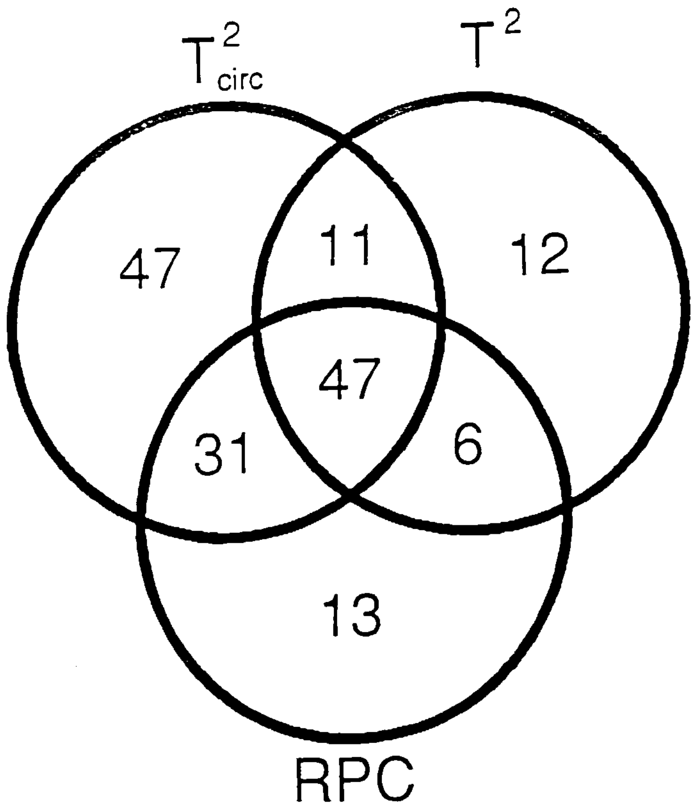


Fig. 5. A Venn diagram summarizing signal detection by T^2_{circ}, T^2, and the RPC. Numbers within each circle indicate the number of episodes in which the corresponding statistic detected a signal first. Numbers in overlaps correspond to ties.

If estimates from correlated subsegments are incorrectly assumed to be independent, then Fourier estimates whose deviation from zero is due to chance alone may be erroneously considered significant. It is therefore reasonable to correct empirically for this error by choosing a more stringent criterion for signal detection.

With this in mind, we compare the performance of the 3 statistics in detecting a signal in a single episode (29.96 sec) of experimental data. The episode is partitioned into 150 subsegments 0.20 sec long. A signal is considered to be detected by the Nth subsegment if the chosen statistic achieves significance for the first N subsegments, and does not lose significance when 1 and 2 additional subsegments are added. With this criterion, no false positives were encountered in analysis of 32 min of EEG recorded in response to an unmodulated pattern. Performance of T^2_{circ}, T^2, and the RPC are summarized in Fig. 5. In 25 of the 192 episodes, a signal was not detected by any of the 3 statistics. Of the remaining 167 episodes, the signal was detected initially (including ties) by T^2_{circ} in 136, by T^2 in 76, and by the RPC in 97. The superior performance of T^2_{circ} was found at all 3 temporal frequencies tested. In those episodes in which a signal was detected by all 3 statistics, the average length of data required to detect a signal was 3.74 sec for T^2_{circ}, 4.28 sec for T^2, and 4.22 sec for the RPC.

