ERP data collection and processing

Overview: scalp potentials to plots

The requirements for ERP experiments vary somewhat with the demands of the different types of experiment and stimuli presented. The following is a general overview, and will be elaborated further in subsequent sections.

To conduct a cognitive ERP experiment, one must be able to present stimuli of the desired types while simultaneously acquiring EEG data several sites on the scalp. In many experimental settings, at least some behavioral data such as response categories or latencies will be acquired from the subject as well. To permit the subsequent calculation of brain activity related to the occurrence of a specific stimulus event, the time course of the stimulus presentation and EEG data acquisition must be synchronized to an accuracy on the order of a millisecond and registered with the stimulus sequence so that the EEG data can be sorted according to the stimulus or stimulus type it is temporally registered to.

Following acquisition, the data for each subject must screened for artifacts, either by an automated procedure or manually. When scalp potentials are of interest, bioelectric artifacts include potentials generated by eye-movements, blinking, heartbeat, and muscles of the head and neck. Technical artifacts include the saturation or "blocking" that occurs when the dynamic range of the amplifier is exceeded, intermittent signals caused by loose scalp-electrode connections, and potentials caused by interfering electrical sources in the environment, like nearby electrical equipment (see Figure 1 for examples). The damage done by some types of artifacts, such as ocular artifacts and high frequency electrical interference, can be mitigated after the fact. However, other types of artifacts, such as large myoelectric potentials and amplifier blocking, necessitate excluding the data from subsequent analysis. The remaining data may be further processed in a variety of ways – for example by digital filtering, rereferencing the electrode derivation, normalizing, or computing a Laplacian transform. The choices here are dictated in part by the character of the data and in part by type of signal that is to be extracted from the continuous EEG. When the data has been appropriately processed, the desired experimental measure is calculated: e.g., time domain average potentials or frequency band power, or estimates of neural electrical sources. These data for each subject are then typically reduced by measuring a selected feature, such as peak or mean amplitude or r.m.s. power, within a specified post-stimulus interval. The reduced data for each subject are subjected to a statistical analysis and tabulated. The EEG data are visualized at each stage of this process. Behavioral data acquired in the course of the experiment

are treated in a similar fashion – sanitized, reduced, analyzed statistically, and visualized. This handling of the behavioral data is similar to non-EEG experiments with the proviso that the behavioral and EEG data should be coordinated in subsequent analyses.

Data Acquisition

The general techniques for recording EEG data for an ERP experiment are much the same as for clinical EEG recordings.

Electrodes: material, positions, placement.

Many different types of electrodes are available for recording from the scalp, including surface electrodes and subdural needle electrodes. Each has advantages and disadvantages (Regan 1989) but subdural electrodes are rarely used for experimental research. Surface electrodes are typically 4mm to 10mm in diameter and may be shaped as disks or cups. The electrodes are attached to wires that are connected to the input of the electroencephalograph on one end and the subject's head on the other (via collodion, paste, or a gauze/paste combination for individual electrodes or embedded in an elastic cap or net for a set of electrodes).

Ideally, electrodes would be placed over precisely the same areas of cortex for each subject in an experiment, but differences between individual brains make this difficult. Solutions involving brain imaging are impractical in the majority of cases, and, therefore, as an approximation, placement of electrodes is calculated with reference to the location of bony landmarks of the skull, with interelectrode distances scaled to take into account differences in head size. The International 10-20 system (Jaspers 1958; see Figure 2), and its variants, are in wide use. Measurements are made from the nasion (top of nose) to the inion (bony protrusion on back of head) across the top of the head, and then between the preauricular depressions (just in front of the ears) through the midpoint of the nasion-inion line. The nasion-inion distance and the interaural distance across the surface of the scalp provide a frame of reference, and subsequent head measurements and electrode positions are calculated as percentages of these distances along circles, with the top middle of the head as the center of the circle, and are marked on the scalp with a grease pencil. This procedure permits electrodes to be placed at similar relative positions on different sized heads. Locations are named to indicate position on the head; each site is a combination of a letter indicating a general region (frontal or F, central or C, parietal or P, occipital or O, temporal or T) followed by the letter z for locations along the midline (Cz - central midline, also known as the vertex), or a number, where odd numbers refer to the left side of the head (e.g. C3 - left central) and even numbers refer to the right side of the head (e.g. C4 - right central).

Electrodes may be positioned and affixed to the scalp individually but when more than a few electrodes are involved, measuring the locations individually is error prone and time consuming. A common alternative is to mount multiple electrodes on an elasticized cap or other flexible device. In this case, the head is again measured and a small number of reference electrode positions measured – e.g., above the orbits. These reference positions constrain the orientation of the cap, and the remaining electrodes are placed by isotonically stretching the cap over the head. This technique preserves consistent relative electrode positions over a reasonable range of head sizes, while eliminating the need to calculate each electrode position individually. Electrodes at locations that are not covered by the cap – for example, those around the eyes – are placed individually (usually with a sticky adhesive collar).

With the electrode locations determined, the scalp is prepared by parting the hair, cleaning if possible, and then lightly abrading the area that will be under the electrode. Abrasion displaces dead skin cells and ruptures the membranes of the top layer of skin cells. The abrasion process may result in redness of the skin, but when performed properly does not. Alternatively the scalp may be punctured with a sterile needle, although this will leave small marks on the scalp. The aim of cleaning and abrading is to reduce the electrical resistance of the electrode-scalp junction since the ability of the differential amplifier circuits in the typical electroencephalograph to reject common mode noise, e.g., ambient electrical noise from electrical mains, varies inversely with the electrical impedance of the circuit and since large differences in the impedances can result in standing potentials between electrodes. By scratching (breaking the skin), one also reduces the contribution of changes in electrodermal activity (EDA; also known as the galvanic skin response or GSR), which reflects changes in skin resistance. EDA or GSR responses can be due to sweating if the person gets hot or nervous and may on occasion be time-locked to the event of interest as when a person is surprised or has some emotional response to the stimulus.

Individual electrodes may be secured by commercially available adhesive rings or hospital grade paper tape. For sites located above the hairline, individual electrodes can be temporarily secured with a tacky conducting electrode paste and cotton ball, but such connections are mechanically weak and best suited to recording sessions that involve virtually no movements of the head. Collodion may be used for a more secure placement, but both collodion and the solvent used to remove it should only be used with adequate ventilation because a spark from electrical devices such as light switch could ignite the vapors and cause an explosion. For electrodes mounted in a cap or similar elastic device, the scalp is abraded and the electrode-scalp junction formed by applying a small amount of electrolytic gel between the scalp and the electrode.

Whether gel or paste is used, the scalp-electrolyte-electrode junction is not an electrically inert connector, but has an intrinsic resistance (R) and capacitance (C). The RC properties of the junction act as a high-pass filter that attenuates lower frequency fluctuations in the scalp potentials. The specific frequency response is determined by the electrical properties of the materials used, which, together with other factors about the experimental measure of interest, constrains the choice of electrode materials. Surface electrodes are commonly made of gold (Au), silver (Ag), chlorided silver (Ag-AgCl) and tin (K). Because of their resistance to artifacts, ease of use, and economy, tin electrodes are widely used for ERP research in which the potentials of interest are above 2 Hz; this includes most of the cognitive potentials that we will be discussing, such as the P3, N400, P1, N1, and MMN. However, the time constant of the electrode-gel-scalp junction for tin is sufficiently low that for recording DC shifts or potentials that evolve over the course of several seconds or more, chlorided-silver electrodes are required. So if the brain response were a square wave of a few seconds duration and we wanted to record the square wave without distortion, we would use AgCl and not tin electrodes. Bias and polarization potentials can result in standing potentials between the electrodes and between the electrodes and the scalp if dissimilar metals, e.g., electrodes of different types, are used or if a direct current is passed through the junction as would occur if electrode-scalp impedances were measured with a DC ohm meter. In other words, at any given time all the electrodes on a subject's head should be of the same material, and if for some reason that is not possible, at minimum any two electrodes that are referred to one another **must** be made of the same material.

