ELSEVIER

International Journal of Psychophysiology 32 (1999) 129–150

INTERNATIONAL JOURNAL OF PSYCHOPHYSIOLOGY

Identifying and reducing noise in psychophysiological recordings

Tim R.H. Cutmore[a,*], Daniel A. James[b]

[a] *School of Applied Psychology, Griffith University, Nathan, Queensland, Australia 4111*
[b] *School of Micro-Electronics, Griffith University, Nathan, Queensland, Australia 4111*

Received 21 April 1998; received in revised form 26 November 1998; accepted 10 February 1999

Abstract

Psychophysiology continues to be a widely used methodology in the study of human behaviour, emotion and cognition. The new researcher is faced with a number of problems in the recording process since the desired physiological signal must be isolated from a variety of noise sources. Precautions and strategies that can be implemented in setting up the recording equipment and isolating the subject from interference are described. There are also a number of software techniques that can be applied to improve signal quality after the data have been acquired. An overview is provided of hardware and software methods used to maximise the signal quality.

Keywords: Noise; Psychophysiology; Hardware; Software; Filter; Data processing

1. Introduction

The recording of psychophysiological signals has provided important insights into the relationships between cognition, behaviour and physiology. This technology has, therefore, been taken up by researchers in a diversity of fields including sports medicine, experimental psychology, neurology, neuropsychology and clinical psychology. This overview is intended primarily for those who are relatively new to the use of psychophysiological methods in research and practice. Thus, it provides introductory material, particularly with respect to noise sources. On the one hand, the paper is intended to provide an accessible account of the engineering and technical basis of noise production and reduction as well as practical guidelines on useful approaches. On the other

* Corresponding author.

0167-8760/99/$ - see front matter
PII: S0167-8760(99)00014-8

hand, the paper does not attempt to be comprehensive in sampling the very wide range of disciplines that use psychophysiological methods. Some general references are provided here to help bridge this gap. Andreassi (1995) and Hugdahl (1995) each contain a broad overview, including chapters on a wide range of measurement techniques and applications; Jennings and Coles (1991) is an excellent general reference from a cognitive perspective including topics on attention, information processing and action; Surwillo (1990) provides a perspective from clinical psychology, including discussion of applications to emotion, biofeedback and deception detection; Nasralla and Weinberger (1986) contains chapters that address the utility of psychophysiology in the study of schizophrenia.

Ideally, one would like to isolate only those signal characteristics which are important to understanding an underlying cognitive or physiological process. Typically the process involved in arriving at such signal information involves a series of steps. Each of these steps has the potential to degrade the signal by introducing 'noise' or by occluding important signal components. This paper examines this problem by classifying noise sources and describing basic methods for dealing with the issue of extracting signal from noise at each critical stage of data processing, from data acquisition to summary measures. It is intended as an introductory overview which should be helpful to new researchers by providing, in the final section, a practical guide and troubleshooting checklist for noise identification and reduction.

2. Defining terms for noise characteristics

By transducing a bodily change into an electrical form, a psychophysiological recording consists of a series of voltages called a waveform. Noise is defined as the set of unwanted components within a waveform, whereas the signal is the set of components one wishes to isolate and quantify. Equipment used by psychophysiologists for the measurement of biological signals is both complex and technical in nature. Devices like transducers, electrodes, amplifiers and other recording equipment need to be carefully examined to ensure they perform adequately in this specialised field. To comply with this, manufacturers produce specifications for their products, which need to be examined carefully to determine their suitability and limitations. Below are some important defining terms used throughout the electronics industry.

2.1. Signal comparison and gain

Often it is useful to compare two signals to determine the functionality of a device. Gain refers to a comparison of input and output characteristics of certain devices. A gain of unity means the output matches that of the input. A gain greater than unity is an output larger than the input by that factor. Voltage gain is expressed as:

$$\text{gain} = V_1/V_2 \tag{1}$$

where V_1, and V_2, are the voltages of two signals (e.g. these may be the output and input voltages of an amplifier). During signal acquisition V is a potential difference between two spatially separated points (e.g. between an active electrode and a reference electrode). Gains of biological amplifiers are typically of the order of many thousands since biological signals are small relative to typical voltage ranges that are used to map the signal into a representation (e.g. in computer memory). Gains of less than one are typically produced by filtering processes to reduce undesirable signals. A widely used term for gain is the full scale deflection (FSD). This refers to the maximum amount of voltage input that can be accommodated by an amplifier before there is distortion or clipping of the signal (i.e. the amplifier fails to function properly).

Since the range between signals for comparison may be many orders of magnitude, sometimes the comparison is expressed on a logarithmic scale. In this case the comparison is expressed in power units (i.e. voltage squared with the result expressed in decibels (dB).

$$\text{gain} = 10 \log_{10}(V_1^2/V_2^2) \quad \text{decibels} \tag{2}$$

For every doubling (or halving) of the power of one signal relative to the other, this results in a gain of approximately ± 3 dB, a useful increment to remember. For example, a gain of 24 dB is indicative of one signal being 256 (2^8) times the power of the other.

2.2. Measures of signal strength: peak-to-peak, base-to-peak, rms

Signal (and noise) amplitudes vary greatly over time and are made up of many frequency components. For example, consider a heart beat as measured by the electrocardiogram (ECG). The voltage measured is near zero most of the time except for large excursions known as the PQRS complex. This segment of the ECG contains peaks not present in the baseline segment of the signal, which consists mainly of slowly changing low amplitude components. Since one usually wishes to express the waveform magnitude of a segment of the signal (as opposed to a single point), and the voltage is varying over the segment, a summary measure is needed. One summary measure that may be useful (e.g. when considering FSD for an amplifier) is the peak-to-peak value. This value over a period of time is used to represent the maximum signal excursion and is a useful measure when there are frequent peaks appearing in the signal. The base-to-peak measure of voltage is the maximum excursion from the average voltage ($\overline{V}$). Both the above measures of voltage can be misrepresentative if the signal is predominantly small with only occasional large excursions or peaks. An alternative quantity, the root-mean-square (rms), is a measure of signal central tendency that has been widely adopted throughout the scientific and technical communities. It is, essentially, a mean deflection about the average voltage. For a series of n voltage measurements

$$\mathrm{rms} = \sqrt{\left(\left(\Sigma(\mathrm{V} - \overline{\mathrm{V}})^2\right)/n\right)} \tag{3}$$

Many signals are quite variable. For example, the power generated from the commercially available grid in Australia is 240 V. This is actually an rms value, and the corresponding base to peak voltage may be nearly 340 V. If the signal is a sine wave then the rms value is computed simply by dividing the base to peak value by the square root of two (Diefenderfer and Holton, 1994).

2.3. Signal to noise ratio

It is clear that noise that is small in relation to the signal is desirable. It is useful, therefore to quantify the relative amount of noise in a signal. For example in an ECG waveform of 1000 μV rms there may be a noise component of only 10 μV rms. The gain Eq. (2) is used to compute a signal to noise ratio (SNR) and, in this example, the SNR is 40 dB which is not likely to be of concern unless there is a small signal component within the ECG segment one wishes to identify. If, however, we are measuring event-related brain potentials (ERPs) that are only 10 μV rms in amplitude and the noise component is 20 μV rms then there exists significant signal degradation due to noise (SNR = -6 dB). Note that the negative sign indicates that the signal is smaller than the noise.

2.4. Frequency domain and bandwidth

The Fourier theorem asserts that any finite, continuous waveform segment may be described to an arbitrary degree of precision as the sum of sine and cosine waves of specified amplitudes and phases (Brophy, 1990). A signal may thus be conceptualised as consisting of many frequency components (i.e. a spectrum). The Discrete Fourier Transform (DFT) operates on digitized data and an efficient algorithm, the Fast Fourier Transform (FFT), has been defined for its computation (Karl, 1989). Press et al. (1986) provide a reasonably accessible account of the mathematical basis of the Fourier transform and its associated spectral methods in chapter 12.

While sine and cosine waves of differing amplitudes may be used to represent the frequency information in a time series, equivalently a series of shifted sine waves of different frequencies will also make the same frequency content explicit. The input to the DFT takes N points from a time series sampled at equal intervals (or, equivalently

at a particular frequency, f) and the output is an array of N complex numbers. Suppose we have 5 s of signal composed of only 10 Hz alpha rhythm and 50 Hz line noise and we sample this 500 times per second. The sequence of complex numbers describes a set of discrete frequency components (sine waves) in terms of their amplitude and phase (or shift). In the example, only two complex numbers will have significant amplitudes. A complex number (C) can be conceptualized as a point in the complex plane with two dimensions (real and imaginary). The amplitude of a discrete frequency component (e.g. 10 Hz) is defined by the length of the vector from the origin to the point. Taking the natural logarithm of C yields another complex number $\ln(C)$. The absolute value of the imaginary dimension of $\ln(C)$ provides the phase information for the sine wave of interest. Given that the continuous signal has been sampled at a rate f, the maximum frequency component is $f/2$, the Nyquist frequency. The set of complex numbers spans from the negative Nyquist frequency to the positive Nyquist frequency. In the example, the Nyquist frequency is 250 Hz, components will span from -250 Hz to $+250$ Hz and there will be 2500 of them (5 s $\times$ 500 Hz). Often the phase information is not of interest. The negative frequency components are sometimes of interest (for example when performing Fourier filtering or computing the cross-correlation between two signals). Considering only the positive frequencies, the DFT results in $N/2+1$ frequency components from 0 Hz to the Nyquist with a resolution of 1/(epoch duration). In the example, the positive frequency components would be from 0 to 250 Hz with a resolution of 0.2 Hz.

