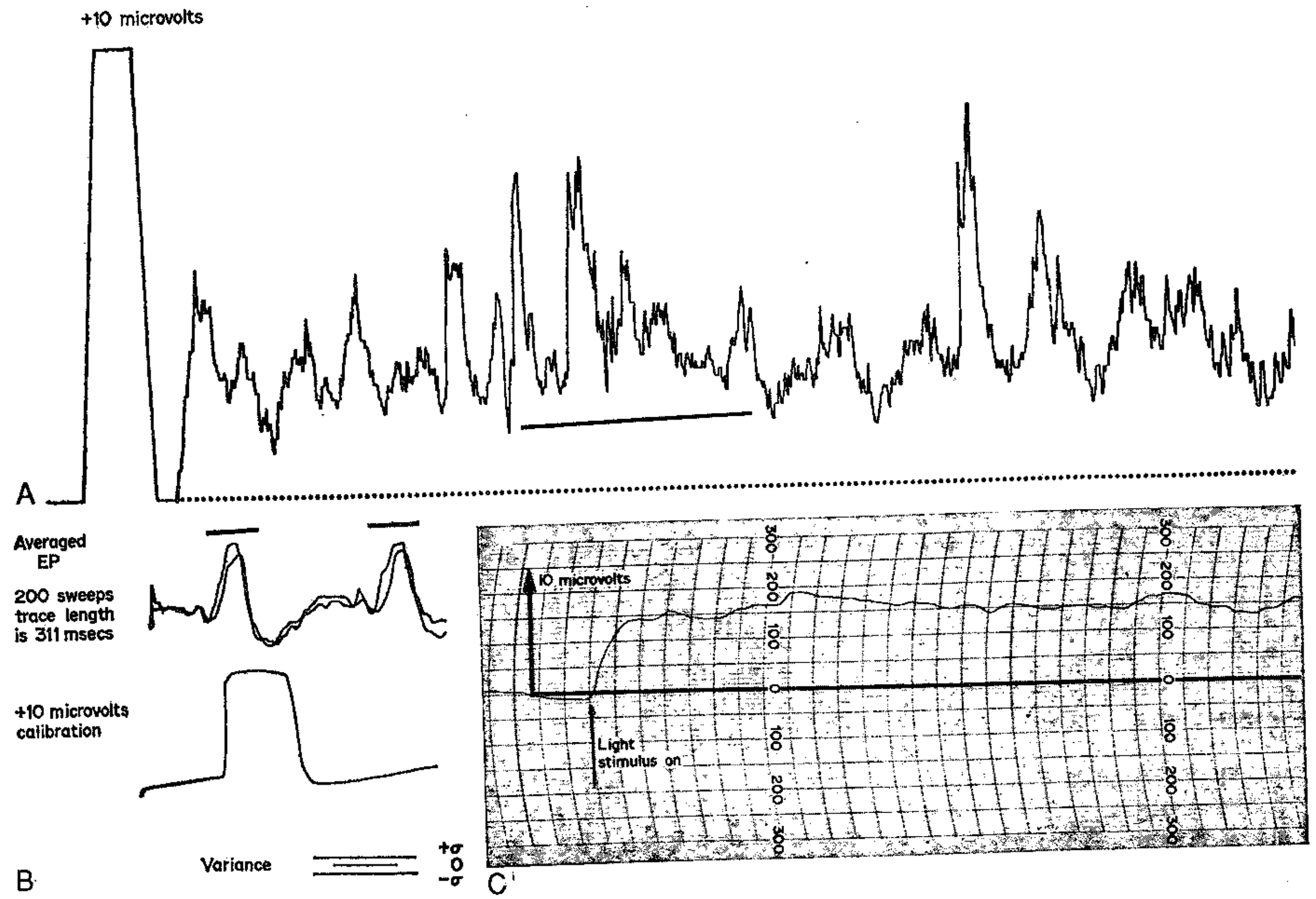


Figure 1.33
Two Ways of Detecting Progressive Change in the EP
The subject fixated a checkerboard pattern for a period of several minutes. There were six reversals per second. A computer integrated the EEG for a 50-msec period centered on a latency of 90 msec, and a running average was formed of each 16 successive responses. Panel A shows this running average throughout 320 sec of continuous stimulation. The two averaged VEPs in panel B were also recorded during this 320-sec period; each trace comprised 200 sweeps, that is, 67 sec of stimulation. The black bars in panel B show the part of the averaged VEP whose constancy with time is shown in panel A. Panel C shows the amplitude of the 6-Hz component of the EEG throughout a recording of 430-sec duration. The stimulus conditions and subject were the same as in panel A. In panel C the 6-Hz frequency component was recorded by the Fourier series analyzer similar to that described in Reference 1973. Moment-to-moment variability seems to be less in C than in A, presumably because much of the apparent variability in A was due to additive noise rather than being true signal variability; noise rejection was better for the Fourier analyzer than for the averager because of its narrower bandwidth. This conclusion is supported by the very narrow bandwidths of the components in Figure 1.29B, 1.32, 1.70, 2.149, and 2.189. (Dr. J. Axford provided records A and B. From Regan D: *Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine*. London, Chapman & Hall, 1972. Reproduced by permission.)

to sensory adaptation that occurred, at least in part, at the retinal level. A similar "running-average" technique was recently used to demonstrate adaptation to contrast using grating stimuli of more moderate luminance (Fig 2.184). Note that a conventionally averaged trace would not have revealed these systematic EP changes that occur during the course of a single 30-sec or 1-min recording session.

Several methods for dealing with EP variability and nonstationarity are discussed below. One approach is to devise rapid techniques so that measurements are completed within a shorter time (Section 1.9). Another approach is to interleave the various stimuli in random order so that EP variability and nonstationarity do not systematically bias the data. A third approach is to use short recording sessions while recording the EEG on an FM tape recorder for subsequent off-line computation of grand averages.[2023]

1.5 Methods for Recording Transient Evoked Potentials. I: Superimposition and Signal Averaging in the Time Domain

1.5.1 Superimposition

The principle of the superimposition technique was used more than a century ago by Galton. One of his aims was to identify common features in the faces of murderers and violent criminals, so that potential offenders could be recognized before the crime was committed. He superimposed photographs of the faces of convicted murderers on the assumption that common features (e.g., close-set eyes) would be preserved in the superimposition, while random variations would reduce to a blur (Fig 1.34). No common physical features emerged. Galton's failure to identify a common physical element in murderers' faces would not occasion surprise today,[1207] though in superimposing photographs of the heads of the most successful racehorses of the 1870s, in a scientific attempt to identify future winners, he might seem well ahead of his time (Fig 1.35).