In addition to the placement of the electrodes, the type of montage must be determined, and this includes selecting an appropriate reference electrode or electrodes (see Nunez 1981 for discussion). In principle, electrical potentials are measured against an electrically neutral reference. Montages are often described as monopolar or bipolar. In a bipolar montage, electrodes are grouped in non-overlapping pairs and potentials recorded between the members of each pair. So, for a bipolar montage for recording eight EEG waveforms, sixteen electrodes grouped into eight pairs would be used. In the case of bipolar recordings, it is assumed that both locations are active, but the interest is simply recording the difference in activity between the two locations. If the activity at the two locations were identical, whether it was large or small, the difference or bipolar recording would be a flat line. On the other hand, if the potential at two sites were of opposite polarity, the difference or bipolar recording would be either a negativity or a positivity (larger than either of the potentials at each site), depending on the direction of the subtraction. Generally, recordings from opposite sides of a dipole/generator will be of opposite polarities. So, if one knew that location of a generator which they wished to manipulate, one could choose to accentuate the

ERP by choosing a bipolar recording from a pair of electrodes on either side of it. A good example includes bipolar recordings for blinks, with one electrode above the eye referred to an electrode below the eye (where the eyeball is the known source of the electrooculogram).

In the monopolar montage that is far more common in cognitive ERP research, electrodes are again grouped in pairs, but in this case, each pair has one member in common with all the others, the "common reference" electrode. So, for a monopolar montage for recording eight EEG waveforms, nine electrodes would be used: one for each of eight scalp sites of interest and a ninth for the common reference. Although the terms "monopolar" and "bipolar" are widely used, all montages are strictly speaking bipolar, since even putatively monopolar montages involve recording the potential between two electrodes, both of which are active. The pressure to treat one such type of montage as monopolar comes from the assumption that the reference electrode is electrically "inactive". If this were so, the EEG waveform would reflect solely the contribution of the "active" electrode. However, inactive references are idealizations that are not realized in practice, and much discussion has centered around the justification for selecting the best practical alternatives. Candidate locations for cephalic reference electrodes include the nose tip and chin (as they are both on the midline and therefore not asymmetric), an earlobe, and the scalp over the left or right mastoid process. However, the location of these on the head allows for the possibility of volume conductance of the same sort that is responsible for the scalp potentials of interest. The electrical sinks and sources in the brain will generate an electrical field defined, in principle, with respect to a point at an infinite distance. In practice, this in-principle field will typically be non-zero for most sites on the scalp at most times (and the zero lines will change rapidly over time). Because, in practice, different candidate reference sites are not electrically neutral, using a common cephalic reference at different points would generally result in the measurement of different potentials – that is, these potentials would be reference dependent.

Non-cephalic reference sites such as the arm or sternum are possible, but these are liable to artifacts such as heart beat. Furthermore, Nunez (1981) argues that for the electrical sinks and sources in the brain that are of interest in EEG recording, net current flow below the neck is nil for practical purposes. Therefore, with respect to these electrical sources in the head, there will be no difference in potentials recorded with respect to a non-cephalic reference electrode on the arm and a reference on the upper neck. In other words, there is no place on the body that can be used as an inactive reference and the difference between a reference on the neck and one on the big toe is not worth the effort of removing the subject's shoes.

In general the choice of a single electrode or linked pairs of electrodes results in EEG data whose properties (for example, the distribution of potentials across the scalp) depend on the

choice of reference position. However, reference free EEG data – that is, data that are insensitive to the choice of a particular reference – can be calculated. For instance, one common approach is to use the average of all the sites on the head as the reference. This works best when there are many electrodes; indeed it would be great if we had the entire head covered, not just on top but from all angles including from below (chin, neck, etc.). Multi-electrode references of all sorts can be obtained by combining individual electrodes. For instance, one can electrically link the mastoids or link electrodes on the sternum and vertabrae near the base of the neck. Linking in this way creates an electrical shunt that can distort the distribution of scalp potentials (because it forces the activity at the mastoids to be identical to each other even if they are not) but this can be avoided by adding parallel resistors sufficiently high to ensure that the current flow between the linked electrodes is negligible. Linked mastoids are a useful conservative choice for a reference when it is likely that the potential of interest is asymmetric. If a potential appears laterally asymmetric even with linked mastoids, it probably is! The use of linked mastoids as a reference for cognitive ERPs was very popular until recently. Now more and more researchers are using soft linking of mastoids (offline averaging of two mastoids), the average reference, or a noncephalic reference.

Recording scalp potentials

Currently, the most common systems for cognitive ERP research record continuous EEG and process it off-line. In such systems, scalp potentials are amplified by the electroencephalograph, the analog output of the electroencephalograph is converted to a digital representation by an analog-to-digital converter, and the resulting data are stored on a computer hard disk for subsequent processing.

Differential Amplification

The issue of choosing an electrical reference electrode or electrodes (as previously discussed) should not be confused with what is known as differential amplification. To measure a time-varying electrical potential, two leads of a voltmeter can placed across a source of electrical current. This is the case whether one is checking the voltage across the terminals of a 110V wall socket or the scalp. However, simply applying the leads of a sufficiently sensitive voltmeter to two electrodes affixed to the scalp – say the vertex and a mastoid reference – would not reveal what we recognize as characteristic background EEG because of high amplitude interference from ambient electro-magnetic fields. However, scalp electrodes are in fairly close proximity, so this interference tends to be fairly constant across all recording sites. This means that if one were to select a **third** scalp location, the "ground", and measure the voltage between the mastoid and the

ground as well as the vertex and the ground, the interference would make about the same contribution to both measurements at any given point in time. Then, if the mastoid-ground potential were subtracted from the vertex-ground potential, the effect of the interference that is common to both would tend to cancel out, leaving the difference between mastoid-ground potential and the vertex-ground potential. Differential amplifiers of the sort widely used in EEG research to amplify biopotentials in EEG research use this technique to reject common mode interference and selectively amplify just the difference between the potentials at the two leads of interest.

Analog to digital conversion

Using the same sort of digital technology that allows the analog acoustic signal produced by an entire symphony to be represented as a sequence of numbers on a CD, the continuous analog output of the electroencephalograph can also be digitized. The basic technique is to "sample" the analog signal – i.e., to measure, within some finite numerical precision, at periodic intervals. In other words, sampling is the process of taking a continuous time signal and representing it by a series of discrete measurements. Both the rate at which the signal is sampled and the precision of the numeric representation are important.

Choosing an appropriate sampling rate is a tradeoff. Higher sample rates allow higher frequencies present in the signal to be accurately represented. However, since each sample is a data point, higher sampling rates generate larger sets of digitized data, which require more computer storage and take longer to process during the data analysis phase of the experiment. In addition, beyond a certain point, higher sampling rates do not lead to any better representation of the signal. This exception to "the more the merrier" may be surprising, since the values of the signal that occur between samples are simply ignored. The initial intuition that something important might be happening between the samples is generally correct, but if the signal varies sufficiently smoothly from sample to sample, measuring the intermediate values adds no additional information. The minimum degree of "smoothness" required is given by the "Nyquist frequency", which is one half of the sampling rate. In other words, according to the Nyquist Sampling Theorem, the sampling rate must be greater than twice the maximum frequency present in the input signal. This critical sampling rate is known as the Nyquist rate. For instance if data are sampled at 250Hz, a signal that is band-limited to frequencies below 125Hz is, in principle, sufficiently smooth. In practice, however, experimenters typically choose to oversample data at 2.5 to 5 times the Nyquist frequency, so that a sampling rate of 250Hz is considered more appropriate for a signal bandlimited to 100Hz. If the signal is not sufficiently smooth between samples – that is, not appropriately bandlimited – then aliasing artifacts can arise from sampling the peaks and troughs

of the higher frequencies. Such aliasing will result in the spurious appearance in the signal of energy at low frequencies that is not actually present in the original signal (Figure 3).