Spectral analysis is an important tool for examining biological signals and the limitations of equipment used to record them. The power spectral density (PSD) is a commonly used frequency representation which is based on the Fourier transform. Using Welch's method (Welch, 1967), it is computed by taking non-overlapping windows within a waveform, tapering the ends (e.g. with a Welch window), computing the FFT for each window, averaging and finally squaring (to convert from amplitude to power units). Scaling of the resultant spectrum may be provided in terms of μV^2, $\mu V^2/Hz$ or may be normalized, depending on the details of the algorithm used. When combining (e.g. averaging) or otherwise comparing results of Fourier analysis from different sources, possible differences in scaling need to be considered. Different definitions of the Fourier transform result in different scaling factors. The possible effects of windowing on scaling also need to be considered; however, beyond this Press et al. (1986) notes that generally the choice of window results in 'effectively no difference' (p. 425). Finally, Parseval's theorem asserts that the total power of the PSD is equal to the rms^2 (or simply mean squared) amplitude of the continuous time series (Press et al., 1986, p. 384).

A PSD graph of how a transducer or amplifier responds to all frequency components permits a determination of its amplitude fidelity and thus suitability for a particular application. Detection of phase distortion will require examination of the phase components of the Fourier transform. High fidelity across a critical frequency band indicates low signal distortion as it passes through the electronic devices. If a biological signal has a known frequency component (e.g. the alpha rhythm of the EEG) then PSD analysis will permit quantifying its strength in the acquired waveform (e.g. as a percentage of the power within a given frequency band). Frequency filters, described in Section 5.1.1, are also based on concepts related to the Fourier transform.

Some of the above concepts are illustrated in the following example. Fig. 1 illustrates the frequency composition concept using two pure sine waves. Fig. 1a shows a 6-Hz 'signal' at 2.0 μV base to peak (1.414 μV rms). Fig. 1b shows a 50-Hz 'noise', at 1.0 μV base to peak (0.707 μV rms). The composite waveform which would present itself to the psychophysiologist is shown in Fig. 1c and it is 2.98 μV base to peak (1.58 μV rms). It should be noted that rms^2 voltages are additive. The PSD graph for the composite signal (Fig. 1d) clearly shows the presence of the two components.

Finally, Dumermuth and Molinari (1987) provide an excellent overview of the utility of spectral analysis on EEG data, describing several as-

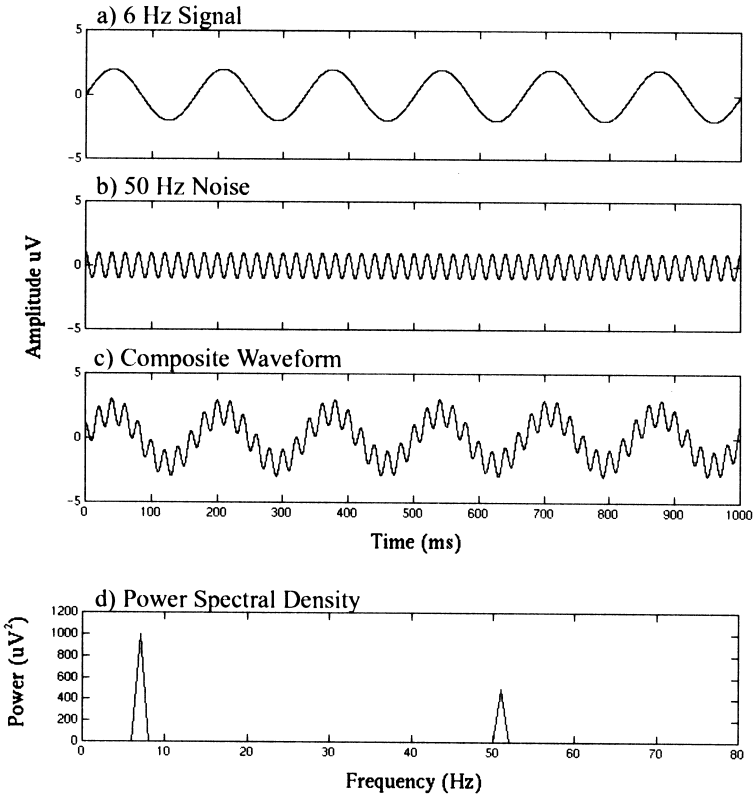

Fig. 1. An example of how a noisy signal may be construed in terms of its frequency components. (a) A 6-Hz signal of interest; (b) line noise from a 50-Hz source; (c) the composite waveform due to superposition of the two sources; (d) the frequency power spectrum showing the relative contribution of the two sources.

sumptions that need to be met for the results to be valid. In general, Fourier analysis techniques assume stationarity of the signal and thus caution must be exercised in applying the DFTs to waveforms which contain transient signals such as occur in ERPs and ECG.

3. Noise sources

There are three main categories of noise: fundamental noise, electromagnetic interference (EMI) and endogenous noise. There is little to be done about fundamental noise once the recording equipment has been purchased. However, knowing its influence will help in determining the limits of resolution in making recordings. EMI and endogenous noise can often be controlled to a significant extent by careful set-up of the recording environment, subject preparation and instruction.

3.1. Fundamental noise

Fundamental noise is that which is due to the physics inherent to the recording process. There are three subtypes: thermal noise, shot noise, and flicker noise (Diefenderfer and Holton, 1994). These can be quantified using the noise figure (NF). This is defined with respect to the amplification relative to the input to an amplifier. It is the ratio of SNR at the input to the SNR at the output. For an ideal amplifier NF equals one. That is, the electronic components of the amplifier do not introduce any additional noise. In practice it will be greater than one. These noise sources intrinsically are associated with electronic

circuitry and therefore may potentially invade psychophysiological recordings depending on the frequency range of interest, particularly when high amplification is used as in ERP research.

3.1.1. Thermal noise

Random movement of charge carriers such as electrons result in fluctuations in voltage being added to the signal. The intensity (V) of thermal noise measured in rms volts is a function of the temperature (T), resistance (R), and bandwidth (B):

$$V^2 = 4kTRB, \tag{4}$$

where k is the Boltzmann's scaling constant (1.38×10^{-23} J/K) temperature is in Kelvin, resistance in Ohms and bandwidth in hertz. This noise is frequency-independent and therefore is referred to as white noise. One of the practical implications of this is the need for adequate cooling of components. Amplifiers and related equipment usually come in ventilated casings with cooling fans. It is best not to operate such equipment in a hot room with the casing removed as this may result in components overheating, introducing thermal noise. A 100-kΩ resistor at room temperature has an rms noise value of approximately 0.5 μV as measured across a 1-kHz bandwidth. This level of noise may seem insignificant but a poor quality pre-amplifier (that is, one with a relatively high NF) could magnify this noise source several fold (Diefenderfer and Holton, 1994).

3.1.2. Shot noise

This noise source (also known as Schottky noise) is due to the presence of a junction in the recording circuit at which stochastic fluctuations of charge can occur. Such a junction exists between the electrode and skin and in many electrical components of amplifiers. The intensity of shot noise measured as current varies as a function of DC current (i), and bandwidth (B) is:

$$I^2 = 2qiB, \tag{5}$$

where the constant q is the charge of an electron (1.59×10^{-19} Coulombs). This noise source is generally negligible below 1 kHz.

3.1.3. Flicker noise

This noise source (also known as $1/f$ noise) varies inversely with frequency, particularly below 1000 Hz and therefore may be of concern to psychophysiologists, although high quality pre-amplifiers (with NF near 1 in the 0–1-kHz range) minimises its influence. This noise source is one of the reasons that DC recordings are more problematic than recordings made with a high-pass filter (or with AC-coupled amplifiers). The physical basis of flicker noise is poorly understood and it is often regarded as the fundamental noise that remains after shot and thermal noise have been accounted for (Diefenderfer and Holton, 1994).

3.2. Electromagnetic interference (EMI)

These noise sources are a direct result of electromagnetic signals generated in the surrounding environs. These manifest themselves as radiated signals that are either wireless or propagated through cabling (the latter is referred to as line noise).

Since these noise sources have known generators that obscure the signal of interest they are also referred to as interference. The sources include mains supply, brushes on electric motors, fluorescent lights, cathode ray tubes (CRTs), computing equipment and so forth. Many of these sources are inherent to most modern psychophysiological recording laboratories. These sources may generate transient impulses, as for example when a solenoid switch engages an electric motor, or continuous, such as mains supply at approximately 50 (or 60) Hz and its harmonics. Some continuous sources may not have stable frequency characteristics, such as an electric motor undergoing a variable load. Typically, biological signals of interest to the psychophysiologist are below 100 Hz so interference from kilohertz (e.g. radio wave) sources and above is not considered a problem since there exist straightforward frequency filter techniques for dealing with them (see later).

3.2.1. Radiated noise

Electromagnetic radiation is present from a variety of sources. Essentially, anywhere that a radio receiver can receive a signal may be an area susceptible to radiation. Neighbouring devices such as computers, fans, electric heaters, light sources and so forth are all potential radiated noise sources. The radiated energy has both electric field and magnetic field components. A fluctuating electric field can induce unwanted voltages in devices whilst magnetic fields may generate currents in any loops formed in circuits. This magnetic interference is sometimes referred to as 'eddy currents'. Electrodes are particularly susceptible to these forms of interference. Any electrode that is poorly connected (or unused but connected to the amplifiers) can act as an antenna for electric fields and loops in the leads can result in magnetically-induced currents. This may permit line noise, cross-talk from other recording channels, and in some cases, drift due to nearby slowly changing voltages sources.