It was the practical problem of recording reliable somatosensory EPs in myoclonic epilepsy that led Dawson[521,522] to use the superimposition technique of signal-to-noise enhancement in EP recording. The procedure is to trigger the sweep of a cathode ray oscilloscope (CRO) synchronously with the stimulus, so that samples of the EEG that occur in the 500 msec or so following successive stimuli are displayed on the face of the CRO tube. The waveforms that are successively displayed on the face of the CRO are then superimposed on a single photographic frame by leaving the CRO camera shutter open throughout the stimulation procedure. The superimposition technique can offer some useful enhancement of the signal-to-noise ratio by virtue of the integrating properties of photographic emulsions. Features of the response that occur at a constant time after the stimulus grow to be more clearly defined after a number of sweeps, while those features that are not regularly related to the stimulus produce no more than a thickening of the baseline (Figs 1.36 and 1.37).

The superimposition technique does not lend itself to precise quantitative measurement of the EP. On the other hand, it does offer direct visual indication of the variability of "raw" individual traces. In view of the considerable dangers of grossly misinterpreting data, and the exposure to a rich variety of artifacts that confront the investigator who relies *exclusively* on the computer's powerful data-reduction procedures, the superimposition technique might still seem to have a useful place in EP research as a routine monitoring procedure.

1.5.2 Signal Averaging in the Time Domain

Although the signal-averaging technique is related to superimposition, there is a major difference: cancelation never takes place in superimposition, whereas cancelation is inherent to signal averaging. To the extent that the final photograph in Figure 1.36 accurately represents the sum of the light intensities passing through each of the individual photographs, then the superimposition procedure can be regarded as a point-by-point summation, but cancelation never occurs because light *intensity* is never negative.

Dawson used a pencil-and-paper method to demonstrate that signal averaging is effective in extracting EPs from the EEG. He recorded a continuous EEG trace from a subject whose ulnar nerve received an electric shock once per second. Using a ruler, he measured the instantaneous EEG voltage at the moment of the first shock, at 3 msec after the first shock, at 6 msec after the first shock, and so on until 40 ordinates had been measured. Then he compiled a similar table of amplitudes after the second shock, a third table after the third shock, and so on up to the twentieth shock. Finally, he added and averaged these 20 tables of EEG amplitudes. In this way he computed a clear averaged EP waveform, although single-trial responses were strongly contaminated by noise in the individual EEG traces.

Dawson pointed out that this pencil-and-paper averaging procedure is essentially the procedure suggested by Laplace in the late eighteenth century for detecting the tiny atmospheric pressure changes that correspond to the sea tides (i.e., the atmospheric tides).[1359] These small changes in pressure are buried in much larger irregular fluctuations and also in periodic changes of solar origin. In 1847, Laplace's idea was put into practice by Sabine who retabulated his barometric pressure records with respect to the lunar cycle and thus succeeded in detecting the lunar tide in the atmosphere.[2162]

Rigorous statistical treatments of signal-averaging theory are available. Appendix 1.1 provides an outline. In brief, averaging N samples of a waveform improves the signal-to-noise ratio by a factor $\sqrt{N}$, providing that the following requirements are satisfied:

Figure 1.34
Galton's Photographic Superimposition Technique for Identifying Common Physical Features Within a Group of Individuals: A Blind Alley
Superimposed faces of violent criminals and murderers of the 1880s. "Composites" are the superimposed faces. (From Pearson K: *Sir Francis Galton 1822–1911*. London, Cambridge University Press, 1914–1930, Reproduced by permission.)

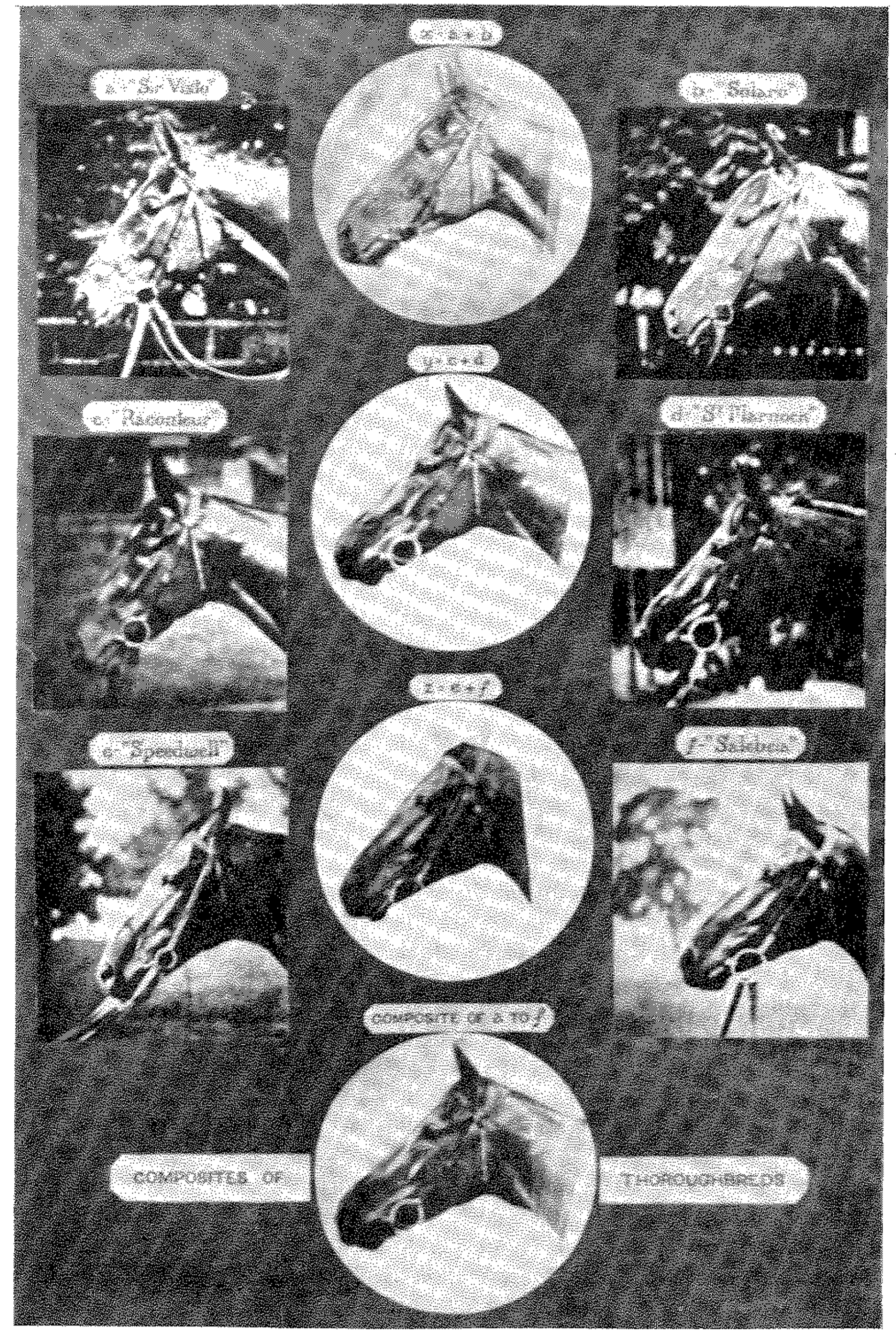


Figure 1.35
Galton's Photographic Superimposition Technique for Identifying Common Physical Features Within a Group of Individuals: The Scientific Sportsman
Superimposed heads of late nineteenth-century winning thoroughbred horses. (From Pearson K: *Sir Francis Galton 1822–1911*. London, Cambridge University Press, 1914–1930, Reproduced by permission.)