The second factor is the degree of numerical precision used to represent the continuously varying analog amplifier output. Analog-to-digital (A-D) converters typically use a fixed number of binary bits to represent the numerical value. Each bit can take on the value 0 or 1 and can be used to represent a power of 2 in a binary representation. The more bits that are available, the more power of 2's and, hence, the greater the range of values that can be used to represent the analog voltages. One bit permits the representation of 2^1 = 2 distinct values, two bits permits the representation of 2^2 = 4 distinct values, four bits permits the representation of 2^4 = 16 distinct values, and in general, n bits permits 2^n distinct values to be represented. The A-D converter divides up the range of input voltages into however many distinct values it has available and maps non-overlapping ranges of voltages into these numeric values. So for instance, if the input voltage can vary between 0 and +2 Volts and the A-D converter uses four bits to represent these voltages, it might treat voltages in the range 0-0.5 as 0, 0.5-1 as 1, 1-1.5 as 1.5, and 1.5-2 as 2. With this degree of numeric precision, the A-D converter can discriminate a 1V signal from a 1.5V signal, but not a 1V signal from a 1.25V signal, both of which it treats as 1V. The error introduced by treating 1V and 1.25 signals as the same is the "quantization error". Clearly, having more bits available for the numeric representation of the signal permits the input range to be divided into smaller ranges, and this allows smaller differences in the input voltages to be distinguished. Eight bits permits allows 2^8 = 256 distinct values and, in the past, many computer-based A-D converters used 8-bit representations. However, 12 bits with 4096 distinct values are now more common and in some instances 14 or 16 bit A-D converters are used.

The Electroencephalograph

The electroencephalograph (EEG) is an electronic amplifier optimized for bioelectric scalp potentials. The electroencephalograph does two main jobs, both of which are aimed at preparing the scalp potentials for analog-to-digital conversion. The first job is to amplify the tiny scalp potentials from their naturally occurring levels of about +/- 50uV to a level that matches the input voltage range of the analog-to-digital converter. The second is to provide some initial signal conditioning of the scalp potentials, and importantly, to attenuate high frequencies in the amplifier output signal to prevent the aliasing artifacts discussed above.

With these two functions – amplification and frequency filtering – in mind, the electroencephalograph can be instructively compared with a home stereo system (thanks to Tyler Lorig for this example). A simple home stereo has two channels (left and right), a knob that

controls the volume, and perhaps two tone knobs, one to adjust the treble and one to adjust the bass. The stereo takes as input the relatively small audio signals that are output by the CD player or radio receiver and amplifies them to a level that is sufficient to drive the speakers. The volume and tone knobs permit adjustment of the signal characteristics before it gets sent on to the speakers. Specifically, the volume adjusts the amplitude of the output signal, the treble knob selectively adjusts the high frequencies, and the bass knob selectively adjusts the low frequencies. The basic functions of the stereo and the electroencephalograph are the same: to amplify multiple channels of very small signals into a usable range and permit some modification of the frequency power spectrum of the output. There are also some differences, of course.

First, whereas the home stereo has just two channels (left and right), the EEG usually has more. There is one for each position on the scalp that data will be acquired from. The number of channels varies from three or four for simple recordings of midline scalp sites and work with children, to 128 for high-density arrays of electrodes. In an optimal world, 256 channels would suffice, given the organization of the human neocortex. In practice, most laboratories in the nineties recorded between 16-32 electrodes. Second, whereas a home stereo amplifies an audio signal enough to serve as input to the speakers, the EEG amplifies the scalp potential to serve as input to a data storage device, typically an analog-to-digital converter that is connected to a computer disk drive or tape. The gain adjustment on the EEG corresponds to the volume knob on the stereo and adjusts the amplitude of the output signal. Third, whereas the treble and bass knobs on a home stereo give a kind of general control over the higher and lower frequencies, the corresponding low pass and high pass filter settings serve to attenuate all frequencies below and above a specified range, permitting only selected frequencies to pass through.

Electroencephalograph settings

The amplifier "gain" is the factor by which the input signal is multiplied. For EEG signals on the order of 50uV, gain factors of 5K or 10K are common. The optimum gain depends on the amplitude of the input scalp potentials and the output voltage range of the amplifier, which is determined by the manufacturer and is often in the 2.5 to 10V range. If the gain is too low, then some of the dynamic range of the amplifier will be wasted. If the gain is too high, then larger input signals will be amplified to a value that exceeds the output voltage range of the amplifier. In this case, the result is amplifier blocking, which looks like a flat line at the maximum or minimum output voltage value. During blocking, information about the input signal is irretrievably lost. The optimum gain is a compromise for which the smallest input signals are adequately amplified and the largest signals do not cause blocking. Blocking is easy to determine and avoid. What counts as "adequate

amplification" depends on the signal of interest and resolution of the analog-to-digital converter. Gains are usually set lower for recordings from children and macaque monkeys who have large potentials at the scalp, and, for an adult, gains are usually set lower for recordings near the eye as they tend to be blocked by large blinks.

Spontaneous EEG potentials recorded at the scalp of healthy, awake subjects generally fall in the 50 to 100 uV range peak-to-peak. A good way to see how large the EEG is likely to get is to ask the subject to close their eyes and relax. The resulting "alpha" activity will synchronize and increase in amplitude and thus provide a sense of the range of the EEG. There can be a good deal of variability between individuals as a result of differences in cortical folding and skull thickness, and in some cases within an individual across scalp locations.

Filter settings: (based on Regan 1989)

The main function of the amplifier's filters is to attenuate certain ranges of frequencies in the changing scalp potentials. Although the decision to deliberately throw away some of the EEG "data" may be initially alarming, at least some signal conditioning is necessary if the analog data are being digitized. At a given sampling rate, frequencies above twice this sample rate will appear in the digital record as spurious low frequency components, i.e., aliasing, as discussed previously. Aliasing can be avoided by bandlimiting the frequency content of the amplifier output – in particular, by ensuring that high frequencies are adequately attenuated. Also, if the cortically generated potentials of interest do not occur at certain frequencies and, as is often the case, artifacts do, there may be grounds for attenuating these frequencies as well. Cortical potentials can range from very slowly changing potentials, essentially DC shifts that evolve over the course of several seconds, to brainstem auditory evoked potentials (BAEPs) which, in addition to lower frequency components, have significant energy at frequencies over 500 Hz. Accordingly, for BAEPs, a frequency pass band of about 20 - 3000Hz is appropriate, but, in this case, slower cognitive potentials or alpha band activity fall below the low frequency cutoff and would be severely attenuated. The question of exactly what passband to select crucially depends on the frequency characteristics of the signal one wishes to observe.

Analog filters do not exhibit ideal behavior in the sense that they do not completely eliminate energy above or below a specified frequency and pass remaining frequencies through unchanged. In the first place, the extent to which frequencies are attenuated varies as a function of frequency, such that gain drops off with frequency in a gradual fashion. The frequency response of a filter is typically characterized by 1) the "3-db corner frequency", i.e., the frequency at which the gain is -3db, and 2) the "rolloff", or rate at which gain changes as a function of frequency, which

can be expressed as change in dB per octave (doubling of the frequency). In addition to changing the amplitude at a given frequency, analog filters also change the phase of the signal. Although this change is greatest outside the pass band, phase lead or lag can also extend into the pass band and distort the timing of peaks and troughs in the waveform. There are many different types of analog filter designs, and they make different types of tradeoffs. For instance, some have particularly flat pass bands, or particularly steep rolloffs, or linear phase vs. frequency characteristics.