3.2.2. Line noise

Line noise is generated by unwanted signals that propagate through cables and connectors. Any device connected to mains supply is also connected by way of power points to all other devices on the same circuit. This inter-connectivity allows any interference from these other devices to be propagated to psychophysiology instrumentation. Computers, devices with motors, etc., all generate some 'back electro-motive force' which, if not properly dealt with, is a source of noise. Typically this may manifest itself as a voltage 'spike', in a recording if a device, connected to the same power circuit as the bio-amplifiers, is switched on or off during the recording process. Some research protocols may require the use of electronic switching for stimulus control and this could be a possible source of interference.

3.3. Endogenous noise sources

While the above two noise sources are due mainly to factors external to the subject, endogenous signals are generated by a number of sources within the human body. These sources may, in fact, be the signal source of interest; otherwise, they may be a nuisance in the recording process and are, therefore, regarded as noise or artefact. For example, the human heart generates an electrical signature with several components, the PQRS complex, which may be regarded as a noise source if one is attempting to record EEG under some conditions. Muscle activity generates high frequency waveforms known as the electromyogram (EMG). While this source may be regarded as a signal in many instances, it is regarded as a noise source when recording EEG, ERPs and ECG. Rotation of the eye in its socket yields a source for electro-oculographic (EOG) signals. The EOG may be used to record eye movements but is a problematic noise source when recording EEG and ERPs. Table 1 provides a summary for several commonly recorded endogenous electrical signal sources. It also includes the typical magnitudes of the voltages recorded (which vary depending on stimulus paradigm, recording sites and other factors), along with common exogenous noise sources and endogenous noise sources that may obscure the signal (Stern et al., 1980; Cacioppo and Tassinary, 1990).

Table 1
Summary of endogenous electrical sources and possible noise influence

Source	Signal magnitude	Exogenous influences	Endogenous influences
Skin potential response (SPR)	2–80 mV	Robust	EMG
Electrooculogram (EOG)	10–500 μV	Line noise	EMG (esp. facial muscles)
Electrocardiogram (ECG)	80–2000 μV	Robust	EMG
Electromyogram (EMG)	10–2000 μV	Line noise	ECG, EEG
Electroencephalogram (EEG)	1–100 μV	Elect. equipment, line noise	ECG, EMG, EOG, ERP
Event-related potential (ERP)	1–20 μV	Elect. equipment, line noise	EEG, EMG, EOG

It is apparent that not all body sources of electrical signals may interfere with recording a particular signal of interest. The psychophysiologist needs to be aware of the variety of electrical sources within the human body and decide whether they need to be dealt with in quantifying the target signal.

3.4. Active noise

During the acquisition (recording) phase, appropriate choice of analog filtering, amplification and digitisation will ensure that noise is not actively added to the incoming time series data. A general overview is given in Gevins (1987). The major steps and precautions in processing are considered in the following sections.

3.4.1. The amplification process

The physical construction of modern amplifiers gives rise to a number of possible sources of noise, apart from the fundamental noise sources noted above. The construction of amplifiers from integrated circuit (IC) components means there is a contribution from each stage of the amplification process. Because ICs are physical devices they have built in limitations such as response times (and delays). Amplifiers can only work within a certain range of voltages and so the gain of the amplifier must be selected carefully such that the output of the amplifier doesn't exceed these limits. When the range is exceeded the amplifier can only output its maximum value and so the output is said to be clipped. The implication here is that a larger biological signal cannot be amplified as much as a smaller signal. Thus, if multiple channels are being recorded, the peak-to-peak signal value should be known for each source (e.g. ECG, EOG, EEG, etc.) and the gain of the amplifier set accordingly so as to avoid clipping. Modern amplifiers used in psychophysiology are usually differential (or bipolar) amplifiers. That is, they measure the difference in voltage between two different inputs. This has some distinct advantages for noise reduction, though it is not without its limitations. Ideally if a noise component is identical on each of the bipolar inputs then they are presented to the amplifier in 'common model' and it 'rejects' the noise. The common mode rejection ratio (CMMR) specifies how well the amplifier rejects signals when they are in common mode. Typically, 100 dB noise reduction or better is seen on modern equipment. The psychophysiologist can minimise any noise that exists in the inputs by taking steps to ensure that electrode leads are as physically similar as possible in length, spatial orientation, and conductivity (see also Section 3.4.3).

3.4.2. Impedance considerations

The input impedance of the amplifier also plays a vital role in the amplification process. This is particularly important in the amplification of biopotentials where the skin impedance (or resistance) is usually large. It is important to ensure that the electrode–skin impedance is at least an order of magnitude smaller than the input impedance of the amplifier. This ensures maximum transfer of the signal to the amplifier. Fortunately, the input impedances to modern amplifiers is on the order of millions of ohms and so skin contact impedance is no longer so critical. Historically the development of amplifiers restricted the maximum allowable electrode–skin impedance to 5 kΩ. Today this is not so important though it is a useful standard to maintain as a 5-kΩ impedance ensures a reliable connection whereas a 20-kΩ impedance may indicate a poor connection and admit spurious signals from EMI sources.

3.4.3. Electrodes

The electrode and gel act as a transducer between ionic currents in the body and electronic currents in the electrode wire. Different ionic–metal interfaces typically have different transduction properties. Thus, it was noted in Section 3.4.1 that physical properties of electrodes should be as similar as possible to take advantage of CMMR (e.g. dirty or non-matching electrodes should be avoided). In this section a few other points are noted regarding electrodes.

1. Placement and fixation of primary recording electrodes is, of course, of central importance to obtaining good SNR and standards have

been published for all of the major recording methods mentioned in Table 1.

2. Choice of appropriate reference and ground electrodes is less well standardised and some trial and error may be necessary for a particular application. For example, improper placement of an EEG reference electrode may permit ECG 'noise' into the recording. Scerbo et al. (1992) point out important differences in electrode locations when recording electrodermal activity.
3. Tassinary et al. (1990) point out the importance of electrode storage considerations for the widely used silver/silver-chloride electrodes, advising against simple dry storage in air. Poor storage can lead to offset potentials and electrode instability (i.e. noise injection) during recording; for example, DC drift.
4. The electrode leads may need to be shielded if EMI presents a problem. This can be recognised as line noise in the recording. Keeping the leads as short as possible also helps to reduce EMI.
5. Electrodes should all be made of the same metal, otherwise unwanted currents may be produced. This may be identified as DC drift in the recording.

3.4.4. The conversion of data to a digital signal

Transducers on the body that respond to changes in physical state give rise to continuous time series waveforms, such as voltage changes in electrodes. Physiological recordings today are inevitably converted to a digital form (discrete time series) so that a computer can be used for signal processing. The conversion process represents a compromise between preserving the signal verbatim and the limitations of processing power and storage available on the equipment used. Discretisation of the signal occurs in both time and amplitude when a signal is digitised. Unlike an analog voltage which varies continuously over time, discretisation involves repeated sampling or measurement to build a semblance of the continuous signal. If a signal is sampled too slowly then important temporal detail in the signal is lost about high frequency components. As a general guideline, the Nyquist sampling theorem asserts that the sample rate should be at least twice the maximum frequency component of interest in the signal. Thus, for an ECG signal, a sample rate of 100 times per second (or 100 Hz) may be sufficient to retain the necessary detail of the PQRS complex though for some applications a greater sample rate may be desirable. Sampling at 500 Hz provides much more temporal detail than sampling at 100 Hz, however, the storage capacity required for a given time interval will be five times greater. A final concern regarding the appropriate sample rate is the possibility of 'alias noise'. This form of noise is introduced when there exists sizeable noise frequency components above the chosen sample rate. This high frequency activity will only be sporadically sampled if a lower sample rate is used: sampling sometimes stores a peak, a trough or intermediate value in a pseudo-random fashion, giving a false appearance of the signal. To avoid this, it is best to apply a low-pass analog filter set well below the Nyquist frequency, (e.g. half Nyquist) to attenuate the high frequency noise to a negligible level before digitising the signal (see also, filter techniques in Section 5.1.1). Alternatively, if storage is not a problem, then sampling at 5–10 times the fastest component of interest will avoid amplitude distortion due to aliasing.

The amplitude resolution of the digitised signal is dependant on the number of bits used by the analog to digital converter (ADC). Modern computers store only 0′s and 1′s which are grouped into 'words' of 16, 32 or 64 bits, depending on the underlying architecture. An ADC typically uses word sizes of 8, 12 or 16 which map a limited voltage range (e.g. ± 5 volts) in the input into a set of discrete values. With an 8-bit word (also known as a 'byte') 256 voltage values may be discriminated; with 16 bits this becomes approximately 64 K values. The range within which the converter operates has a large effect on the amplitude resolution of the digitised representation of the waveform. If the range for the converter is much larger than the signal to be converted (e.g. signal peak-to-peak is 10 mV and the ADC range of conversion of ± 5 V), then many of the bits devoted to the conversion process are not used

and hence wasted. This can effectively turn a 12-bit converter into a 4-bit converter with the corresponding loss in resolution. For this reason the gain of the amplifier needs to be carefully selected such that, after amplification, the voltage range of the input signal is close to (e.g. approx. 50%) the ADC conversion range, yet avoids possible clipping. Finally, caution is needed in using ADCs since, depending on the technique used, temporal desynchronisation between channels can occur (Lutzenberger and Elbert, 1991) when multiple inputs are being recorded simultaneously. The practical advice here is to know the properties of the ADC being used and thus the possible limitation in time shifting across channels.