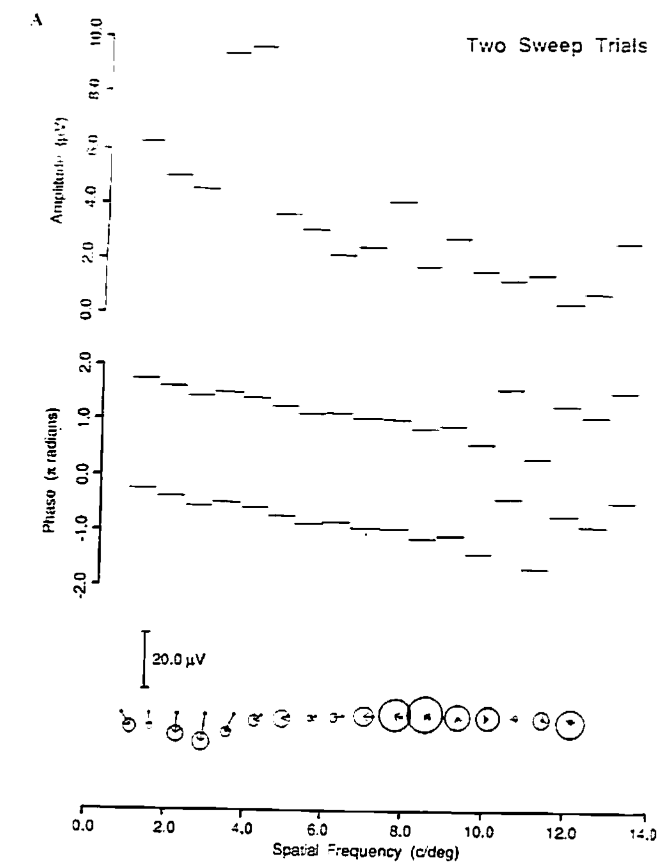
Comparison of confidence regions

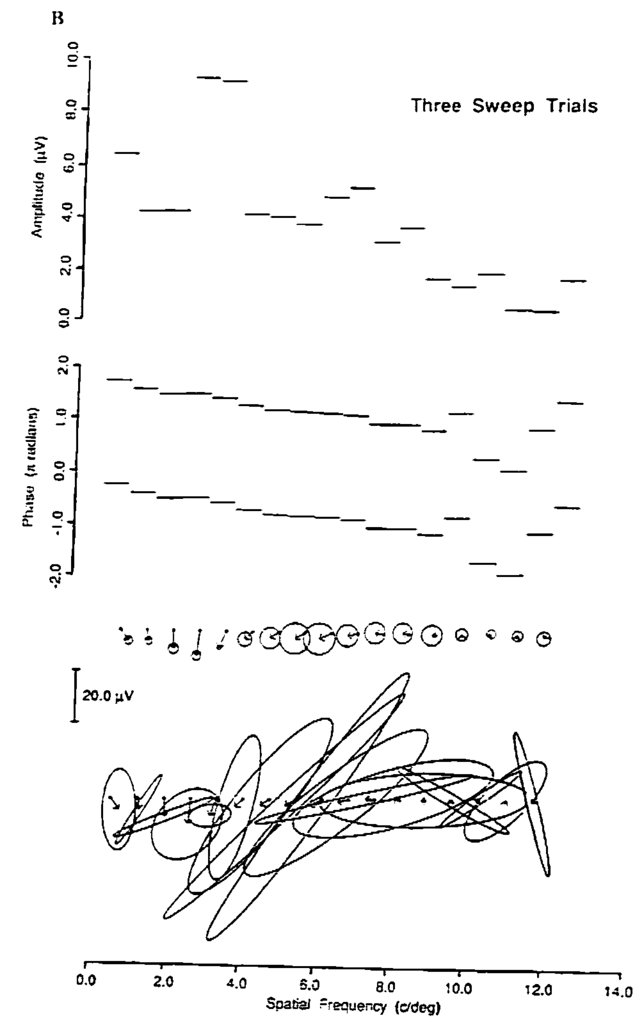
In a typical application, one is interested in knowing not merely whether a signal is present, but also what are the confidence limits on its amplitude and phase. With steady-state responses represented as vectors in the complex plane, confidence regions are represented by regions surrounding the vector. The T^2 statistic provides elliptical confidence regions; the axes of the ellipse are related to the estimated variances of the real and imaginary parts of the responses and their covariance (Anderson 1958). The T^2_{circ} statistic provides for circular confidence regions. The RPC does not provide for confidence regions, since it ignores response amplitude.

We now compare the area of confidence regions provided by T^2_{circ} and T^2. For this comparison, we considered each data collection period to be a single segment (30 sec), without subdivision. Thus, these confidence regions are derived in a manner which respects the need for truly independent estimates. A comparison was made for 8 subjects at 3 temporal frequencies. In 22 of the 24 ($= 3 \times 8$) comparisons, the T^2_{circ} statistic provided a smaller confidence region than that provided by T^2. On the average, the confidence region provided by T^2_{circ} was 22% smaller (geometric mean).