Data Analysis.

There are a wide variety of techniques for analyzing EEG data. This section will outline some of the filtering and time domain averaging techniques one is most likely to encounter in a report of cognitive electrophysiology research.

Schematically, the data in an ERP experiment can be classified hierarchically as follows:

Each experiment = a set of subject groups
Each subject group = a set of subjects
Each subject = a set of within subject experimental conditions/levels
Each within subject condition = a set of trials
Each trial = a set of electrode positions (channels)
Each channel = a set of data samples recorded at regular intervals
Each sample = a digitized version of an analog scalp potential value

Conceptually, then, the data from a typical cognitive electrophysiology experiment consists of a six-dimensional array of digital EEG samples, with the dimensions of the array as follows:

Group (for between subject variables) x Subject x Condition (for within subject variables) x Trial x Channel x Timepoint

The following will outline some common procedures for visualizing and quantitatively analyzing these large data sets. Of primary interest in cognitive experiments are comparisons between groups and between within subject conditions.

<u>Quantitative analysis</u>

Improving the signal to noise ratio

The peak to peak amplitude of background EEG recorded at the scalp is typically in the 25uV to 100uV range, whereas the amplitude of the modulations of potentials elicited by a stimulus event, such as a tone pip or a visually presented word, are often much smaller, often in the range of 5 to 10 uV. Furthermore, the difference in event-related amplitude modulation by an experimental manipulation can be smaller still. It is not uncommon to expect an experimental manipulation to produce only a 0.5-2 uV difference between ERPs. Although certain kinds of very disruptive stimulus events can produce ERPs that are clearly visible in the continuous EEG, in general, the small amplitude ERPs of experimental interest are thus buried in relatively higher amplitude background EEG. Under the assumption that the time-locked potential changes are the signal of interest and the non-time locked background EEG is noise, the ratio of root mean square (rms) ERP signal to rms EEG noise in a single trial is quite low (on the order of 1:10, and, for the even smaller differences between experimental conditions even 1:50 or 1:100 in single trials). If the recorded potentials contain high amplitude noise from non-cortical sources, such as 60Hz interference from power lines or muscle (myoelectric) activity, the signal to noise ratio can be much worse. This signal to noise ratio can be improved in a variety of ways, including selectively filtering in the frequency domain and time-domain averaging.

Frequency filtering

Filtering is the process of altering the frequency content of a signal. It involves selectively removing or altering parts of the frequency content of a signal to create a new signal. Filters, both in continuous and discrete time, can be grouped loosely into one of four categories: lowpass, highpass, bandpass, and notch. A lowpass filter passes low frequencies but stops high frequencies above a cutoff frequency; it is useful for noise removal and data smoothing. A highpass filter passes high frequencies, but stops low frequencies below a cutoff frequency; it is useful for removing DC potentials or low frequency drift. A bandpass filter passes a band of frequencies, stopping all others. A notch filter removes all frequencies between two cutoff frequencies; it is useful for removing noise at a particular frequency, e.g., 60 Hz. Figure 4 shows examples of filtering.

One technique to improve the signal to noise ratio is to empirically or theoretically identify the frequency characteristics of the signal of interest and then selectively filter the data to pass these frequencies while stopping the others. For instance, if the ERPs of interest evolve relatively slowly and have little energy at higher frequencies, then it may be appropriate to low-pass filter the EEG accordingly, say at 35 to 50 Hz (i.e., to pass only frequencies below 35 or 50 Hz). Alternatively, if the energy in the ERPs of interest is distributed largely in higher frequencies, for

example as with auditory brain stem potentials, it may be appropriate to high pass filter the EEG above 100 Hz (i.e., pass everything above 100 hz) to reduce the slower potential shifts which are unrelated to the BAEP. In addition to passing energy carried by the ERP of interest, if the frequency characteristics of the noise are known, then this noise can be selective filtered out. For instance, if the intrusion of 60Hz interference from power mains is a problem, it is possible to notch filter the EEG – i.e., to strongly attenuate frequencies in a narrow neighborhood around 60Hz.

Filtering can be done either with analog filters built in to circuitry of the bioamplifier or with a variety of digital signal processing techniques applied to the digitized EEG data record. As noted above, some analog filtering is necessary to avoid aliasing artifacts. However, in general filtering should be undertaken with caution for several reasons. First, the performance of the filters used should be understood. Filters typically merely attenuate rather than eliminate energy at various frequencies and the extent of the attenuation can vary from filter to filter. As already mentioned, certain sorts of analog filters significantly phase shift frequencies in the pass band, and digital filters often corrupt the signal at the beginning and end of the digital record. How much of the data is corrupted may depend on the frequency response function of the filter; obviously, this can be a problem if there is ERP signal of interest in the corrupted portion of the data. Finally, a difficulty with all frequency based filtering is that some ERPs may have significant energy in the same frequency range as artifacts, in which case attenuating the frequencies characteristic of the artifacts will attenuate the ERP signal as well. Specific examples include the potentials in the first 200 milliseconds, which often have significant energy at the higher frequencies characteristic of myoelectric artifact, and the slower "cognitive" ERPs such as the P300 and the N400, which have significant energy in the 10-12 Hz band characteristic of background alpha activity. Filtering to attenuate frequencies in these ranges would throw out the ERP baby with the noise bathwater.

Time-domain averaging

The most common technique for improving the ERP signal to noise ratio is signal averaging in the time domain. In this technique, the data points from each of N trials in an experimental condition are aligned with respect to the onset of the stimulus. Then, for each of trials for each moment in time, the arithmetic mean of the samples recorded at that point is calculated. The result is the time-domain average of all the ERPs in the experimental condition. Averaging works well as long as the signal of interest is exactly the same in waveshape and timing relative to the time-locking point (usually stimulus onset). Of course, in practice we do not know how badly the single trials averaged to yield a P300, for instance, violate this assumption, but undoubtedly they do, for they probably vary in amplitude, latency, and waveshape.

In principle, the numbers of trials that must be averaged to get a meaningful estimate of an average ERP or the difference between two ERPs depends on the characteristics of both the ERP and the background noise. In practice, for very small signals such as BAEPs, very large numbers of trials, e.g., on the order of a thousand, are used. For cognitive experiments in which a difference between experimental conditions of two or three microvolts is expected, averaging 30 stimuli in each condition for each of about 16 subjects is a good working number. Good results can still be obtained with fewer trials if the difference between experimental conditions is larger or if there is less variability than usual across trials and subjects, as is often the case for ERP deflections before 200 ms post-stimulus-onset. The number of trials needed also depends on what question is being addressed. For example, five trials may be sufficient to show that a semantically anomalous word elicits an N400 whereas a highly predictable, semantically congruent word does not; however, it may be necessary to average 50 or 60 trials in each of two conditions to show that the N400 begins earlier in one condition than the other. Figure 5 shows the result obtained with different numbers of trials are averaged together.

Time domain averaging assumes that the ERP signal in each trial is identical and that the background EEG varies randomly. If this is the case, then at a given moment in time a random positive value due to EEG in one trial will tend to be offset by a random negative value due to EEG in another trial. Hence, as the number of trials averaged together increases, the mean of the random EEG tends toward zero. This is a statistical argument about the improvement of the signal to noise ratio under the assumptions of perfect stationarity of the signal and genuinely random noise. It should be noted that, even in principle, this technique offers an improvement in the signal to noise ratio of 1/sqrt(n) where n is the number of trials being averaged together. The noise does not "cancel out" in the average – it is only reduced. Furthermore, as noted above, both the assumption of a stationary signal and random noise are violated to some degree in the EEG data. And finally, like any other calculation of an arithmetic mean, the time domain average is sensitive to outliers and these must be guarded against in the process of screening the data. Indeed, one way to improve the averaging process would be to check for outliers.