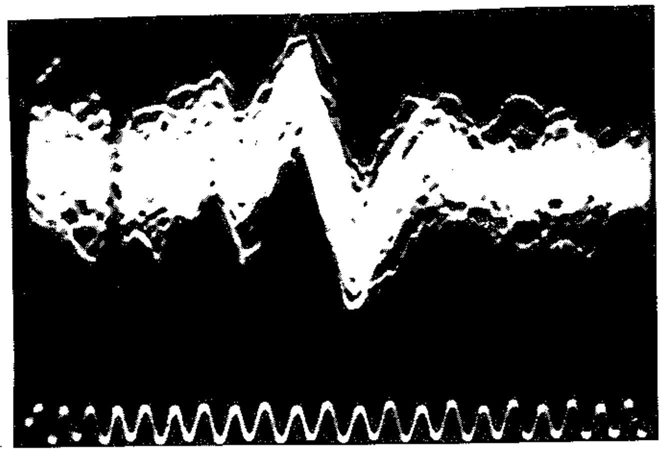


Figure 1.36
Dawson's Superimposition Method for Improving Signal-to-Noise Ratio
Superimposed EEG samples following light flashes. The flashes are marked by the vertical dark band at the left of the white traces. (From Ciganek L: Die Elektroencephalographische Lichtreizantwort der Menschlichen Hirnrinde. Bratislava, Verlag der Slowakischen Acad. der Wissenschaften, 1961. Reproduced by permission.)

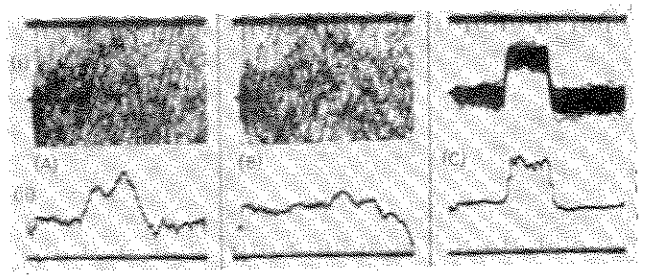


Figure 1.37
The First Published Averaged EPs
This figure compares the effectiveness of Dawson's automatic averaging machine with his superimposition method. A subject's left ulnar nerve was stimulated at the wrist once per second. The upper record (I) shows 55 individual EEG traces superimposed, and the lower record (II) shows the averaged trace. Superimposition provides only a hint of the EP waveform in panel A, but the averaging machine extracted a clear response (II). Panels A and B demonstrate lateralization of the somatosensory EP. Panel A was recorded from contralateral scalp with one electrode on the midline and one over the right central sulcus. Panel B was recorded from ipsilateral scalp between the same midline electrode and an electrode over the left central sulcus. Panel C shows a 5-μV calibration pulse. Large spikes on the time scale are 20 msec apart. (From Dawson GD: A summation technique for detecting small signals in a large irregular background. *J Physiol* 1951;115:1–2P. Reproduced by permission.)

1. the waveform to be averaged is the sum of two independent waveforms, namely, the signal waveform and the noise waveform;
2. the signal waveform is produced by a process that is stationary from trial to trial, and its variance is negligible;
3. the noise waveform is produced by a stationary random process;
4. the N samples of noise are uncorrelated from trial to trial.

Although the pencil-and-paper method is of educational value, it is far from convenient. The field of evoked potentials as we know it today effectively dates from the introduction of the automatic averager. Dawson was painstaking in his efforts to document earlier work that led to the automatic averaging machine.[10] The fact remains that, by building a machine that averaged automatically, Dawson created from the statistical theory of averaging a powerful practical tool, and went on to demonstrate its power by recording microvolt-level EPs that were completely masked by the background EEG. Figure 1.37 shows these first averaged EPs, demonstrated to the Physiological Society in May 1951.

Dawson's automatic signal averager can be seen in the Science Museum in London. The machine operated by successively charging an array of capacitors that acted as integrators; switching was performed by an arm that rotated 10 times per second (Fig 1.38). This machine is the precursor of the thousands of automatic signal averagers in use today. More than 30 years later, Dawson's succinct outlines of the principles of signal-to-noise enhancement are still excellent introductory reading.[524a,525]

Automatic averaging performed by hard-wired circuits or by general-purpose computers has, of course, far wider application than EP recording. It is now a routine tool in cellular physiology, nuclear physics, engineering, and many branches of medicine. Because of its ability to extract low-level signals, the averaging machine can record the transient responses of mechanical, electrical, and other systems to inputs that are so weak that they barely disturb the system.

Several other laboratories developed averaging machines during the 1950s using a variety of devices[116] including barrier grid tubes,[449] magnetic tape,[1475,2122] photographic emulsions,[343a,b,940,2284a,2285] and cathode ray

[10] In his 1951 paper,[524] Dawson acknowledged the advice of Dr. C. J. Hunt, who suggested that "an additive technique would be of value and that a condenser storage system might be used." In his 1954 review[525] he described the earlier instrumental contributions of Martindale (1941) and Foster (1946) and traced the basic idea back to the eighteenth century. The introduction to his 1954 paper says "A certain Dr. Brown, on being rebuked because he had failed to acknowledge some previous work on the subject of his writings replied: 'I make no claim to originality for I have long since found that to consider oneself original one must read nothing at all. All I have done is to describe those methods which I have found to suit me best in practice.'"

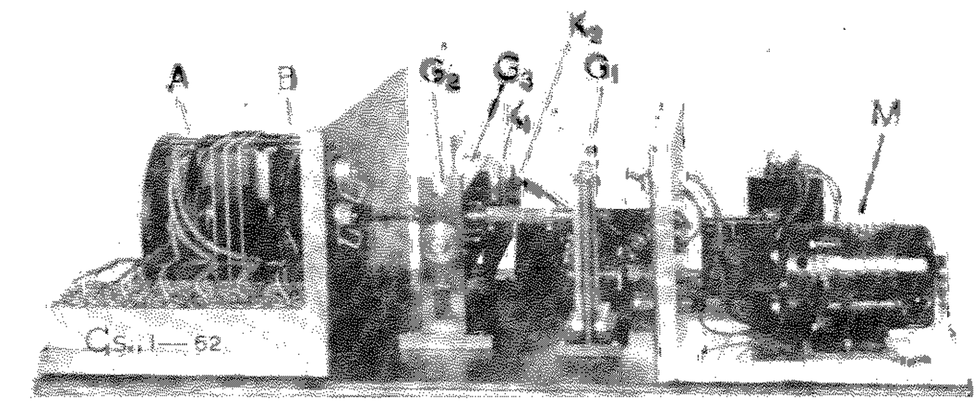


Figure 1.38
Dawson's Automatic Averaging Machine
General view of the storage unit. A and B are the two distributors. C_s1–62, storage capacitors; K_1 and K_2, timing contacts initiating the stimuli, the changing of the store, and the starting of display sweeps; M, driving motor. (From Dawson GD: A summation technique for the detection of small evoked potentials. *Electroenceph Clin Neurophysiol* 1954;6:65–84. Reproduced by permission.)

tubes,[315a,1304a] but it was the digital memory of the average response computer (ARC)[406] that anticipated the course of future technical developments and led to the first commercial averager, an excellent machine designed by Clynes and Kohn.[414a] The commercial availability of reliable averagers made possible the enormous expansion of EP research that took place during the 1960s.