4. Physical barriers to insulate against noise

4.1. Equipotential grounding and reference electrodes

Measurement of biopotentials using electrodes involves the measurement of the difference in electric potential between two sources. To transmit information from one part of the acquisition process to another a third electrode that acts as a common reference point is sometimes required. This common point is usually a ground potential, sometimes called earth or zero volts. If many devices are present in the recording process and they are each referenced to a separate ground point, noise may be introduced because of variations between the different reference points. In addition, when ground points are separated by some distance noise can be induced in this path by ground loops and magnetic interference from other sources. Fig. 2 illustrates these concepts. While in many cases, only a few microvolts may be induced, this is a significant quantity with the measurement of some biopotentials.

To avoid the introduction of noise by the above processes, a common ground technique should be used. The common ground technique requires that all devices be connected to a ground at the same point. This can be implemented simply by connecting all devices to the same ground on the same power distribution point that powers all devices in use. Subjects should also be referenced to this ground point but never directly connected to it for safety reasons as will be discussed below.

For stand alone devices that run on battery power (thus reducing ground loop problems and other EMI problems by not being connected to anything else), all components of the apparatus should be referenced to a common point. By common convention this ground point would usually be the negative terminal of the battery used to power the system.

Typically, reference and ground electrodes are connected to a subject. The reference electrode is used as one input into a bipolar amplifier, whilst the ground electrode is the common ground point. While the subject ground is not necessary in principle, in practice there may be a significant common mode signal without this subject ground and this permits serious line noise contamination. Typically in a biological application the ground electrode is not actually connected to an equipment ground (as this introduces a safety hazard)

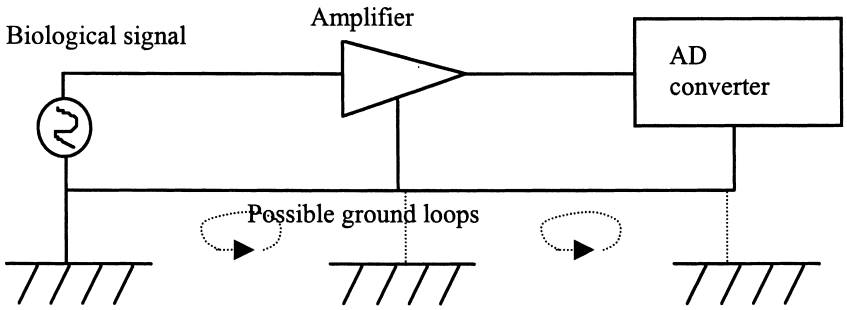

Fig. 2. A shielding strategy to connect all systems to a common ground. Note that there is a single ground point (solid line). If different ground points (dotted lines) were used for different components, ground loop currents could arise.

but rather is attached to an amplifier that is set to zero volts. This is known as the 'driven right leg' configuration (Webster, 1995). If the subject were actually connected to ground and some induced ground loops were to exist, then even these comparatively small induced currents could be lethal. Electrical failures are particularly dangerous when subcutaneous electrodes are used as the normal skin resistance has been bypassed permitting very high currents to be induced. Possible cardiac fibrillation can result. Optical isolation of the subject from line voltage can be achieved using either an opto-isolator or fibre optic transmission. In the latter, each side of the fibre has its own power for its circuit and communication.

Finally, related to the issue of choice of ground electrode is the choice of reference for a recording montage. It is the reference against which other 'active' channels are compared. Nunez (1981) and Nunez et al. (1997) discuss the important practical issue of choosing between different types of 'reference' electrodes. There are three main choices. The common reference is ideally an 'inactive' site defined by either a single electrode or average of a few others (e.g. linked ears) against which the active electrodes are compared as a difference in voltage over time. An average reference uses the set of active sites as a comparison for each one individually. A bipolar recording compares pairs of active sites. An alternative to these reference-based methods is to compute a current source density estimation which can be made with the use of Laplacian approximations provided a sufficient number of electrodes are used (Thickbroom et al., 1984; Nunez et al., 1997).

4.2. Electrostatic shielding

To reduce the amount of external noise that may contribute to induced noise sources in the recording, a shield can be placed between the noise source and recording system. The grounded Faraday cage provides such a shield to the electromagnetic noise in the environment by acting as an energy reflector. The cage need not be a sealed unit, but may have holes that do not permit undesirable frequencies to pass. As a rule of thumb the hole size should not be more than $1/10$ of the wave length of the noise source, where wave length is $1/$frequency times the velocity of the signal (typically propagated at the speed of light, 3×10^8 m/s), easily permitting holes of a few centimetres. Thus, a grounded wire mesh may also serve as a functional barrier, for example, around AC light sources. The degree of attenuation achieved by a Faraday cage depends mainly on the metal used in its construction. If the desired shielding is mainly from electrical sources then aluminium or copper works well.

The leads used for recording should also be shielded to provide a similar barrier to EMI. Similarly, any cables in the recording room (within the Faraday cage) should be shielded using braided co-axial wire that is earthed to the common ground point. Often a CRT is used to present stimuli to the subject. Placing the CRT outside the Faraday cage may degrade the image if it must be viewed through a grid, although a high quality commercial screen filter of conductive glass may be suitable. If the CRT is within the cage then shielding may be achieved by surrounding it by its own miniature Faraday cage and grounding it to the larger one. A low-interference option for stimulus presentation is the liquid crystal display that is common on laptop computers as these generate very little EMI.

Psychophysiology recordings of signals at the microvolt scale (for example, ERPs) typically require Faraday cage shielding from EMI sources. The relatively large signals from ECG, EOG and SCR can be acquired with good (in some cases shielded) electrode leads without the need for a Faraday cage. Test recordings in a particular environment should be conducted to establish what EMI shielding may be necessary to give the desired SNR characteristics.

4.3. Magnetic shielding

Magnetic shielding is difficult to achieve without using large quantities of ferrous materials (steel is an economical choice) or mu metal (which is costly), although the dual nature of EMI (alternating electric and magnetic fields) results in some attenuation of the magnetic field through electrostatic shielding. Magnetic interference arises

from induced currents being formed in loops of a conductor such as an electrode lead (see also Section 3.2.1). This induction effect is maximal when the loop is orthogonal to the direction of the magnetic field. By twisting active and reference leads around each other between devices (including people) equal and opposite currents are induced and these cancel each other out. Another alternative is to use co-axial connections, such as is used in some high quality electrode leads today.

5. Active noise reduction techniques

The techniques described in Section 4 provide effective means for reducing noise external to the subject. These are essentially passive techniques in the sense that a barrier is erected against broad classes of noise. In this section, we examine methods for identifying specific unwanted features of noise and summarise the active steps that are needed to reduce or eliminate them. These techniques are typically applied off-line after the data have been stored in digital form.

Some residual noise from external sources may persist in the recording even after shielding. More importantly, endogenous noise must still be dealt with. These limitations result in the presence of noise in the final stored recording. This section is subdivided into three sections on filtering techniques, statistical techniques and innovative techniques. Although it is useful to discuss these topics separately, there are some commonalities underlying all of these procedures that are worth noting. Nearly all of the techniques described are founded on the operation of a sum of products between a 'recognizer' vector and the raw time series data (also a vector) which is the recording acquired from the human subject. To appreciate how the recognizer responds preferentially to some pattern in the time series, consider the following example. Suppose the recognizer consists of the following pattern: $R = [0\ 1\ 0]$. It is applied to two different time series: $T1 = [0\ 1\ 0]$ and $T2 = [1\ 0\ 0]$. The element-by-element sum of products of the vectors (also known as the dot-product, DP) R and $T1$ is 1, whereas the DP of R and T2 is 0. At this simplest level, it could be concluded that R recognised $T1$ but not $T2$ to have a certain pattern. The pattern for the recognizer may be known in advance or may need to be built from the acquired data itself. In the above example, the recognizer was prespecified and acted as a kind of template against which various time series could be matched. The recognizer may also be referred to as a weight vector. Time domain filtering involves 'sliding' a weight vector along a time series and, at each time step, computing the DP which can be used to produce a new time series. This process is known as convolution. This approach is the basis of frequency filtering, template matching and averaging. Related techniques are used with discriminant analysis and principle components analysis (although these are better considered under the domain of statistical techniques), and neural networks and wavelet analysis which are considered in a final section.

5.1. Filtering[1]

A 'filter' may be conceptualised very generally to be a process that acts on a signal to emphasise certain of its features whilst attenuating others. This general formulation, therefore, includes frequency filters, template filters and signal averaging. An analog filter may be applied to the data prior to digitisation. Analog filters are implemented in electrical circuitry and thus operate rapidly on-line as the data are acquired. Once the data have been stored on a computer then digital filters may be used. Many engineering sources provide precise mathematical formulations of digital filter concepts. However, even introductory texts require familiarity with complex analysis techniques such as the Fourier transform (e.g. Cunningham, 1992). Fortunately, Cook and Miller (1992) provide an excellent tutorial overview of

[1]Although many commercial psychophysiology recording systems include software for basic signal processing, in practice these will need to be extended with custom built software. We have found the MATLAB™ system and signal processing toolbox to be a powerful and flexible programming environment for offline data filtering.

digital filters and Farwell et al. (1993) provide a reasonably non-technical account of digital filter concepts and their application to ERPs. The major elements that are necessary to appreciate the fundamentals, appropriate application and limitations of these filters are summarised below. These sections provide an overview of the three major filtering approaches: frequency filtering, template filtering and signal averaging.

5.1.1. Frequency filtering

It was noted above (Section 3.4.4) that the data acquired from psychophysiological recordings are originally continuous time series and, once digitised, become discrete time series. A time series may be described as consisting of a set of sine and cosine waves (frequency components) with no DC offset. Indeed, this is the basis of the Fourier transform which converts the time series (time domain data) into a set of amplitudes and phases of these simple periodic waveforms (frequency domain data). Frequency filtering consists of transforming the time domain waveform to remove unwanted frequency components. An important characteristic of some noise sources is their frequency characteristic. If the range of frequencies generated by a noise source does not significantly overlap with the desired signal, then frequency filtering can achieve dramatic improvements in the SNR. Caution must be exercised, however; if important attributes of the signal are manifest within the attenuation range of a frequency filter, then signal power may be lost. Frequency distortions can also be introduced inadvertently by misapplication of these filters.