Data Reduction

Working with time domain average data

The process of time domain signal averaging collapses the data recorded at each scalp location in each trial of a given subject into an average ERP for each experimental condition of interest. These waveforms, however, are a time series of data points (i.e., a waveform in time) and

it still remains to reduce them to the value of a dependent variable suitable for statistical analysis. EEG research, in other words, is data rich. Although choosing the appropriate measure to derive from the recorded scalp potentials can be difficult, this is a nice problem to have. Choosing a given measure to use as an independent variable for a specific experiment can be guided by the question at hand, theory, previous experimental work, the usual and customary procedures found in the literature, and practical factors like ease of calculation. The three most common measures used in cognitive experiments are: peak waveform amplitude, peak waveform latency, and mean waveform amplitude (Figure 7).

These measures are often computed for a specific pre- or post-stimulus interval and are typically computed after the average waveform has been normalized by subtracting the mean amplitude of the voltage in a specified prestimulus window (usually 100 or 200 ms prestimulus). This "baseline" adjustment ensures that the ERP waveforms all start, on average, from the same prestimulus voltage. Peak amplitude or latency measures are taken for the largest peak in some pre-specified time window and, as such, are quite sensitive to noise (much more so than area measurements). One alternative for ameliorating the effects of the noise is to apply a filter. However, this is sometimes not possible because, as discussed, the frequency content of the noise and the signal of interest overlap. In this case, one can set additional criteria for the measurement of a peak; for example, values must first increase for a certain number of samples in a row and then decrease for a certain number of samples in a row. This procedure will not be fooled by a short noise spike.

One alternative to calculating amplitude relative to a prestimulus baseline is to calculate amplitude relative to another specified peak – i.e., a peak-to-peak amplitude measure. Of course, for this measure one cannot determine the extent to which each of the two peaks is modulated by an experimental manipulation without an additional base-to-peak measurement of each peak. Alternatively, the peak can be identified by its relative position in a wavetrain (e.g., the first negative or second positive peak), and its amplitude or latency calculated. In this case, the peak is not constrained to occur in some specific time interval, but variability in the wavetrains can still make it difficult to reliably identify the peak of interest.

If the relative timing of cognitive processes is of interest, the latency of waveform features are highly relevant, and the latency of a particular peak might be an appropriate choice. For instance, experiments that examine the association and dissociation of ERPs with response latency have used the latency of the P300 peak as the measure of interest. One can also measure the onset latency of a particular peak or effect (measured in difference ERPs for two conditions).

Amplitude and latency measures of this type are only suitable when the waveforms at issue have clearly defined peaks. This is generally the case for sensory evoked potentials out to perhaps 300 or 400 ms, and, accordingly, these measures are often used in studies of perceptual and attentional processes. However, after about 400 ms, the poststimulus ERP waveforms often show slow shifts rather than clear peaks and valleys, and the motivation for selecting one of several rather variable bumps in the waveform as "the" peak latency is unclear. For these later potentials, mean amplitude in an interval of interest can be used to smooth out smaller variations and characterize the central tendency of the potential at that latency. To characterize the amplitude of a negative peak 400 ms post-stimulus-onset, for example, mean potentials in intervals between 100 ms and 300 ms in width have been used. Measuring the area under the curve or the mean amplitude essentially gives the same value, except the latter takes into account the duration of the window for which the area has been calculated (mean amplitude 300-500 = area 300-500/duration of window = area/200). Since the area/mean amplitude measure is less sensitive to high frequency noise than the peak measure, one can combine the strengths of the two measures by measuring the mean amplitude of a narrow window (~50 ms) around peak amplitude.

There are many alternatives to these simple time-domain voltage measures that might be appropriate, and some of these are reviewed below. Nevertheless, peak latency, peak amplitude, and mean amplitude are the most common measures for a wide variety of cognitive ERP research. That said, one of the most difficult decisions involves determining where a measurement window should begin and end. As yet, there are no commonly-accepted, objective criteria for setting these values.

Other analyses

Although considerable attention has been given to transforming time domain potentials into other domains in order to get different perspectives on the data, there are techniques in addition to averaging that operate on time-domain potentials.

Calculating difference potentials is a simple but useful technique. Difference potentials are calculated by subtracting one waveform from another, time-point by time point. The result is a single waveform that represents the difference in potential between the two original waveforms. It says nothing about localization or interpretation; it is simply a difference between two ERP waveforms. The polarity of the potentials across time depends on what condition is subtracted from what other condition. The reverse subtraction would yield opposite polarity components. Moreover, a flat line (i.e., zero difference) could result because there is no potential in either case, or from two very large potentials (both negative or both positive) that are the same. Likewise, a large potential

could reflect no activity in one condition and a large potential in the other, two conditions with large potentials of opposite polarity, and so on. This technique has the effect of canceling out amplitude deflections that the waveforms share, although, again, such shared similarity in surface form does not necessarily imply shared underlying sources. Subtraction is a useful way of seeing whether two ERPs are the same or different, and if different, when the differences start, stop, and peak. It is often used for comparing time domain averaged waveforms in different experimental conditions. In a typical visual or auditory experiment, the stimuli in both conditions will evoke a series of peaks and troughs and the effect of the experimental manipulation will be seen as an amplitude or latency modulation of one of these features. Calculating the difference wave can eliminate the shared peaks and troughs and highlight just the modulation resulting from the experimental manipulation. Difference ERPs are not without their problems and are no panacea to complexity; they should never be interpreted without recourse to the original waveforms. Both amplitude and latency variation between two conditions will lead to a difference, although they obviously have different implications for processing. Moreover, if the two ERPs that go into the subtraction are fundamentally different, then the difference ERP would be at best descriptive of that difference.

Special analytic techniques are also sometimes employed to deal with the issue of latency jitter. Latency jitter poses a problem for conventional time domain averaging because it violates one of the assumptions of averaging. If the same waveform is jittered in time, the amplitude of the average peak will underestimate the value of the true peak of the original waveform. If only one component of the ERP varies in latency from trial to trial, and its amplitude remains the same, then that component will be less peaked (i.e., broader in duration), and smaller in amplitude; all else being equal, however, the area under the curve for that component should be unaffected by the latency jitter. Thus, peak amplitude measurements are not the best way to compare a component in two conditions known to vary in latency from trial to trial; comparing mean amplitudes is likely to give a truer answer. If reaction times are recorded along with the ERPs and the reaction times vary from trial to trial, then so might some ERP components (such as N2 and P3, for instance). Just how much variation there is can be estimated by breaking the ERP average into subaverages based on mean RTs – for example based on a median split on RTs or a split into thirds or quartiles.

Woody (1967) has proposed an adaptive filtering approach to latency correction for jittered components on single trials. The procedure starts with a template waveform (resembling the presumed waveshape of the component of interest) and in each single trial determines the latency at which the template best fits the waveform (i.e., has the highest correlation). The signals are then shifted and lined up according to these latencies and averaged, and the process repeated. If

started with an appropriate template, this technique can converge on latency corrected average waveforms. The procedure is reiterated until there is no improvement in the average signal by shifting the latency of the component on any individual trial. However, since there is always a "best" match even if it isn't very good, the procedure can line up random noise in a way that looks most like the template and, through a series of iterations, eventually extract a something from nothing if the signal to noise is too low (i.e., below about 0.5RMS (Ruchkin, 1987)).

Visualization: Know thy data

As should be clear by now, ERP research is an information-rich technique, and a crucial part of understanding the data is being able to see it. The more ways one can visualize the data, the better, since there are different patterns in the data and different renderings emphasize different aspects of these patterns. Given the enormous quantities of data involved, data visualization has historically been a non-trivial enterprise. However, the advent of the microcomputer, and, more recently, the "multimedia" computers designed to accommodate digitized sound and video, has made processing multi-megabyte files routine. As software catches up with these hardware advances, data visualization is becoming much more tractable than it was even a few years ago.