The modern digital averager does not process the continuously varying EEG voltage directly. The EEG waveform is first sampled or digitized by an analog-to-digital converter (ADC). As described in Section 1.2.10 the digitization rate must be more than twice (in practice 2.5–4 times) the highest frequency in the EEG if "aliasing" is to be avoided. Section 1.2.10 also notes that, in practice, if the effective resolution of the ADC is 8 bits or more, the quantal noise associated with digitization is of negligible importance if sufficient noise is present and a sufficient number of responses are summed.

Now we turn from the resolution of the ADC converter to the question of memory resolution. This question has been discussed by Picton et al[1868] and others.[369] The resolution required depends on the algorithm used in the averaging calculations. The simplest algorithm involves reading the inputs from the ADC as positive integers and adding the values from each trial to the sum of the values from previous trials. When averaging N trials we assume that the amplitude of the summed noise will increase by $\sqrt{N}$, and the amplitude of the summed signal will increase by N (Appendix 1.1). Because the amplitude of the summed signal eventually becomes greater than the amplitude of the summed noise, the determining factor for memory resolution will be the amplitude of the summed signal. The required memory resolution is $\log_2(NS/A)$ bits more than the resolution of the ADC, where N is the number of trials averaged, S is the peak-to-peak amplitude of the EP signal, and A is the voltage range of the ADC.[1868] Optimally, the EEG voltage should match the voltage range of the ADC, and S/A would then be the same as the signal-to-noise level in the EEG before averaging. For example, if 4,096 trials are averaged to extract a 1-μV signal from 16 μV of noise, then the memory resolution required is 8 bits more than the resolution of the ADC, provided that the 16-μV signal is amplified to match the range of the ADC.[1868]

A problem with this simple summing algorithm is an artifactual progressive rise of mean DC level and eventual "rollover" in the memory. Figure 1.39 illustrates this effect, so familiar to users of the original Mnemotron averager ("Computer of Averaged Transients," or CAT). Rollover occurs as the summed signal, continuously rising up the cathode ray tube (CRT) monitor's face, goes over the maximum value that can be stored in memory and reappears at the bottom of the screen as a small number. In practice, these rollovers in the CAT were not a serious problem. The final averaged trace could be reassembled by electronically adding a constant to the trace that completed or reversed the rollover. Commercial averagers that followed the CAT did not have this continually rising display; rollover was prevented by automatically subtracting the signal mean from the output of the ADC. In some cases, however, the rollover cannot be eliminated so easily because the averaged signal is larger than the range of the memory. The familiar result

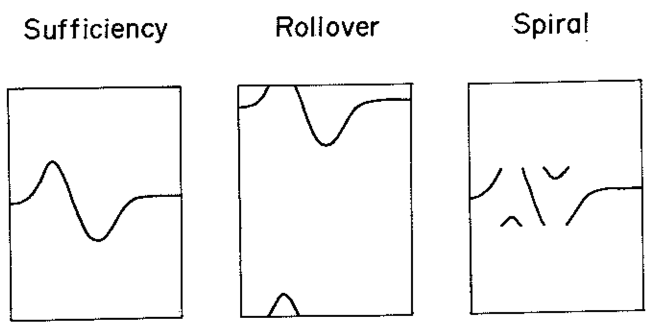


Figure 1.39
Memory Limitations During Averaging
On the left is shown the average EP recorded when the memory resolution is sufficient for the calculations involved in averaging. In the middle is shown the "rollover" of memory that occurs when unsigned integer (binary offset) arithmetic is used and the number of trials involved in calculating the average exceeds 2^{b+1}, where b is the number of extra bits in the memory compared with the ADC. On the right is shown memory "spiral." This occurs using either signed or unsigned integer arithmetic when the number of trials involved in the average exceeds 2^{b+r}, where b is the number of bits in the memory greater than the ADC and r is the base 2 logarithm of the ratio of the range of the ADC converter to the peak-to-peak amplitude of the EP. The waveforms illustrated in this figure are diagrammatic. (From Picton TW, Hink RF, Perez-Abalo M, Linden RD, Wiens AS: Evoked potentials: How now? *J Electrophysiol Technol* 1984;10:177–221. Reproduced by permission.)

is memory "spiral," also illustrated in Figure 1.39. Three ways of handling this problem are[1868] (1) to reduce the gain of the input amplifier, (2) to increase the resolution of the memory, and (3) to reconstruct the average using an algorithm that detects when the average waveform jumps by a factor greater than half the memory limit divided by the number of trials, and then shifts the following waveform into its appropriate position.[2458]

Time-domain averaging is illustrated in Figures 1.40 and 1.41. The subject fixated the upper boundary of a pattern of black and white checks of side length 50 min arc and of near-100% contrast that subtended 15° (vertical) × 25°. The checks exchanged places 1.88 times per second. Mean luminance was 3 cd/m². Electrodes were placed on the inion and vertex with the right mastoid grounded. The EEG was amplified by a Nicolet Model HGA-200A amplifier and, together with trigger pulses, recorded on an FM tape recorder (DC–600 Hz bandpass). The amplifier bandpass was DC–100 Hz. Ten separate 1-min recordings were made. Figures 1.40A–D show four single-trial EEG samples. These four samples are superimposed in Figure 1.40E. No VEP waveform is evident in this superimposition; all that emerges is some indication of rhythmic activity at about 9–12 Hz. Figure 1.41A shows the result of averaging four single-trial samples. Four such averages are super-