A comparison with practical importance was made from 'sweep VEP' (Regan 1973; Norcia and Tyler 1985) data collected from human infants. The data shown here (Fig. 6) are from a normal 14 week infant. In this implementation (Mast and Victor 1989) of the sweep technique, a grating is modulated at a reversal rate of 13.5 Hz as spatial frequency is incrementally stepped every 0.89 sec during the course of a 15 sec 'sweep' from 0.7 c/deg to 10.6 c/deg. For the present analysis, each 0.89 sec period of constant spatial frequency was regarded as a data segment, and corresponding segments were compared across 2, 3, or 4 successive sweep trials.

Strictly speaking, we cannot regard responses within a single data segment as a 'steady-state' VEP, since there is no reason to expect that response phase is constant in time. But we can use the T^2_{circ} statistic to assay consistency of Fourier coefficients measured from corresponding segments in different sweep trials, and to estimate confidence limits. Since these corresponding segments are well separated in time, the hypothesis of independent estimates is met. Responses and confidence regions as determined by T^2_{circ} and T^2 are shown in Fig. 6. For comparisons across two trials (Fig. 6A), only the T^2_{circ} statistic may be applied. Responses were found to be significant by the T^2_{circ} statistic for the first 9 sweep segments and bordered on significance in the tenth segment (approximately 5.3 c/deg). For compari-





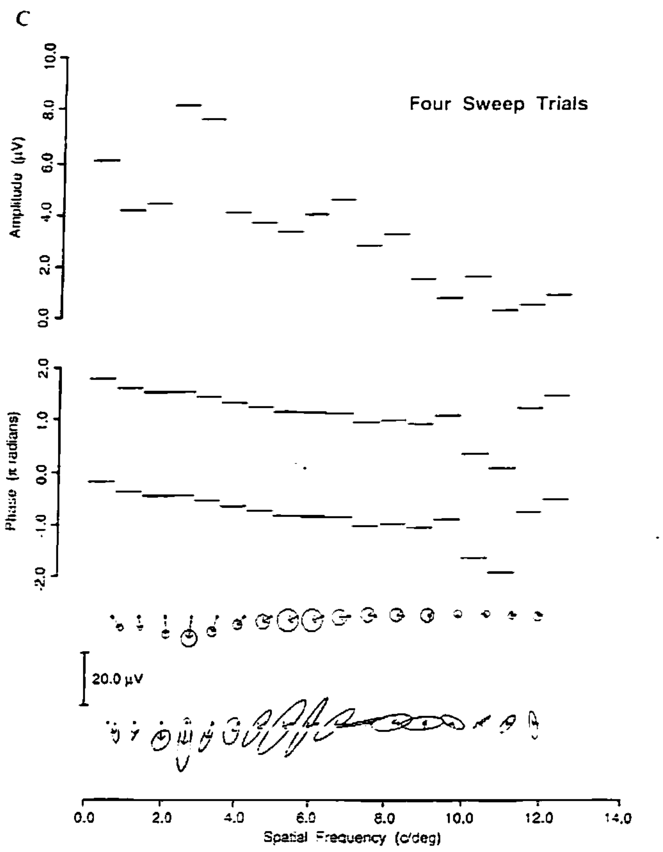


Fig. 6. Analysis of 'sweep-VEP' data by T^2_{circ} and T^2 statistics. A: 2 sweep trials. B: 3 sweep trials. C: 4 sweep trials. In the top 2 graphs of each part, the amplitude and phase (in π radians) of the second harmonic of the VEP are plotted as a function of spatial frequency. In the third graph of each part, this Fourier component is plotted as a vector, with the 95% confidence circle as determined by T^2_{circ}. The 95% confidence ellipse as determined by T^2 is shown as the lowest graph in B and C, but not in A, since confidence regions cannot be determined from T^2 with only 2 repeats.

sons across 3 trials (Fig. 6B), responses bordered on significance into the eleventh segment (approximately 6.0 c/deg) by the T^2_{circ} statistic. However, the T^2 statistic indicated significance only in 3 of the initial segments, and the confidence regions were many times larger than those determined by T^2_{circ}. For comparisons across all 4 trials (Fig. 6C), responses bordered on significance into the twelfth segment (approximately 6.7 c/deg) by the T^2_{circ} statistic. The T^2 statistic indicated significance as far as the eleventh segment. Confidence regions determined by T^2 were approximately twice as large as those determined by T^2_{circ}.