As noted above, in their most basic form, the data of a single ERP experiment can be thought of hierarchically. The key to understanding these vast quantities of data is to be able to visualize it at these various levels. Some of the levels of interest include single trial waveforms, condition average waveforms for individual subjects (collapsing across single trials; this is the most common average), and condition average waveforms for a group of subjects (collapsing across subjects, also known as a grand average). Rarely can one see an ERP in a single trial EEG epoch without additional filtering, as the data are too noisy. However, with some filtering, the larger potentials can be unveiled. Also, note that while the grand average is a nice way to look at the data, and is thus often used for illustrative purposes in publication, it is not the basis for measurements, as it provides an estimate of the mean but not an estimate of the variance across subjects. Generally, it is assumed that each subject contributes equally to the average. This assumption may be violated if different subjects have very different numbers of trials in their averages or very different noise levels. The presence of an effect in grand average comparisons, therefore, cannot by itself be taken as evidence for a reliable effect because they can be very misleading, on occasion. Such comparisons are, however, typically a good way to begin to see the effects of a particular manipulation, as long as this is followed by careful scrutiny of the same comparison for each individual subject's data.

Waveform data can be represented in a variety of ways; the most common are time-series plots, surface maps, and 3-D rendering. In a typical time-domain plot, for example, two condition averages for several channels in an entire experiment may be overplotted. One may position the electrode locations on the figure in order to roughly represent, in two dimensions, the relative locations of the electrodes in three-dimensions. Although distances are not accurately represented, the relative positions of the electrode positions lateral to the midline and from front to back are preserved. Such a plot makes it easy to compare the timecourse of the waveforms in the two conditions at different post-stimulus latencies. Another option is to select the values of the voltage data from all channels at a single moment in time and represent the level of this voltage with a color (e.g., positive is red, negative is blue); this makes it easy to see how these values are distributed across the scalp. This technique is not restricted to voltages and can be used to represent the scalp distribution of any single quantity: mean voltage, peak voltage, alpha band power, etc. Although this type of visualization accurately represents electrode positions in three dimensions, it provides only a "snapshot" of the measure of interest and does not represent the whole time course. However, the time-course can also be visualized in three dimensions by constructing the 3D image for each data point in the series and then displaying them in sequence. If this sequence can be displayed at a sufficiently high rate, the result will be an animation of the time course of the potential at all points across the scalp. Of course this type of visualization is limited to dynamic media, such as computer displays and video tape. A static approximation of the animation can be acheived in print by selecting a small number of images from the sequence, e.g. at regular intervals or points of experimental interest.

Single trials

At the level of individual trials, time series plots can be used as computer renderings of conventional pen-and-paper EEG strip charts. In this technique, the data from each recording electrode is displayed on an area of the computer screen. There are two main contexts for visualizing single trial data and they have different computational requirements. First, during the acquisition of EEG data, there must be some way to monitor the signals being recorded so that problems, e.g., myoelectric artifacts and equipment failures can be detected and fixed as soon as possible. The display might take the form of a continuous strip chart or a sweep display of the sort familiar from oscilliscopes. Displaying the EEG data in real-time as a continuous strip chart requires redrawing the portions of video display fast enough to animate the waveform. With the current generation of high-performance video display adapters and fast (e.g., 200MHz) microprocesser CPUs and multimedia software tools, programming this sort of real-time animation

is less challenging than it was in the past. The second main application of single trial visualization is the review of the EEG data during which artifacts can be identified and data integrity ensured. In this case, previously recorded EEG data, typically stored on a computer disk, are displayed on a screen and the user has controls that allow navigation through the data. The visualization of single trial time series at this stage is important since these are the basic EEG data and the results of all subsequent analysis are strictly determined by their properties. Like any time-series data, a range of statistical properties can be quantified. It can be useful to have descriptive statistics for the data including their range and standard deviation and other properties such as the power spectrum. This kind of information can be instructive, but does not replace the need for inspecting the raw data itself.

Reduced data

As noted above, the signal processing technique of time-domain averaging takes a given subject's individual experimental trials, aligns them with respect to stimulus onset, sorts them into the different (within subject) experimental conditions, and then computes a single waveform that is the mathematical average of the individual trials in each condition. These single subject condition averages are waveforms with the same structure as single trial, i.e., some number of channels with some number of samples per channel. The same resources used to visualize single trial data should be able to accommodate single subject condition averages with at most minor modification. These condition averages for each subject may then be reduced according to the measure of interest, and the inferential statistics that support hypothesis testing are conducted on these individual subject measures.

Line and bar graphs serve the same purpose in ERP research as they do in behavioral research where they are a useful way to represent the main effects or interactions of the experimental variables. So for instance, in an experiment in which the measure is mean amplitude in a post-stimulus interval between 300ms and 500ms and there are three levels of an experimental variable, the mean amplitude and standard errors for each level can be represented in a bar chart. If there is more than one multilevel independent variable, then interactions can also be represented. A summary bar chart of means and standard errors of the reduced data for the entire design is the most condensed form of representation of the ERP; this is derived from single subject data.

FIGURE CAPTIONS

Figure 1: Examples of artifacts in EEG

Sample EEG traces from multiple electrode sites showing the appearance and distribution of various types of biological and technical artifacts. The top trace of each plot is derived from an electrode placed below the eye, referenced to the mastoid; the second trace is derived from electrodes placed on the sides of each eye and referenced to one another. The rest of the traces are from scalp electrodes, labeled according to their locations in the 10-20 system. Figure 1-A shows a *horizontal eye movement*, which appears as an abrupt voltage deviation most prominent in the horizontal eye channel, but also visible in recordings from electrodes near the front of the head (e.g., f3 and f4). Figure 1-B shows a *blink*. Note the reversal of polarity in the recording from the electrode placed under the eye and those from frontal scalp sites, such as fz, f3, and f4. Figure 1-C shows *EKG (heartbeat)* artifact. It is being picked up by the reference electrode and is therefore visible in all channels. Figure 1-D shows the high frequency activity associated with *movement (muscle activity)*. Technical artifacts are shown in Figures 1-E and 1-F. *Amplifier blocking* (where the signal has exceeded the range of the amplifier) appears as a flat line in Figure 1-E. Figure 1-F shows the *ambient 60 Hz electrical noise* that is picked up by an "open channel" -- that is, a channel in which a complete circuit between electrode and amplifier has not been made, due to, for example, a faulty wire or loose connection, etc.

Figure 2: The 10-20 system of electrode placement

The top left of the figure shows a frontal view, illustrating how the central line of electrodes is measured relative to the left and right preauricular points, while the top right of the figure shows a side view, illustrating how midline electrodes are measured relative to the nasion and inion. The bottom of the figure shows all the positions projected onto a top view of the head. From Jasper 1958.

Figure 3: Aliasing

The relationship between sampling rate and signal frequency is illustrated in this figure, where the left side shows sine waves that are being sampled (dots show sampling points) and the right side shows the sampled points connected to give the estimated measured signal. When the sampling rate exceeds two samples per cycle (as given by the Nyquist frequency), as in A-D, the result is a fairly faithful representation of the original signal. In contrast, when the sampling rate falls below two samples per cycle, as in E-L, the result is *aliasing*. That is, the measured signal falsely

suggests the presence of lower frequencies than those contained in the original signal. From Regan 1989.