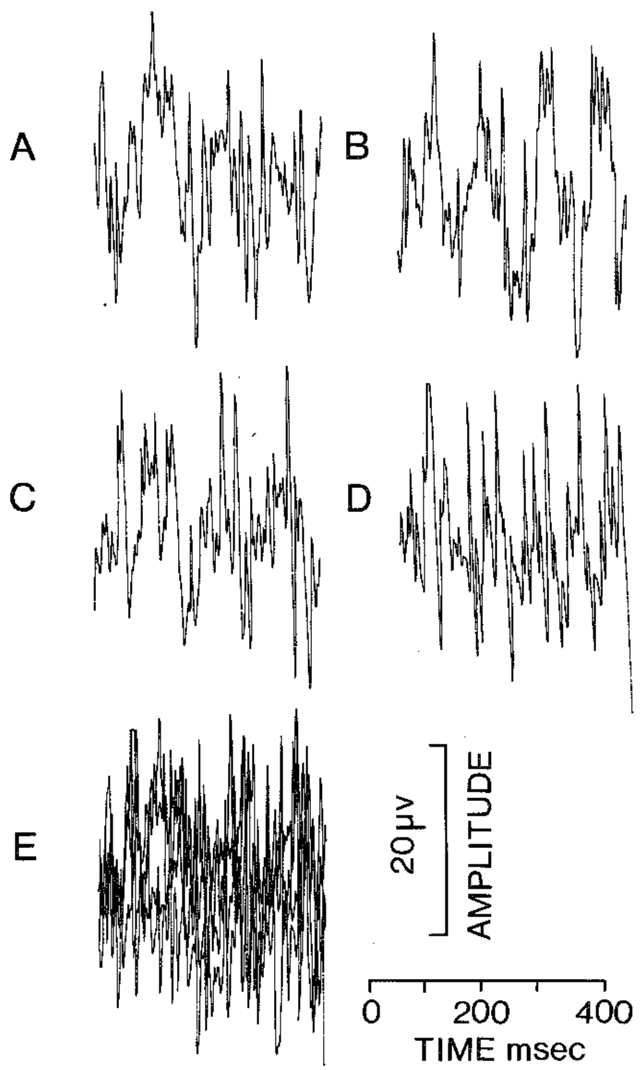
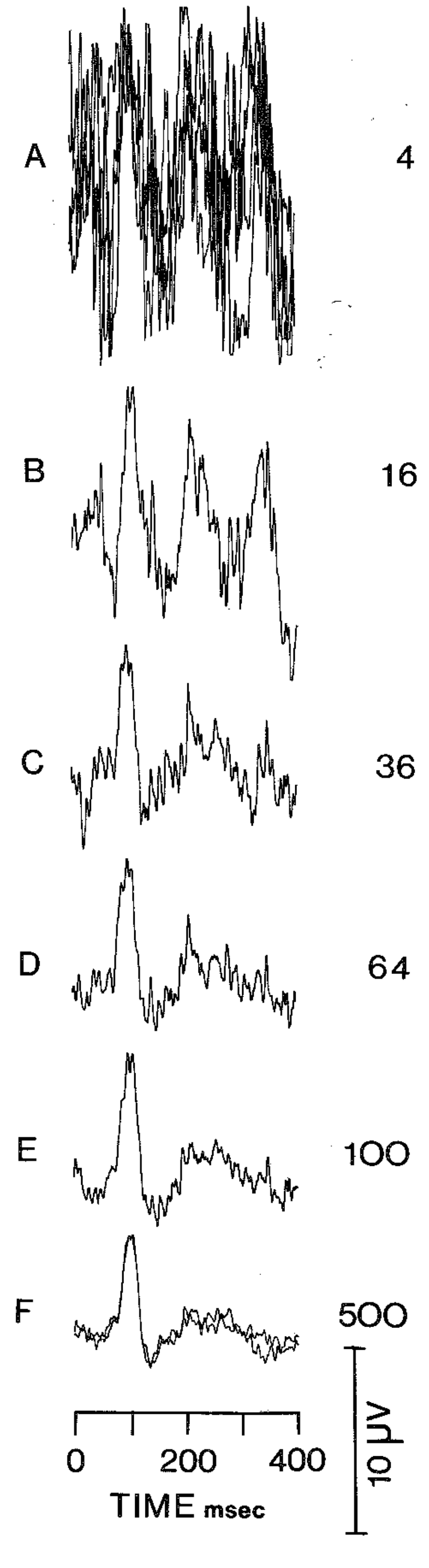


Figure 1.40
Masking of Visual EP by Noise
(A–D) Four successive samples of EEG recorded while a subject viewed a black and white pattern of checks. Each record is 400 msec long, and started the instant that the black and white checks exchanged position. The four traces are superimposed in panel E. Panel E shows that the EP is not evident in the single-trial responses; all that emerges is some rhythmic activity at about 10 Hz. Stimulation and recording conditions are described in the text.

imposed. Again, rhythmic activity at about 10 Hz can be seen, but it is not clear whether this activity is correlated with the stimulus. The same conclusion held after 16 averages (Fig 1.41B), but after 36 averages (Fig 1.41C) the EP became (rather suddenly) clearer. A closely reproducible VEP was obtained after 500 sweeps. (Figure 1.41F shows two superimposed averages of 500 sweeps.) By comparing Figures 1.41F and D, it can be seen that in this case a tolerably adequate estimate of VEP waveform



was available after only 64 sweeps even though, as shown in Figure 1.40, the VEP waveform was not evident in single-trial samples.

1.5.3 Signal Averaging in the Time Domain: Caveats

There is a temptation to visualize the waveform recorded by a conventional averaging computer as some sort of enlarged representation of an EP to a single stimulus. It is easy to forget that an averaging computer records the result of performing a certain statistical procedure on "raw" EEG. The simple treatment of signal averaging given in Appendix 1.1 applies to an ideal situation where the several conditions listed on page 50 are satisfied. Practical electrophysiological signals often violate one or more of these conditions.

Latency Jitter
The effect of trial-to-trial jitter of signal latency is to reduce the amplitude and *increase the latency* of the averaged EP. Figures 1.42 and 1.43 demonstrate these effects by means of a simulation. Figure 1.42 shows averages of 64 identical waveforms, each of 640-msec duration, but subject to random latency jitter according to the variable shown in Figure 1.43. Figure 1.42 clearly brings out the progressive increase of latency as mean jitter is progressively increased, an effect that should be borne in mind when interpreting VEP abnormalities in patients with multiple sclerosis and other diseases (Sections 3.2.1, 3.3.4, and 3.4.9).

Statistical Independence of Signal and Noise
The assumption of statistical independence between signal and noise is important. Failure to meet this assumption can have a large effect (usually adverse) on the signal-to-noise enhancement produced by time-domain averaging. In this context several authors have questioned the spontaneity of the so-called "spontaneous" EEG.

It is common practice to make a clear conceptual distinction between the EP signal and the "background EEG." From this point of view, changes in the "spontaneous EEG" that are triggered by the sensory stimulus are distinct from the EP. Although this view may not be stated clearly, it is implicit in the procedures of an experimenter who ignores stimulus-induced changes in the

Figure 1.41
Signal-to-Noise Enhancement by Signal Averaging
(A) Each trace is the average of four single-trial samples similar to the four single-trial samples shown in Figures 1.40A–D. (B–E) Average of successively more samples (16, 36, 64, 100). (F) Two separate averages of 500 samples superimposed to show the excellent reproducibility. Recording conditions were the same as in Figure 1.40.