To understand the operation of frequency filters and thus their advantages and limitations a few concepts need to be introduced. Frequency filters can be classified into four main types: low-pass, high-pass, band-pass and band-stop. A low-pass filter permits frequencies below a cutoff to pass with little attenuation. A highpass filter permits frequencies above a cutoff to pass with little attenuation (see Fig. 3). Band-pass and band-stop filters attenuate frequencies either outside or within a frequency range, respectively. If there exists a particular feature of the signal to be detected that has known frequency characteristics, then a band-pass filter which selects only that frequency range can be applied with advantage (Farwell et al., 1993). To appreciate the importance of the choice of filter method consider the following example: If frequency components of EEG between 30 and 80 Hz are of interest, but there exists a 50-Hz line noise in the recording, the use of a low-pass filter set at 45 Hz to attenuate the noise may leave frequencies near 30 Hz unattenuated but result in severe loss of information above 50 Hz. In this case, a better solution would be the use of a band-stop filter set at 50 Hz (also known as a 50-Hz 'notch filter').

Frequency filters (particularly analog filters) do not abruptly delimit the frequency band into admitted and omitted frequencies, but rather, provide a gradual attenuation with change in frequency. The 'cutoff frequency' used to describe a filter is not, therefore, an absolute measure. In fact, the cutoff is usually defined as the frequency at which the signal power has already been attenuated by −3 dB (that is, by one-half, Section 2.1). Fig. 4 shows two representations of a band-pass filter. In (a) the change in amplitude gain as a function of frequency is shown. It is clear that frequencies in the range of approximately 2–4 Hz are admitted with little or no reduction in gain, whereas frequencies above and below this are severely attenuated. To more precisely define the nature of this filter another representation is required which plots attenuation in decibel units. As noted above, the cutoff is defined as the point at which the signal is attenuated by −3 dB, so a decibel scale is necessary to see this directly. In Fig. 4b, these values can be read off the graph as approximately 1.8 Hz (low cutoff) and 4.5 Hz (high cutoff). In digital filter design the stop band is defined by setting a parameter. In this case it was set to −40 dB. In these frequency ranges (above and below the pass band) the signal is severely attenuated. Between the pass band and the stop bands there are transition bands (upper and lower). The 'rolloff' of a filter is a term used to describe this transition band. It is typically given in −dB per octave (or doubling of the frequency). Thus in (b) it can be seen that the rolloff to the upper stop band is approximately an attenuation of −37 dB between 4.5 Hz and 8 Hz

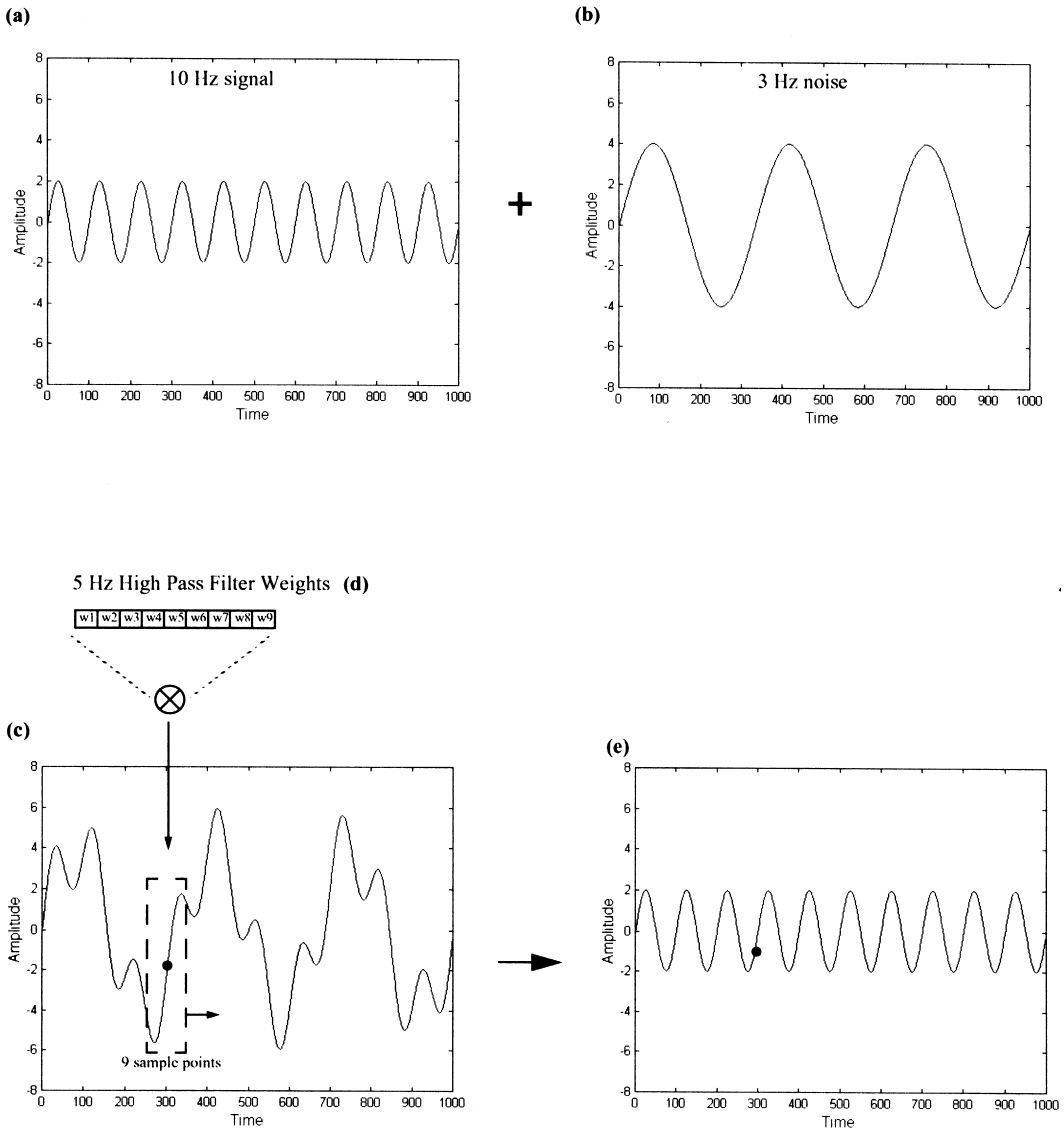

Fig. 3. The application of an ideal 5-Hz high-pass filter to remove a low frequency noise component. (a) A 10-Hz sine wave signal with an amplitude of 2 units; (b) a 3-Hz sine wave signal with an amplitude of 4 units; (c) the composite signal with a selected point at 300 ms having the 5-Hz high-pass filter weight vector (d) applied to it; (e) the filtered signal showing a 10-Hz wave form.

(or 0.83 octaves). Therefore the rolloff is $-37/0.83$, or -45 dB per octave. Similar considerations reveal rolloff for the lower transition band (1.1–1.8 Hz, or 0.71 octaves) is -52 dB per octave. Finally, Fig. 4a illustrates a small amount of pass band 'ripple'. That is, changes in degree of attenuation with frequency. Fig. 4b illustrates a small amount of stop band ripple (note that b is in logarithmic units so the stop band ripple has negligible effect in terms of gain change). Some filters may produce sizeable ripple when a steep roll off is used (e.g. over 60 dB per octave) and this can produce distortion in the filtered output waveform. It is therefore, advisable to examine a filter's characteristics as has been done here, to determine its possible limitations.

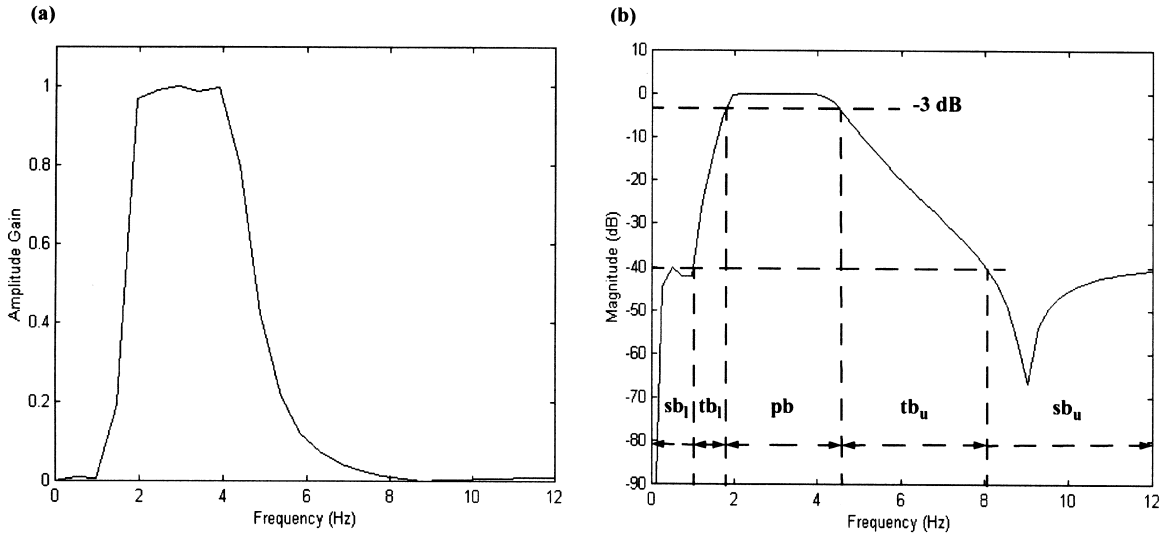

Fig. 4. Two representations of a band-pass filter. (a) The amplitude gain representation has intuitive appeal since it illustrates how much the signal will be altered by the filter weights. It also shows some band-pass ripple between 2 and 4 Hz; (b) the magnitude representation is better for identifying the various types of bands. Filter cutoffs are typically described in terms of the -3 dB signal attenuation, defining the pass-band (pb). The stop band (sb) here is defined at -40 dB. The stop-band has two ranges, lower (sb_l) and upper (sb_u). There is a transition band between the pass band and each of these stop bands, on the lower side (tb_l) and the upper side (tb_u).