Discussion

Comparison of methods of signal detection

While the need to have rigorous yet efficient methods to assess the reliability of Fourier components estimated from experimental data is clear, current practice is to adapt standard statistics from other contexts. For example, Norcia and Tyler (1985) and Stapells et al. (1987) determined the presence of signal in individual epochs on the basis of measures of phase coherence (Mardia 1972). Picton et al. (1987) made a rigorous comparison of phase-coherence methods with the T^2 statistic (Anderson 1958), in the extraction of steady-state auditory evoked potentials, and found no significant advantage for either method: both methods were somewhat superior to the ($\pm$) average initially introduced for transient signals (Schimmel 1967).

In Figs. 2 and 3, we used simulated data to compare the two methods found most efficient by Picton et al. and the T^2_{circ} statistic described here. The results of Figs. 2 and 3 may be summarized as follows. There are 2 regimes: a regime in which the number N of analysis segments is limiting, and a regime in which the SNR of individual analysis segments is limiting. In the first regime, T^2_{circ} and the RPC are superior, and T^2 is relatively inefficient. In the second regime, T^2_{circ} and T^2 are superior, and the RPC is relatively inefficient. For intermediate values of the SNR and the number of analysis segments, T^2_{circ} performs better than either statistic.

The first regime (a limitingly small number N of analysis segments) can be understood by counting up the number of parameters available from the data, and the number of parameters which must be estimated by the statistical technique. Each analysis segment provides 2N pieces of data: a real and an imaginary component of the estimated Fourier component. To apply the T^2_{circ} statistic, a total of 3 quantities must be estimated: the real and imaginary parts of the actual response, and the radius of a confidence circle. Thus, with N Fourier estimates, there are $2N - 3$ effective degrees of freedom. However, to apply the T^2 statistic, a total of 5 quantities must be estimated: the real and imaginary parts of the actual response, their variances (which are not assumed to be equal), and their covariance. Thus, with N Fourier estimates, there are $2N - 5$ effective degrees of freedom. Finally, the RPC essentially attempts to estimate one parameter (phase coherence), but only makes use of one parameter from each segment (phase), and hence has $N - 1$ effective degrees of freedom. The T^2 statistic cannot even be applied unless there are at least 3 segments; the RPC and T^2_{circ} statistic can be applied if there are only 2 segments. For a small number of segments, the number of degrees of freedom determines the relative efficiency of the statistical methods.

The second regime is that of low SNR. When the SNR is low, many segments must be averaged in order to reach significance. Here, it becomes crucial to use all of the data available from each segment: the effective number of degrees of freedom is no longer limiting. The statistics T^2_{circ} and T^2 use both real and imaginary parts of the Fourier estimate. The RPC, however, depends only on phases of individually estimated components. In ignoring amplitude information, it throws out a portion of the already degraded signal and hence cannot perform as well as the other statistics.

In both regimes, T^2_{circ} performs as well as (or better than) T^2 and the RPC. This enhancement is attained without an increase in the rate of false positives. Fundamentally, the enhancement relies on amplitude information not used by the RPC statistic, and a priori information not used by the T^2 statistic. The critical piece of a priori information is that fluctuations of real and imaginary parts of Fourier estimates are of equal variance and zero covariance.

There are 2 ways of justifying this assumption. First, it is a consequence of the hypothesis that the EP signal and the EEG noise are independent. However, while this is a tacit assumption of the T^2 statistic as well, it is not typically tested explicitly (see Mast and Victor 1991 for a review). When we searched for interactions of the VEP and the EEG, we found that additivity and independence were only approximations (Mast and Victor 1991). However, despite additivity failure, deviations from the assumptions of independent real and imaginary parts of equal variance could not be detected in approximately 90 min of combined VEP/EEG data. Thus, although the assumption of additivity is not entirely valid, the central assumptions underlying the T^2_{circ} statistic were verified.

The theoretical advantages of the T^2_{circ} statistic were borne out in practice as well as in numerical simulations. T^2_{circ} detected signals earlier than T^2 and the RPC, and confidence regions derived from T^2_{circ} were consistently smaller than those derived from T^2. The improvement offered by T^2_{circ} depended on the amount of data and its signal-to-noise in a manner consistent with the theoretical analysis: the largest difference was obtained under circumstances in which signal-to-noise was high and only a few segments were required to achieve statistical significance (Fig. 6). Another practical advantage of T^2_{circ} over T^2 is that calculation of confidence regions is simpler: they are circular and are derived from the variances of the real and imaginary parts of the Fourier estimates. In contrast, confidence regions derived from T^2 are elliptical and require calculation of covariances and determinants as well (Anderson 1958).