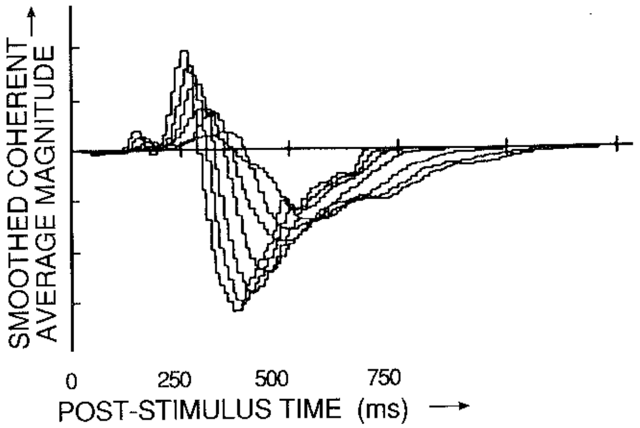


Figure 1.42
Latency Jitter Attenuates the Averaged EP and Increases Peak Latency
The traces are averages of 64 identical waveforms, each of 640-msec duration, subjected to progressively increasing amounts of latency jitter. (From Sayers BMcA, Beagley HA, Ross AJ: Auditory evoked potentials of cortical origin, in Beagley HA (ed): *Auditory Investigation: The Scientific and Technological Basis*. Oxford, Clarendon, pp. 489–506. Reproduced by permission.)

"background" EEG activity. Unfortunately, the averager does not share this viewpoint. Such EEG changes do not necessarily "average out."

Several investigators have examined the relationship between the VEP and alpha activity. Regan[1973] claimed that the steady-state flicker EP is virtually unaffected by

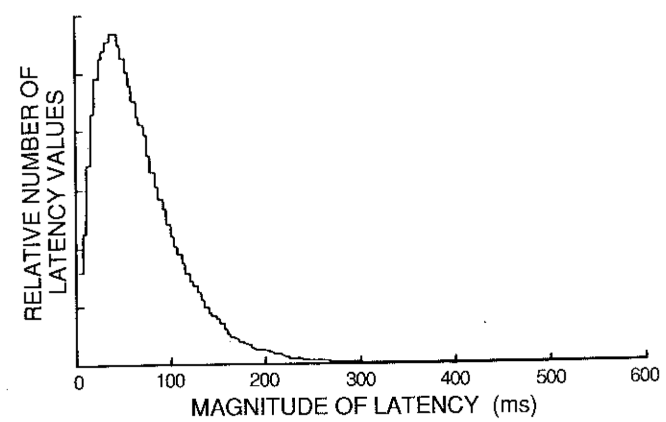


Figure 1.43
The Gamma-Distributed Random Latency Variable Used to Generate the Traces in Figure 1.42
Before averaging, the signals were added to an "EEG background" obtained by filtering white noise according to the average EEG amplitude spectrum. (From Sayers BMcA, Beagley HA, Ross AJ: Auditory evoked potentials of cortical origin, in Beagley HA (ed): *Auditory Investigation: The Scientific and Technological Basis*. Oxford, Clarendon, pp. 489–506. Reproduced by permission.)

alpha bursts, and a recent study by Jones and Armington[1175] confirmed that the VEP is independent of alpha activity and that the familiar "alpha driving" by flicker stimulation may be an example of a 10-Hz VEP adding to a free-running rhythm that occasionally has the same phase as the VEP. On the other hand, some investigators have concluded that visual stimulation can produce evoked changes in alpha activity. For example, it has been claimed that alpha phase is related to the timing of transient visual stimuli.[660,1299,1300,2535] Again, Kaufman and Locker reported that 1.5-Hz flicker stimulation modulated alpha amplitude so as to create power at sum and difference frequencies, that is, bands of power centered on about $10 + 1.5$ Hz and $10 - 1.5$ Hz.[1215a]

Jones and Armington[1175] found that a useful improvement in averaging efficiency could be obtained by eliminating periods of high-amplitude alpha activity, and this procedure did not distort the waveform of the averaged EP. Their paper includes a diagram of an alpha-rejection circuit.

Rejection of Evoked Potential Components by Time-Domain Averaging: Throwing the Baby Out With the Bathwater

Even if all single-trial EPs are identical, and even if background noise is negligible, it is still possible that the averaged VEP waveform is quite different from any single-trial VEP. This is because a time-domain averager acts as a comb filter,[524a] in that it selectively emphasizes components of frequency F, $2F$, $3F$, $4F$ Hz, and so forth, where F Hz is the trigger frequency (Fig 1.44). An immediate consequence is that any subharmonic components are rejected; for example, components of frequency $F/2$ Hz do not appear in the averaged waveform. A second consequence of the comb-filter action is that an averager rejects any frequency components that are not harmonically related to the trigger frequency. Consequently, the averaged waveform can be quite misleading as the following example shows.

It is well known that if an averager is triggered at F_1 Hz while a subject views a light flickering at a frequency F_1 near 5 Hz, the averaged VEP approximates a 10-Hz frequency-doubled sine wave of large amplitude, but when a light flickering at some other frequency F_2 Hz is superimposed on the 5-Hz stimulus light, the averaged VEP may become very small (Fig 2.148). It is less well known that this weak F_1-Hz sine wave response can give a very misleading impression of the actual VEP: the real VEP may still be large. Figure 1.32F shows that the real response can contain a substantial component of frequency $F_2 + F_1$ Hz, and also other discrete components including those of frequencies $F_1 + 2F_2$, $2F_1 + F_2$, $3F_1 - F_2$, and $3F_1$, none of which appear in the averaged waveform because they fall between the "teeth" of the

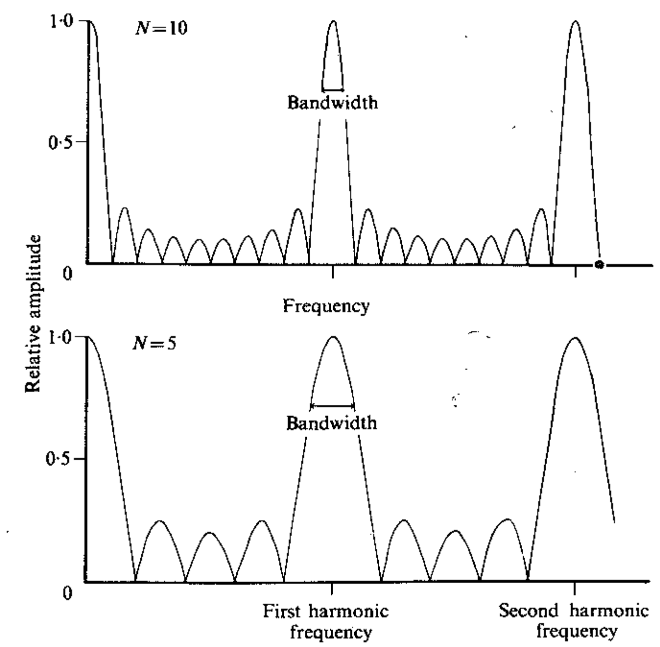


Figure 1.44
A Time-Domain Averager Is, in Effect, a Comb Filter
The comb filter action is illustrated for 10 sweeps (upper panel) and for 5 sweeps (lower panel). The peaks are located at multiples of the sweep frequency. (From Spekreijse H, Estevez O, Reits D: Visual evoked potentials and the physiological analysis of visual processes in man, in Desmedt JE (ed): *Visual Evoked Potentials in Man*. Oxford, Clarendon, 1977, pp 16–89. Reproduced by permission.)

comb filter. Figure 1.32D shows the time-averaged VEP to such a stimulus, and Figure 1.32F shows the VEP obtained from the same EEG sample as Figure 1.32D but by frequency-domain averaging (i.e., spectrum averaging) (Section 1.8.7). Note that this form of response distortion is inherent to the process of time-domain averaging and is quite different from the waveform distortion caused by overfiltering the EEG illustrated in Figure 1.22. The phenomenon is not specific to visual EPs. Figure 2.65 shows auditory EP components that would be rejected by time-domain averaging.