Digital filters fall into two classes: finite impulse response (FIR) and infinite impulse response (IIR) and are the techniques implemented in the hardware of digital signal processing boards. To compute the element at time t for the filtered signal, an FIR operates on a finite segment of the digitised input signal points. The DP of this vector (a segment of the overall waveform) and the filter weights is computed. The convolution process then moves to the next time point and repeats the computation, thus producing a new point for each step in this process. Thus, the computation of a filtered signal (SF) from a raw signal (S) with a FIR weight vector (W) is given by:

$$\mathrm{SF}_t = \sum_{i=-N}^{+N} W_i S_{t-i} \tag{6}$$

where i ranges over a segment in time about t with $2N+1$ filter weights. Fig. 3c,d shows a convolution in process at time point $t = 300$ ms on a 1-s signal segment, sampled at 100 Hz. Nine points (four before t, t and four after t), are used in this simple filter. The FIR filter array shown in Fig. 3d is cross-multiplied on the signal interval $260 \leq t \leq 340$ to compute the filtered time point shown in Fig. 3e. The completed convolution in this example has eliminated the low frequency noise component, acting, therefore, as a high-pass filter. Such 'clean' filtering is ideal. In practice, some distortion may be introduced by filter ripple (see later). Filters that fall under the FIR category include moving average (also known as boxcar or smoothing filters), Hanning, Kaiser and polynomial.

The IIR filter retains an infinite memory of prior inputs by using them recursively to compute the signal at a given time t. Computation of a filtered signal, SF_t, with an IIR filter is thus given by:

$$\mathrm{SF}_t = \sum_{i=1}^{+N} W_i SF_{t-i} \tag{7}$$

Note the recursive nature of this filter type in terms of its prior outputs (SF_{t-i}).

Care must be exercised in using IIR and analog filters as both can introduce significant phase distortion (different frequency components are phase shifted by different amounts) into the signal. However, by appropriate filtering in both directions of the time series, phase distortion can be essentially eliminated. For example, the MATLAB™ signal processing toolbox includes such functions for applying IIR filters. FIR filters, on the other hand, typically have zero phase distortion. However, they typically require many more weights (and thus computation time) to achieve the same filter characteristics as an IIR filter.

Modern digital filters can generally match or outperform their analog counterparts and, because of their implementation in software, can be flexibly adjusted to tune their characteristics. However, built-in analog filters are still useful when only rudimentary filtering is necessary. For example, to avoid aliasing noise (Section 3.4.4) without undue oversampling, a signal should be low-pass filtered prior to digitisation. Signal drift can be effectively removed with analog high-pass filtering and thus maintain the signal in a range acceptable to analog amplifiers and the word size used to quantise the signal (see Section 3.4.4). The main precaution to exercise is not to analog filter more than is necessary, since information will be irretrievably lost.

While such narrow band filters as described in the example above have problems of signal distortion due to ripple and may not eliminate noise outside the pass band, nevertheless, Farwell et al. (1993) have shown that careful selection of a filter can help identify single trial ERP components, a challenging signal detection task. An algorithm was used to find an optimal FIR filter and this was found superior to boxcar filters (i.e. smoothing filters) for detection of a P300 component. Similar advantages in improving SNR may be found for other types of psychophysiological data, although this has been little explored. Litvack et al. (1995) compared spectral analysis and a frequency filtering method for analysis of ECG data and caution about possible distortion introduced by filtering.

5.1.2. Signal averaging

The signal of interest is often related to a stimulus event either under the experimenter's control, or to some marker reliably identifiable post hoc. Regularities in the signal are embedded in noise and thus may not be easily identified in single trials.

However, averaging the signals from several trials (aligned on the stimulus or marker event) can improve SNR in proportion to the number of trials being averaged. For example, the SNR for an average based on 80 trials will be 19 dB (Eq. (2)) better than that of the individual signals. This averaging process is commonly performed on ERP data in which a relatively small signal may be embedded in noise. For heart rate data acquired under several different conditions, it may be of interest to examine the shape of the ECG signature within each condition by aligning on a feature on each heart beat and averaging these waveform segments. Parameters of the averaged waveform may then be extracted, such as peak amplitudes and areas, or inter-peak intervals. These may be regarded as the 'signal' information that was systematic trial-to-trial. Comparison of methods for quantifying this information for ERP averages are summarised by Callaway et al. (1983). Donchin and Heffley (1976) compare these average measures with multivariate methods. Woldorff (1993) identifies some problematic issues for averaging methods when underlying components overlap. Pfurtscheller and Cooper (1975) describe a method for selectively averaging the individual trial data. Finally, an alternative to peak measures is the use of difference potentials with associated significance testing to help identify the discriminating signal (Gutherie and Buchwald, 1991; Blair and Karniski, 1993).

5.1.3. Template filtering

Signal averaging may fail to provide much improvement in SNR if the signal has variability which precludes alignment on the basis of a simple marker, such as the occurrence of a stimulus. In such a case, averaging temporally 'smears' the signal. Template filtering is a method which involves detecting (to some degree) the presence of the signal in each trial. Woody (1967) described

an algorithm to perform template matching on the signal to detect a pattern of interest. This method detects the signal using an iterative process which starts with the average signal (or some other estimate of the shape of the signal) and temporally shifts this past the signal while computing an index of match ρ (e.g. correlation or covariance, which are both based on DP). A ρ is thus calculated for each shift increment. If it is assumed that the waveform is present to some extent in each trial, then the maximum value of ρ can be used as evidence of its occurrence and the latency noted. Computing an average on the latency aligned waveforms now provides the signal to noise reduction benefit as described in Section 5.1.2. Wastell (1977) found that if the initial template is good then few iterations are needed to achieve optimal performance. More recently, this template algorithm has been shown to perform favourably when compared with other techniques (e.g. Gratton et al., 1989).

5.2. Statistical techniques[2]

5.2.1. Principle components analysis

Principle components analysis (PCA) is a general data reduction technique which identifies major 'components' in a data set (also known as the Karhunen–Loeve transform). For time series data which the psychophysiologist acquires, there is considerable correlation between nearby points. This correlation is also influenced by underlying events which give rise to the time series. For example, ERP time series data is related to underlying neural activity. This neural activity will, hopefully, be related to the manipulated conditions of the experimental paradigm. The experimental manipulations will be manifest in the ERP and cause intervals of this time series data to covary (according to the presence or absence of the events which drive the neural activation). PCA will identify these components amidst random noise influences, which by definition do not result in covariation between time series points over a set of trials. The derivation of principle component scores provides a reduction in the number of data points required to describe the signal and reduces the influence of random noise. The set of raw waveforms, provided as a two-dimensional matrix of time points (indexed by i) by waveforms (indexed by j), is used to build a matrix which represents covariation (e.g. a correlation or cross-products matrix). A relatively small number (e.g. $n = 3–8$) of orthogonal components are extracted by an iterative process which decomposes each raw voltage time point into a product of a principle component score (PCS) and a principle component loading. To compare this with the filter equations presented above, a new score can be computed as a DP of raw time series voltage (V) and a weight vector (c). Thus, the nth PCS for the jth waveform can be construed as:

$$\mathrm{PCS}_{jn} = \sum_{i=1}^{T} c_{in} V_{ij} \tag{8}$$

where each c_{in} is a score coefficient vector (yielded as part of a PCA, and is analogous to a template filter) and V_j is a raw voltage vector for trial j which consists of T time points. By applying each 'template' to the time series the relative presence of each component in a single ERP trial can be quantified. Unlike the frequency filters (e.g. Eqs. (6) and (7)) the template does not 'slide' along the raw waveform, but is only applied once to yield a single number. Since a relatively small number of components account for the variance of the data matrix, each waveform of, perhaps, several hundred points is reduced to a small number of PCSs. Each PCS indicates the strength or presence of a particular, orthogonal signal (component) in the voltage time series V_j. Analysis of variance can then be performed on this reduced set of scores to determine the effects of experimental treatments on these signals. Eq. (8) has the properties of a filter and the mathematical basis is similar. Rather than being derived from a priori considerations such as frequency range, a PCA filter is derived from the statistical properties of a set of waveforms. A tutorial overview

[2] There are many statistical packages available for implementing these techniques. The MATLAB™ system also includes a statistics toolbox and flexible programming and display environment.

applying the PCA approach to ERPs is given by Donchin and Heffley (1976), and an evaluation of PCA's strengths and weaknesses with ERPs is given in Gutherie (1990). One of the more serious problems with PCA is that the loadings are not unique and the varimax solution may not reflect theoretically relevant factors. Collet (1989) points out that the autoregressive structure of ERP data (and this applies generally to psychophysiological time series) should be taken into account and shows that a simpler model can be used to account for the same variance as a PCA. A comparison of traditional PCA to related techniques for ECG data is provided by Zarzoso et al. (1997).