The potential advantages of the T^2_{circ} statistic apply to sinusoidal analysis of steady-state responses in the general physiological context, and not only to VEPs. Indeed, this statistic has been previously used in the analysis of single-unit responses as a criterion for whether responses were significantly different from zero (Victor and Shapley 1979).

Not the last word

We pointed out that although T^2_{circ} represents an improvement over T^2 for analysis of steady-state responses in a background of additive noise, it does not represent the last word. One situation in which further gains may be anticipated is that in which the variance of a Fourier estimate is known from a priori data. Furthermore, our analysis only treats statistics of a single harmonic. In situations in which more than one harmonic is present, extensions of T^2_{circ} may need to be considered.

Conclusion

The standard T^2 statistic (Hotelling 1931; Anderson 1958) is the optimal statistic for determining the confidence limits for a vector quantity corrupted by additive gaussian noise of *unknown* variance/covariance. We have exploited the additional information that real and imaginary parts of an estimate of the Fourier component are independent and of equal variance to improve on the T^2 statistic.

This work was supported in part by Grants EY1428, EY6871, EY7977, and HD819.

We thank Mary Conte and Annemarie Canel for their assistance.

References

Anderson, T.W. The generalized T^2-statistic. In: An Introduction to Multivariate Statistical Analysis. Wiley, New York, 1958.

Hotelling, H. The generalization of Student's ratio. Ann. Math. Statist., 1931, 2: 360–378.

Hughes, G.W. and Maffei, L. Retinal ganglion cell response to sinusoidal light stimulation. J. Neurophysiol., 1966, 29: 333–352.

Mardia, K.V. Statistics of Directional Data. Academic Press, New York, 1972.

Mast, J. and Victor, J.D. Sweep-VEP response phase: normative data and developmental trends. Invest. Ophthalmol. Vis. Sci., 1989, 30 (Suppl.): 311.

Mast, J. and Victor, J.D. Fluctuations of steady-state VEPs: interaction of driven evoked potentials and the EEG. Electroenceph. clin. Neurophysiol., 1991, 78: 389–401.

Milkman, N., Schick, G., Rossetto, M., Ratliff, F., Shapley, R. and Victor, J.D. A two-dimensional computer-controlled visual stimulator. Behav. Res. Meth. Instrum., 1980, 12: 283–292.

Norcia, A. and Tyler, C.W. Spatial frequency sweep VEP: visual acuity during the first year of life. Vision Res., 1985, 25: 1399–1408.

Picton, T.W., Vajsar, J., Rodriguez, R. and Campbell, K.B. Reliability estimates for steady-state evoked potentials. Electroenceph. clin. Neurophysiol., 1987, 68: 119–131.

Pringle, J.W.S. and Wilson, V.J. The response of a sense organ to a harmonic stimulus. J. Exp. Biol., 1952, 29: 220–234.

Regan, D. Some characteristics of average steady-state and transient responses evoked by modulated light. Electroenceph. clin. Neurophysiol., 1966, 20: 238–248.

Regan, D. Rapid objective refraction using evoked brain potentials. Invest. Ophthalmol. Vis. Sci., 1973, 12: 669–679.

Regan, D. Human Brain Electrophysiology. Elsevier, New York, 1989.

Schimmel, H. The (±) reference: accuracy of estimated mean components in average response studies. Science, 1967, 157: 92–94.

Sokal, R.R. and Rohlf, F.J. Biometry: The Principles and Practice of Statistics in Biological Research. Freeman, San Francisco, CA, 1969.

Stapells, D.R., Makeig, S. and Galambos, R. Auditory steady-state responses: threshold prediction using phase coherence. Electroenceph. clin. Neurophysiol., 1987, 67: 260–270.

Strasburger, H. The analysis of steady state evoked potentials revisited. Clin. Vision Sci., 1987, 1: 245–256.

Van der Tweel, L.H. and Verduyn Lunel, H.F.E. Human visual responses to sinusoidally modulated light. Electroenceph. clin. Neurophysiol., 1965, 18: 587–598.

Victor, J.D. and Shapley, R. Receptive field mechanisms of cat X and Y retinal ganglion cells. J. Gen. Physiol., 1979, 74: 275–298.