Time-domain averaging requires a trigger signal. A different kind of averaging can be carried out in the frequency domain rather than in the time domain. This kind of averaging, called spectrum averaging, improves the signal-to-noise ratio of any periodic components in the signal but does not need a trigger. It has the feature that it can detect a signal consisting of many different frequency components that an averager in principle cannot faithfully record. Spectrum averaging is discussed in Section 1.8.8. Section 1.8.7 describes how computer analysis by discrete Fourier transform can detect all the frequency components present in the VEP and indicate their powers. Sections 1.8.3 and 1.8.6 describe a simple method that gives both phase and amplitude, not only for harmonic but also for subharmonic and cross-modulation (e.g., $F_1 + F_2$ and $F_1 + 2F_2$) components.

Trial-to-Trial Stationarity

So far it has been assumed that both "noise" and signal are *stationary processes* from trial to trial, in other words that a description of either noise or signal that is valid at any moment is equally valid at any other moment. In practice, confusion may result from treating records of EEG noise and EPs as though they were ideal stationary processes. For example, the signal-to-noise enhancement obtained in practice by summing a large number (N) of EEG samples may fall well below the predicted value of $\sqrt{N}$. Examples of situations where serious departures from stationarity may be observed are (1) for late cognition-related components whose amplitude depends on surprise, selective attention, or arousal (see Section 2.2); (2) when N sweeps are summed during experimental procedures that are so taxing and prolonged that the subject grows fatigued so that the EP either changes in amplitude of phase or becomes increasingly variable; (3) summations of N sweeps during which the subject becomes so bored and sleepy that alpha activity grows progressively more obtrusive.[11]

The comparison of EPs with different stimuli is often hindered by slow EP variations resulting from physiological changes. This is especially true for experiments that extend over several hours and for experiments that are interrupted. *Order effects* may be introduced if stimulus parameters are varied in a regularly progressive manner (e.g., from dim to bright illumination). If regularity is unavoidable it is best to repeat the stimulus sequence in reverse order. A widely used method for minimizing these problems is randomization of the sequence of stimulus presentations.[603]

Conclusions

This section has emphasized the point that an averaged VEP does not necessarily give even a tolerably accurate indication of any single-trial response. One consequence

[11] Perry and Childers (Ref 1841, p 16) show a rate of increase of summed EP amplitude that is roughly proportional to N from a total of 120 to 960 flashes delivered at 4 flashes per second. Clearly, their subject's state remained effectively constant from 30 to 240 sec during the experiment. On the other hand, and quite apart from changes in the subject's state, for auditory stimuli it seems that stimulus repetition rate in itself may affect the way in which EP amplitude grows with an increasing number of stimuli. Milner (Ref 1621a, Fig 9.5) shows that summed EP amplitude rises roughly proportional to N for repetition rates slower than one stimulus in 7.5 sec. For rates faster than one in 3 sec, however, the higher the stimulus rate the greater is the departure from proportionality to N. At a rate of two stimuli per second, the amplitude of the EP rises roughly proportional to $\sqrt{N}$.

of this fact is that if two averaged VEPs are found to differ, it is not immediately obvious whether the difference reflects (1) a difference between two groups of approximately homogeneous single-trial responses, or (2) different trial-by-trial variability of VEP amplitude or latency within the two samples, or (3) different noise characteristics under the two different conditions. Conversely, two averaged VEP waveforms may look similar only because the averaging process has rejected some substantial component of the VEP as, for example, in Figure 1.32.

1.5.4 Signal-to-Noise Ratio of the Averaged Evoked Potential: Reliability, Noise Intrusion, and Signal Variability

It is sometimes asserted that noise averages to zero. This is not so. Signal averaging gives a $\sqrt{N}$ improvement in signal-to-noise ratio if the basic assumptions listed in Section 1.5.2 are satisfied, and this rate of improvement often seems frustratingly inadequate. An averaged VEP *always* contains a noise component, and this is often not negligible. In practice, the signal-to-noise ratio in averaged VEPs is typically between about 2 : 1 and 6 : 1 and, although this is a considerable improvement over the 0.05 : 1 to 1 : 1 ratios in the unprocessed EEG, by no means does it represent complete rejection of noise.

The residual of random noise in the averaged EP should be distinguished from a second cause of EP unreliability: trial-to-trial variability of the EP signal itself. Evoked potential amplitude has some variance about the mean and the same holds for EP latency. In some experimental situations this signal variability is of central interest (Sections 1.7.1 and 1.7.5). Progressive changes with time of the mean and the variance (i.e., signal nonstationarity) are discussed in Section 1.4.5.

This section discusses ways and means of estimating the reliability of averaged EPs. Recording a visual EP with the stimulus light occluded or an auditory EP with the sound muffled provides only a rough estimate of the noise level in EPs recorded during stimulation, because spontaneous EEG activity can be quite different when the subject is paying attention to the stimulus and when the stimulus is not present. In particular, alpha activity is often stronger in the unstimulated case. Rather than providing a noise estimate, the chief value of recording a light-occluded or sound-muffled average is to confirm that the waveform recorded by the averager really is evoked by the sensory stimulus rather than being electrical or magnetic pickup. It is good practice to record at least one such control average early in any recording session.