Finally, traditional PCA uses a two-mode decomposition of raw scores where subjects are assumed to have similar distributions of component generators. Mocks (1988) provides an extension to PCA methods by introducing a means of representing subject variability with a three-mode model of the time course of a set of waveforms.

5.2.2. Discriminant analysis

Waveforms recorded in a psychophysiology experiment can generally be grouped according to different experimental conditions and thus may be regarded as belonging to different 'classes'. Although the individual waveforms may have poor SNR, if there is sufficient numbers of them in each class it may be possible to use discriminant analysis (DA) to derive discriminant functions each with a relatively small number of coefficients which represent the invariants in the waveform classes. With a suitable interpretation, these invariant features may be regarded as the signal to be extracted from the original noisy data. A good general introduction to DA is provided by Hair et al. (1995). An example of the application of DA to ECG data is Gomis et al. (1997); for ERPs, Horst and Donchin (1980); and for EEG, Lind et al. (1997). The importance of proper cross-validation has been emphasised by Daruna and Karrer (1981) and the potential for misuse by Lachin and Schachter (1974).

5.2.3. Regression techniques for noise reduction

When an endogenous noise source gives rise to a relatively stable albeit complex waveform, and an estimate can be made of the waveform shape (e.g. a template can be created by averaging), then regression techniques can be used to remove the contribution of this noise source to the primary recording channels. This problem has received considerable attention in ERP research due to the contamination of ERP channels by eye movement 'noise'. In the simplest case, a template of the eye-related noise is built. For example, using an EOG electrode placed above one eye, blinks can be detected and averaged. This can then be used to determine the presence of eye-related noise in individual trials. One method to achieve this is to regress the EOG template on each ERP channel to determine how much variance can be accounted for by the artefact. This information can then be used to derive propagation factors (Gratton et al., 1983) and these used to attenuate the presence of the artefact in the ERP channels (i.e. filter it out). More sophisticated techniques have been developed in recent years and a comparison is provided by Kenemans et al. (1991).

5.3. Other signal analysis techniques

In this last section brief mention is made of some relatively new techniques. These are wavelet analysis, neural network classification and fractal analysis.

5.3.1. Wavelet decomposition

Wavelet decomposition of a time series offers an alternative conceptualisation to the Fourier spectral representation. Wavelet analysis processes the signal at different scales or resolutions. In essence, the wavelet functions have a role analogous to sine and cosine functions in the Fourier series. However, unlike Fourier analysis, both temporal and frequency analysis of the signal is possible since wavelets are finite waveforms, whereas the sine wave in Fourier decomposition is infinite. Samar et al. (1995) provide a tutorial overview of the utility of wavelet analysis for ERP research. They conclude that the strengths of wavelets in this domain are the ability to locate in time the occurrence of an ERP component against background noise and discriminant analysis of ERP features. Jobert et al. (1994) discuss the merits of wavelet analysis for analysis of continu-

ous EEG data. Wavelets have been recognised as a useful tool for examining ECG signals following myocardial infarction (Reinhardt et al., 1996). It also has potential for preserving good SNR in data compression (e.g. Thakor et al., 1993), which is noteworthy given the voluminous data generated in psychophysiology research. It has been shown superior to traditional Fourier-based filtering methods for ECG data (Gramatikov and Georgiev, 1995). Wavelet analysis and recomposition can also be used to implement low-pass, high-pass, band-pass and band-stop filters.

5.3.2. Neural network analysis[3]

Artificial neural networks (ANNs) have been recognised for their ability to learn to classify noisy patterns. A segment (e.g. a single trial) of psychophysiological data may be construed as a pattern and a collection of these segments may form a few classes which are defined by the experimental conditions under which the data were acquired. Gupta et al. (1995) provide a tutorial overview of the use of an ANN in the classification of ERP waveforms using a feed-forward network. Anderer et al. (1994) used a similar ANN with EEG data and compared it to DA, finding the ANN superior. A feed-forward ANN computes the DP of the network input array with a weight array at each neuron, and thus can be considered to be performing a function similar to a collection of template filters. Other properties of neural network computations permit them to perform non-linear functions (e.g. non-linear discriminant analysis) and this may provide them with greater signal processing capability than the linear techniques discussed in the previous sections (e.g. standard DA). In the classification task, the network learns connection weights which will result in a maximal response at an output to one class of inputs, while attenuating this response to all other inputs. Examining the weight profiles at each neuron may provide important information regarding the signal features to which the network is sensitive once it has been trained. A theoretical analysis of the feed-forward ANN trained with the backpropagation algorithm is given in Ruck et al. (1990).

5.3.3. Non-linear dynamics

Finally, Pritchard and Duke (1995) apply concepts from non-linear dynamics to describe EEG data. They propose that if the underlying source of the EEG is generated (at least partially) by feedback loops, then this will give rise to periodic and aperiodic behaviour manifest in the EEG signal which is indicative of self-organisation. This self-organisation occurs in the impulse behaviours of the interconnected neurons. The 'fractal dimensionality' of the EEG signal, a measure which is a product of dynamical analysis, is indicative of the nature of this connectivity and thus a clue to the neural architecture that gave rise to the EEG signal.

6. Summary and troubleshooting

A variety of hardware and software techniques have been described which can reduce the presence of noise in psychophysiological recordings. This final section summarizes some general guidelines when considering which of the above methods may be most suitable in a given situation. The issue here is not which type of recording (ERP, EMG, etc.) should be matched with a noise reduction technique since application of noise reduction or signal enhancement techniques is driven by the functional characteristics of the noise and signal to be discriminated. Thus, three general guidelines are provided for deciding which filtering, statistical or innovative technique to consider. This is followed by a troubleshooting table which lists most of the commonly occurring symptoms of noise and suggested solutions.

The three general guidelines for selection of an appropriate active noise reduction technique are:

1. There exists an identifiable frequency-stable noise source (e.g. as revealed by spectral analysis (Section 2.4) or the signal is known to occur within a particular frequency band. In these cases digital filtering (low-pass, high-pass, band-stop and band-pass) are effective using either the conventional fourier-based

[3] Neural networks and wavelet toolboxes for MATLAB™ offer flexible approaches.

Table 2
Observed noisy signals and possible solutions

	Waveform observation	Possible causes	Possible solutions
1	Repetitive peaks at 50 or 60 Hz in several channels (FFT analysis helps to identify frequency)	Line noise, monitor noise, light sources (Section 3.2)	Check/install shielding (Section 4.2) Analog notch filter on-line (Section 5.1.1) Digital band-stop filter off-line (Section 5.1.1)
2	Repetitive peaks at approximately 10 Hz in EEG, particularly in occipital regions	Alpha rhythms (Section 3.3)	Increase subject alertness Digitally filter off-line (Section 5.1.1)
3	Most channels are fine, some have broadband or 50 or 60 Hz noise	Poor electrode impedance (Section 3.4.2) Aliasing (Section 3.4.4)	Check impedance, reduce to 5 kΩ Low-pass or notch analog filter prior to ATDC (Section 5.1.1)
4	Loss of signal (flat line) in one or more channels	Connection lost Amplification too high (clipping) (Section 3.4.1) Severe DC drift (Sections 3.2.1 and 3.4.3)	Check electrodes (Sections 3.4.2 and 3.4.3) Reduce amplifier gain (Section 3.4.4) Check/reduce electrode impedance, high pass filter (Section 5.1.1)
5	Waveform has poor resolution of detail	Amplification too low (Section 3.4.1) ATDC word size too small (8 bit) (Section 3.4.4)	Increase amplifier gain (Section 3.4.4) Use a larger word size (e.g. 16 bit) (Section 3.4.4)
6	High amplitude 'spikes'	In EEG, eye blinks (Section 3.3) General: electric switches in nearby equipment (Section 3.2)	Amend subject instruction, attenuate using regression technique (Section 5.2.3) Shielding or eliminate at source, ensure common grounding (Sections 4.1 and 4.2)
7	Bursts of high frequency (30 Hz +)	EMG due to subject movement (Section 3.3)	Instruct/practice subject to reduce
8	High amplitude 'rolling' waves, EEG recording	Eye movements (Section 3.3)	Amend task/instructions Attenuate using regression technique (Section 5.2.3)
9	Broadband noise several channels	Induced currents from external sources (Section 3.2)	Ensure electrodes are similar (Section 3.4.3) Check impedances (Section 3.4.2)
10	Poor signal to noise ratio, above sources not the problem	Endogenous sources (Section 3.3)	Digital filter off-line (Section 5.1.1), signal averaging (Section 5.1.2), PCA (Section 5.2), other off-line signal detection methods (Section 5.3)
11	Intermittent problems	The majority of all electrical problems are in the connectors	Impedance testing of cables and connectors Unplug and reconnect connectors Reseat data acquisition cards in computers
12	None of the above, failure to get any recognisable signal	Improper equipment configuration	Disassemble and rebuild complete Psychophysiology recording apparatus

methods (Section 5.1.1) or the newer wavelet approach (Section 5.3.1).

2. One has used a repetitive stimulus driven paradigm where the signal has a known, perhaps complex, structure and is masked by background noise. Simple averaging (Section 5.1.2) time locked to the stimulus is effective if the signal is latency stable, otherwise a template filter (Section 5.1.3) can be used prior to averaging.
3. The signal consists of multiple unknown components and it is desired to identify the components or classify groups. Pattern recognition techniques, either linear such as PCA (Section 5.2.1) or non-linear such as with certain neural networks (Section 5.3.2) can be used provided cross-validation of the solutions is applied to ensure they are not spurious. DA (Section 5.2.2) and neural networks can be applied where it is desired to discriminate groups on the basis of the signals.