Figure 4: Filtering

The result of filtering in the time domain (4-A) and in the frequency domain (4-B). The left-most column of Figure 4-A shows simulated EEG containing, in the top row, no alpha activity, and in the remaining rows constant alpha, alpha in the early part of the time window (in two different phase relationships), and alpha in the late part of the time window, respectively. Identical high and low frequency noise have been added to all five signals. Low pass filtering at 8 Hz (second column) pulls out the (identical) low frequency signals in all five cases. Similarly. high pass filtering at 20 Hz (third column) pulls out the (identical) high frequency signals in all cases. Bandpass filtering in the critical frequency range for alpha activity (10-14 Hz, fourth column) reveals the critical differences between the signals in each case. Figure 4-B shows the same data, but in the frequency domain -- where the x-axis gives the frequency and the y-axis shows the power at that frequency. Low pass filtering (second column) removes all the power from frequencies above (in this case) 8 Hz, while high pass filtering (third column) removes all the power from frequencies below 20 Hz. Bandpass filtering (fourth column) removes the power from frequencies outside the range of interest (here, 10-14 Hz). Note that there is some power in the alpha frequency range even when no alpha is in the signal; note also that this type of representation is free of time, so that traces containing alpha in the early and the late part of the signal look identical when transformed into the frequency domain.

Figure 5: Time-domain averaging

The left side of the figure shows single trial EEG data, creating by adding random noise to the same "target" ERP signal. Overlapped are 4, 16, 32, and 64 such trials, respectively. The right side of the figure shows the resulting time-domain average, with the target signal overlapped. When only four trials are averaged, as at top, it remains difficult to pull out the target signal from the noise. Once 64 trials have been averaged (bottom), however, much more noise has been eliminated, and the resulting signal closely approximates the "target" ERP.

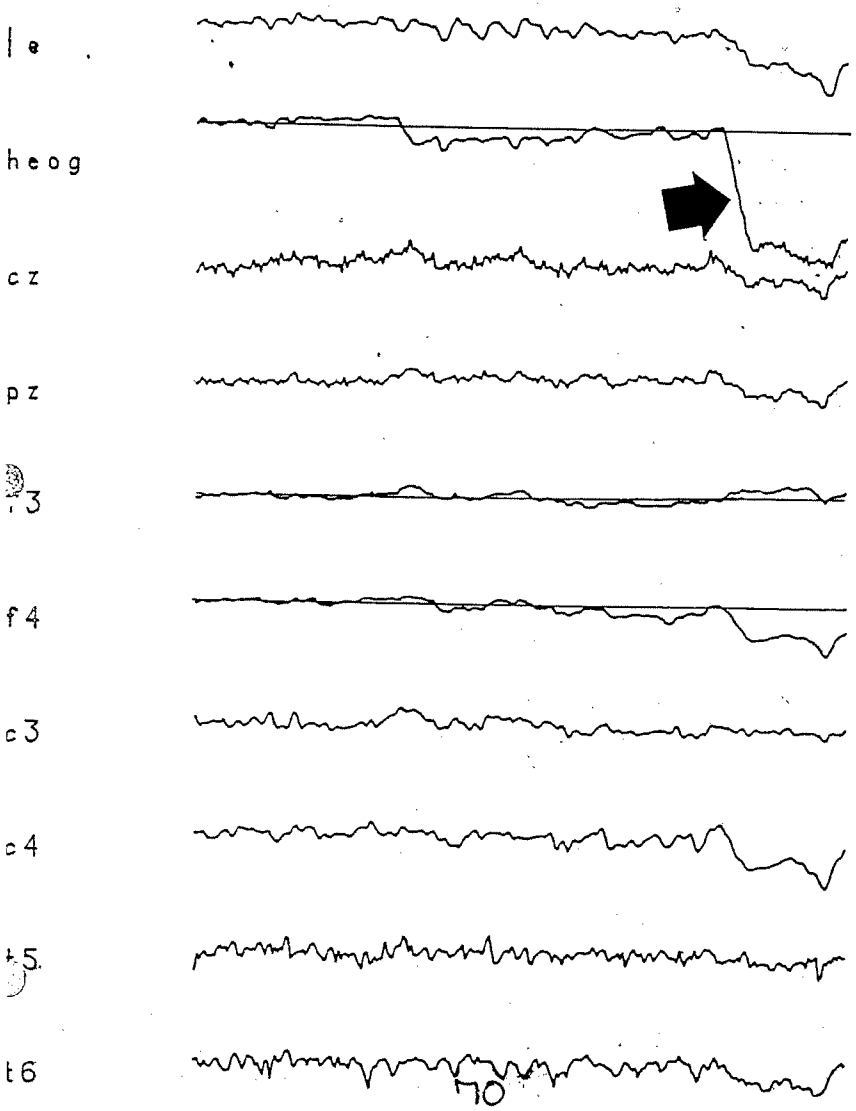
Figure 6: Examples of data reduction measures

This figure illustrates various data reduction measures on a sample ERP trace. Measurement is done relative to a pre-stimulus baseline (i). Measuring peak latency (ii) involves finding the largest amplitude deviation in a given time window (here, 75-110 ms) and then determining its timing

relative to stimulus onset -- in this case, 110 ms. Measuring peak amplitude (iii) also involves finding the largest response in a given time window (here, 150-300 ms) and reporting its amplitude (4.5 uV). One can also measure the area under the curve in a particular time window (iv); that, divided by the time interval, gives a mean amplitude (v).