Noise intrusion in the averaged EP can be estimated by recording with the stimulus still present, but with the trigger rate set slightly different from the stimulus rate *without the subject's knowledge.*

Better still, a noise estimate can be obtained from the same EEG record as the averaged EP. Some averaging computers indicate the reliability of the averaged EP by displaying standard deviations computed from the same

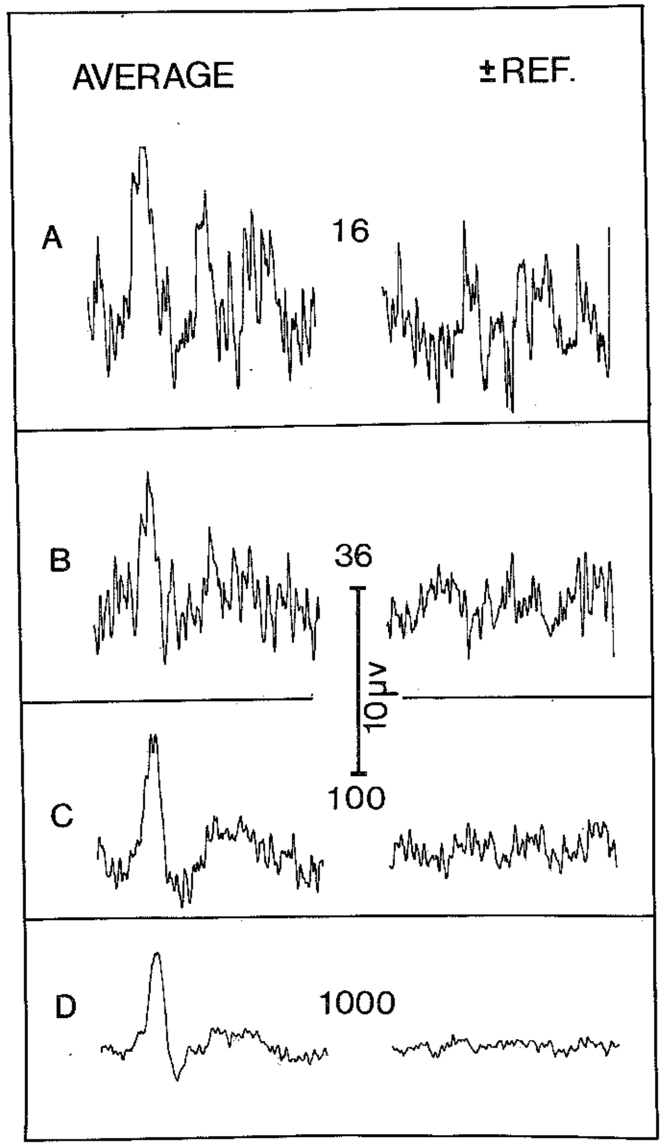


Figure 1.45
Noise Intrusion in the Averaged EP
The traces in the left-hand column show the progressive emergence of the VEP after first summing, then normalizing 16 (A), 36 (B), 100 (C), and 1,000 (D) single-trial samples of the EEG. These data are reproduced from Figure 1.41. The traces in the right-hand column are waveforms computed by adding and subtracting successive single-trial EEG samples before normalizing, so as to eliminate the time-locked signal. In each panel, the right-hand trace can be regarded as the noise level in the left-hand trace.

N samples of EEG from which the average is computed. (Ruchkin[2142] indicates how a general-purpose computer can be programmed to provide standard deviations.) An alternative way of assessing EP reliability is to compute a quantity called the "plus and minus reference" (i.e., ±Reference) as originally suggested by Schimmel.[2211a] All the traces in Figure 1.45 were obtained from the same tape-recorded EEG sample as the traces in Figure 1.41. The left-hand column of traces in Figure 1.45A was computed in the standard way by adding 16 successive samples of EEG, then normalizing the sum. The right-hand column of traces was computed in a different way: the 16 successive EEG samples were alternately added and subtracted, then the sum was normalized. If we regard the EEG as being composed of a sequence of 16 identical single-trial EPs added to unrelated noise then, because of the successive additions and subtractions, the right-hand trace in Figure 1.45A will contain no EP, but the noise power will be approximately the same as in the averaged waveform shown in the left-hand trace. Because the right-hand trace in Figure 1.45A has almost as high an amplitude as the left-hand trace we conclude that residual noise amplitude dominated the averaged left-hand trace. Presumably, this is because only 16 trials were averaged, giving a theoretical signal-to-noise enhancement of only about four times (i.e., $\sqrt{16}$ times). The amplitude of the ±Reference was considerably smaller after 100 trials, giving a theoretical 10-fold enhancement (i.e., $\sqrt{100}$ times) of signal-to-noise ratio (Fig 1.45C). After 1,000 trials had been averaged the ±Reference was smaller still, giving a theoretical 32-fold enhancement of signal-to-noise ratio.[12] A comparison of the left- and right-hand columns in Figure 1.45 shows that the empirical signal-to-noise ratio in the averaged EP was about 3 : 1 after 100 sweeps and about 9 : 1 after 1,000 sweeps.

The ±Reference traces of Figures 1.45A–D provide a useful indication of the degree to which the averaged VEPs of Figures 1.45A–D are contaminated by noise. On the other hand, the ±Reference provides little indication of possible signal nonstationarity such as, for example, progressive changes in the amplitude of single-trial VEPs.

More particularly, the ±Reference does not directly indicate which features of the averaged waveform are reproducible. This vital question is best addressed by the simple and popular expedient of repeating each recording one or more times and superimposing different traces. For example, the superimposition of the 500-trial averages in Figure 1.41F demonstrates the excellent reproducibility of the VEP's main features. It makes little sense to even think of deciding whether the two EPs are the same or different before the reproducibility of the two waveforms has been firmly verified as illustrated in Figure 1.41F. Note, however, that, although this procedure can detect both stimulus-linked changes in the EP and slow progressive changes in the EP, neither it nor the ±Reference detects progressive changes in the EP due, for example, to adaptation that may occur *within each N-trial averaging period.* This and other problems arising from nonstationarity are discussed in Section 1.9.

[12] If an artifact reject facility is used, the ± Reference technique is less straightforward. It may be necessary to ensure that the number of sweeps added is equal to the number subtracted, for example, by feeding a "reject" signal to the hardware device that multiplies the EEG by +1 and −1 during alternate sweeps.

1.6 Analysis of the Averaged Transient Evoked Potential Waveform

1.6.1 Data Reduction and Basis Functions

Two stages of data reduction can be recognized in EP research: (1) extraction of the evoked potential from the EEG and (2) description of the evoked potential in terms of the linear sum of a few *basis functions.*

There are two main classes of basis function. The first class is fixed basis functions that do not depend on the data. These include sine/cosine (Fourier), Walsh (square waves), Haar (square waves), and polynomials. The second class of basis functions depends on the data, and is created by multivariate procedures such as principal component analysis. The familiar "peak analysis" procedure (Section 1.6.4) is a hybrid in which experimenters attempt to describe a prototypical set of averaged EP waveforms as the sum of a few peaks, and then use these same empirically defined peaks as basis functions to describe averaged EP waveforms recorded in subsequent experiments.

Fixed Basis Functions: Sine/Cosine, Walsh, Haar, Decaying Sinusoid, and Polynomials

CHOOSING BASIS FUNCTIONS. Here we consider the analysis of a waveform (e.g., an averaged EP) into some fixed set of basis functions. The waveform (the averaged EP in this case) is analyzed into components whose *linear sum* recovers the original waveform.

There are several criteria for choosing a particular set of basis functions, including: (1) orthogonality (i.e., the basis functions are independent of one another in the sense that movement along one direction in space is independent of movement along a perpendicular direction); (2) ease of mathematical manipulation; (3) a well-developed and thoroughly explored mathematical basis,