Finally, Table 2 presents a summary checklist of observed problems and suggested solutions. While this table is not exhaustive in scope or possible solutions, it is intended to provide a basis for checking some of the frequently encountered problems that can arise in data acquisition and signal analysis.

Acknowledgements

Preparation of this paper was supported by the Australian Research Council grant A79601859.

References

Anderer, P., Saletu, B., Kloppel, B., Semlitsch, H.V., Werner, H., 1994. Discrimination between demented patients and normals based on topographic EEG slow wave activity: Comparison between z statistics, discriminant analysis and artificial neural network classifiers. Electroencephalogr. Clin. Neurophysiol. 91, 108–117.

Andreassi, J.L., 1995. Psychophysiology: human behavior and physiological response 3rd edn,. L. Erlbaum, Hillsdale NJ.

Blair, R.C., Karniski, W., 1993. An alternative method for significance testing of waveform difference potentials. Psychophysiology, 30, 518–524.

Brophy, J.J., 1990. Basic electronics for scientists 5th edn., McGraw-Hill, New York.

Cacioppo, J.T., Tassinary, L.G., 1990. Principles of Psychophysiology: Physical, social and inferential elements, Cambridge Univ. Press, Cambridge.

Callaway, E., Halliday, R., Herning, R.I., 1983. A comparison of methods for measuring event-related potentials. Electroencephalogr. Clin. Neurophysiol. 55, 227–232.

Collet, W., 1989. Doubts on the adequacy of the principal component varimax analysis for the identification of event-related brain potential components: a commentary on Glaser and Ruchkin, and Donchin and Heffley. Biol. Psychol. 28, 163–172.

Cook, E.W., Miller, G.A., 1992. Digitial filtering: background and tutorial for psychophysiologists. Psychophysiology, 29, 350–367.

Cunningham, E.P., 1992. Digital filtering: An introduction, Houghton Mifflin, Boston.

Daruna, J.H., Karrer, R., 1981. On the validation of discriminant functions: An empirical analysis using event-related potentials. Psychophysiology, 18, 82–87.

Diefenderfer, A.J., Holton, B.E., 1994. Principles of electronic instrumentation, Harcourt Brace, Sydney.

Donchin, E., Heffley, E.F., 1976. Multivariate analysis of event-related potential data: A tutorial review. In: Otto, D. (Ed.), Multidisciplinary perspectives in event-related brain potential research, US Government Printing Office, Washington DC, pp. 555–572.

Dumermuth, G., Molinari, L., 1987. Spectral analysis of EEG background activity. In: Gevins, A.S., Rémond, A. (Eds.), Methods of analysis of brain electrical and magnetic signals. EEG handbook (revised series, Vol. 1), Elsevier, Amsterdam.

Farwell, L.A., Martinerie, J.M., Bashore, T.R., Rapp, P.E., Goddard, P.H., 1993. Optimal digital filters for long-latency components of the event-related brain potential. Psychophysiology, 30, 306–315.

Gevins, A.S., 1987. Overview of computer analysis. In: Gevins, A.S., Rémond, A. (Eds.), Handbook of Electroencephalography and Clinical Neurophysiology, Elsevier, Amsterdam, pp. 31–83.

Gomis, P., Jones, D.L., Caminal, P., Berbari, E.J., Lander, P., 1997. Analysis of abnormal signals within the QRS complex of the high resolution electrocardiogram. IEEE Trans. Biomed. Eng., 44, 681–693.

Gramatikov, B., Georgiev, I., 1995. Wavelets as alternative to short time Fourier transform in signal averaged electrocardiography, 482–487.

Gratton, G., Coles, M.G.H., Donchin, E., 1983. A new method for offline removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55, 468–484.

Gratton, G., Kramer, A.F., Coles, M.G.H., Donchin, E., 1989. Simulation studies of latency measures of components of the event-related brain potential. Psychophysiology, 26, 233–248.

Gupta, L., Molfese, D.L., Tammana, R., 1995. An artificial neural network approach to ERP classification. Brain Cognit. 27, 311–330.

Gutherie, D., 1990. Intergroup and intrasubject principle com-

ponents analysis of event-related potentials. Psychophysiology, 27, 111–119.

Gutherie, D., Buchwald, J.S., 1991. Significance testing of difference potentials. Psychophysiology, 28, 240–244.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1995. Multivariate data analysis, 4th edn., Prentice-Hall, Englewood Cliffs, NJ.

Horst, R.L., Donchin, E., 1980. Beyond averaging. II. Single-trial classification of exogenous event-related potentials using stepwise discriminant analysis. Electroencephalogr. Clin. Neurophysiol. 48, 113–126.

Hugdahl, K., 1995. Psychophysiology: the mind-body perspective, Harvard Uni. Press, Cambridge, MA.

Jennings, J.R., Coles, M.G.H., 1991. Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches. John Wiley & Sons, Brisbane.

Jobert, M., Tismer, C., Poiseau, E., Schultz, H., 1994. Wavelets: A new tool in sleep biosignal analysis. J. Sleep Res. 3, 223–232.

Karl, J.H., 1989. An introduction to digital signal processing. Academic Press, San Diego.

Kenemans, J.L., Molenaar, P.C.M., Verbaten, M.N., Slangen, J.L., 1991. Removal of the ocular artifact from the EEG: A comparison of time and frequency domain methods with simulated and real data. Psychophysiology, 28, 114–121.

Lachin, J.M., Schachter, J., 1974. On stepwise discriminant analyses applied to physiologic data. Psychophysiology, 11, 703–709.

Lind, J.C., Koles, Z.J., Flor-Henry, P., Soong, A.C.K., 1997. Frequency domain discriminant analysis of the electroencephalogram. Psychophysiology, 34, 358–364.

Litvack, D.A., Oberlander, T.F., Carney, L.H., Saul, J.P., 1995. Time and frequency domain methods for heart rate variability analysis: a methodological comparison. Psychophysiology, 32, 492–504.

Lutzenberger, W., Elbert, T., 1991. A/D conversion in DMA-Mode: Problems and solutions. Psychophysiology, 28, 607–608.

Mocks, J., 1988. Decomposing event-related potentials: A new topographic components model. Biol. Psychol. 26, 199–215.

Nasralla, H.A., Weinberger, D.R., 1986. The neurology of schizophrenia. Elsevier, Amsterdam.

Nunez, P.L., 1981. Electric fields in the brain: the neurophysics of EEG. Oxford Univ. Press, New York.

Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M., Silberstein, R.B., Cadusch, P.J., 1997. EEG coherencey 1. Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr. Clin. Neurophysiol. 103, 499–515.

Pfurtscheller, G., Cooper, R., 1975. Selective averaging of the intracerebral click evoked responses in man: An improved method of measuring latencies and amplitudes. Electroencephalogr. Clin. Neurophysiol. 38, 187–190.

Press, W.H., 1986. Numerical recipes: The art of scientific computing. Cambridge Univ. Press, Cambridge.

Pritchard, W.S., Duke, D.W., 1995. Measuring chaos in the brain: A tutorial review of EEG dimension estimation. Brain Cognit. 27, 353–397.

Reinhardt, L., Makijarvi, M., Fetsch, T., Montonen, J., Sierra, G., Martinezrubio, A. Katila, T., Borggrefe, M., Breithardt, G., 1996. Predictive value of wavelet correlation functions of signal averaged electrocardiogram in patients after anterior versus inferior myocardial infarction. J. Am. Coll. Cardiog. 27, 53–59.

Ruck, D.W., Rogers, S.K., Kabrinsky, M., Oxley, M.E., Suter, B.W., 1990. The multilayer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Trans Neural. 1, 296–298.

Samar, V.J., Swartz, K.P., Raghuveer, M.R., 1995. Multiresolution analysis of event-related potentials by wavelet decomposition. Brain Cognit. 27, 398–438.

Scerbo, A.S., Freedman, L.W., Raine, A., Dawson, M.E., Venables, P.H., 1992. A major effect of recording site on measurement of electrodermal activity. Psychophysiology, 29, 241–246.

Stern, R.M., Ray, W.J., Davis, C.M., 1980. Psychophysiological recording, Oxford University Press, Oxford.

Surwillo, W.W., 1990. Psychophysiology for clinical psychologists. Ablex Publ., Norwood, NJ.

Tassinary, L.G., Geen, T.R., Cacioppo, J.T., Edelberg, R., 1990. Issues in biometrics: Offset potentials and the electrical stability of Ag/AgCl electrodes. Psychophysiology, 27, 236–242.

Thakor, N.V., Sun, Y.C., Rix, H., Caminal, P., 1993. Multiwave: A wavelet based ECG data compression algorithm. IEICE Trans. Info. Syst. 12, 1462–1469.

Thickbroom, G.W., Mastaglia, F.L, Carroll, W.M., Davies, H.D., 1984. Source derivation: application to topographic mapping of visual evoked potentials. Electroencephalogr. Clin. Neurophysiol. 59, 279–285.

Wastell, D.G., 1977. Statistical detection of individual evoked responses: An evaluation of Woody's adaptive filter. Electroencephalgor. Clin. Neurophysiol. 42, 835–839.

Webster, J.G., 1995. Medical instrumentation: application and design, 2nd edn., Houghton Mifflin, Boston.

Welch, P.D., 1967. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Transactions on Audiology and Electroacoustics, 15, 70–73.

Woldorff, M.G., 1993. Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. Psychophysiol. 30, 98–119.

Woody, C.D., 1967. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. Biol. Eng. 5, 539–553.

Zarzoso, V., Nandi, A.K., Bacharakis, E., 1997. Maternal and Foetal ECG separation using blind source separation methods. IMA J. Math. Appl. Med. Biol. 3, 207—225.