FIGURE 1-A

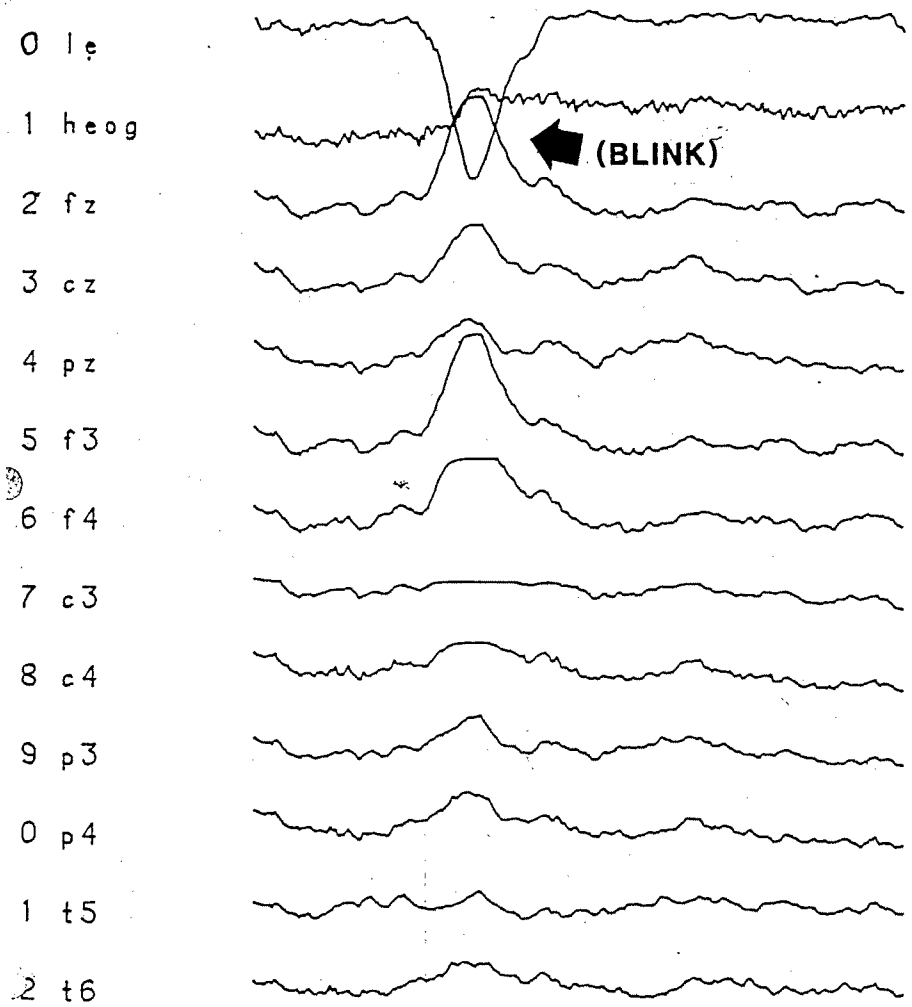
LATERAL (HORIZONTAL) EYE MOVEMENT



70

FIGURE 1-B

VERTICAL EYE MOVEMENT



71

FIGURE 1-C

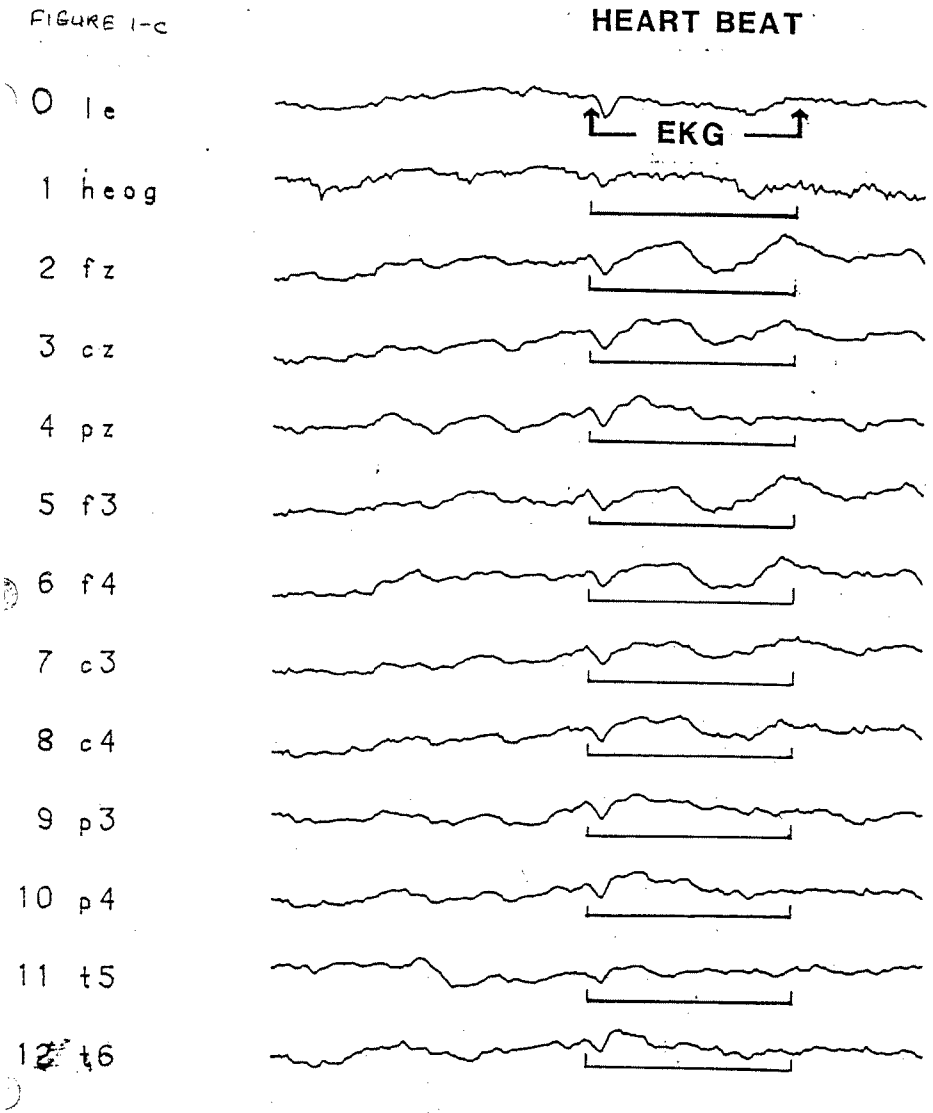


FIGURE 1-D

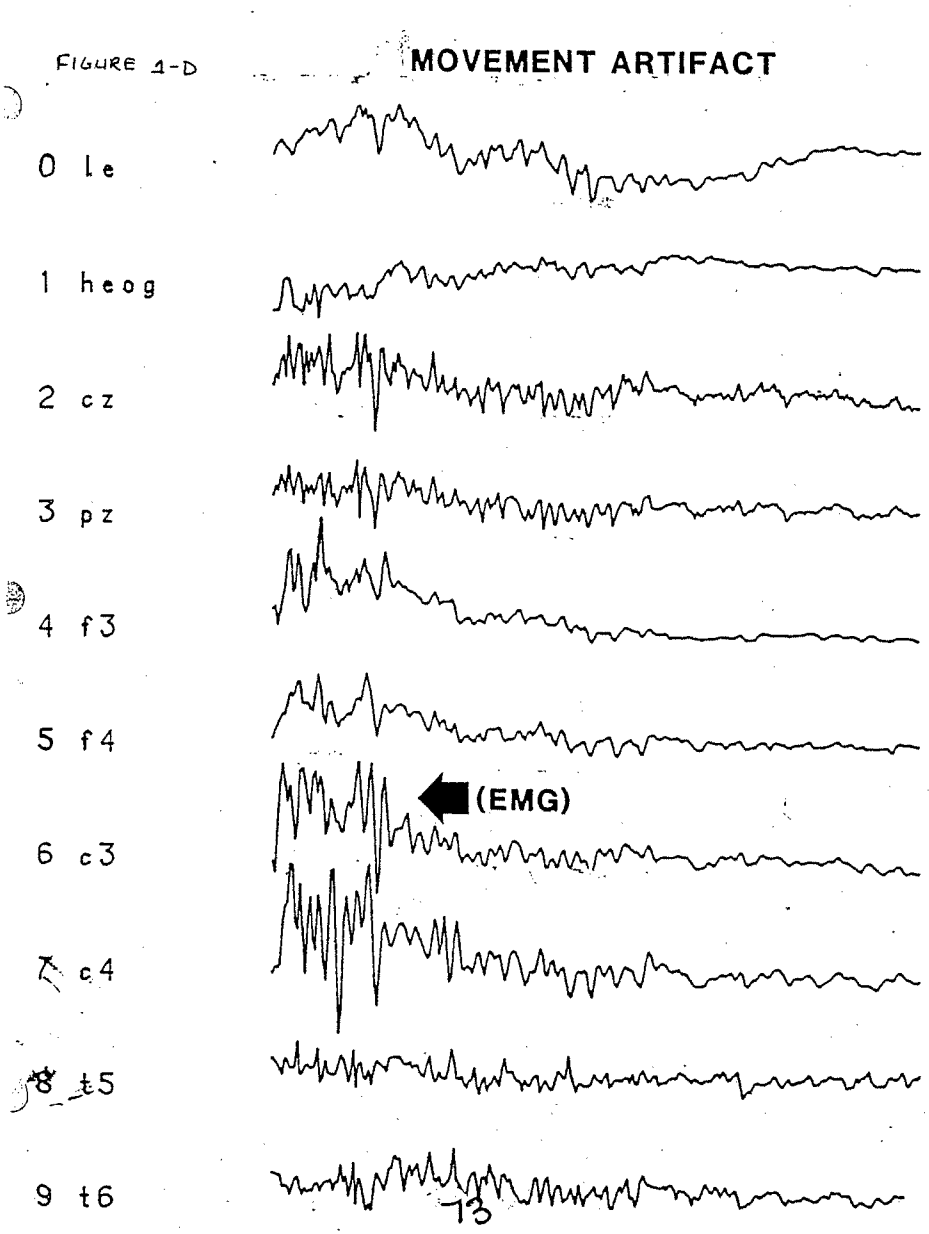


FIGURE 1-E

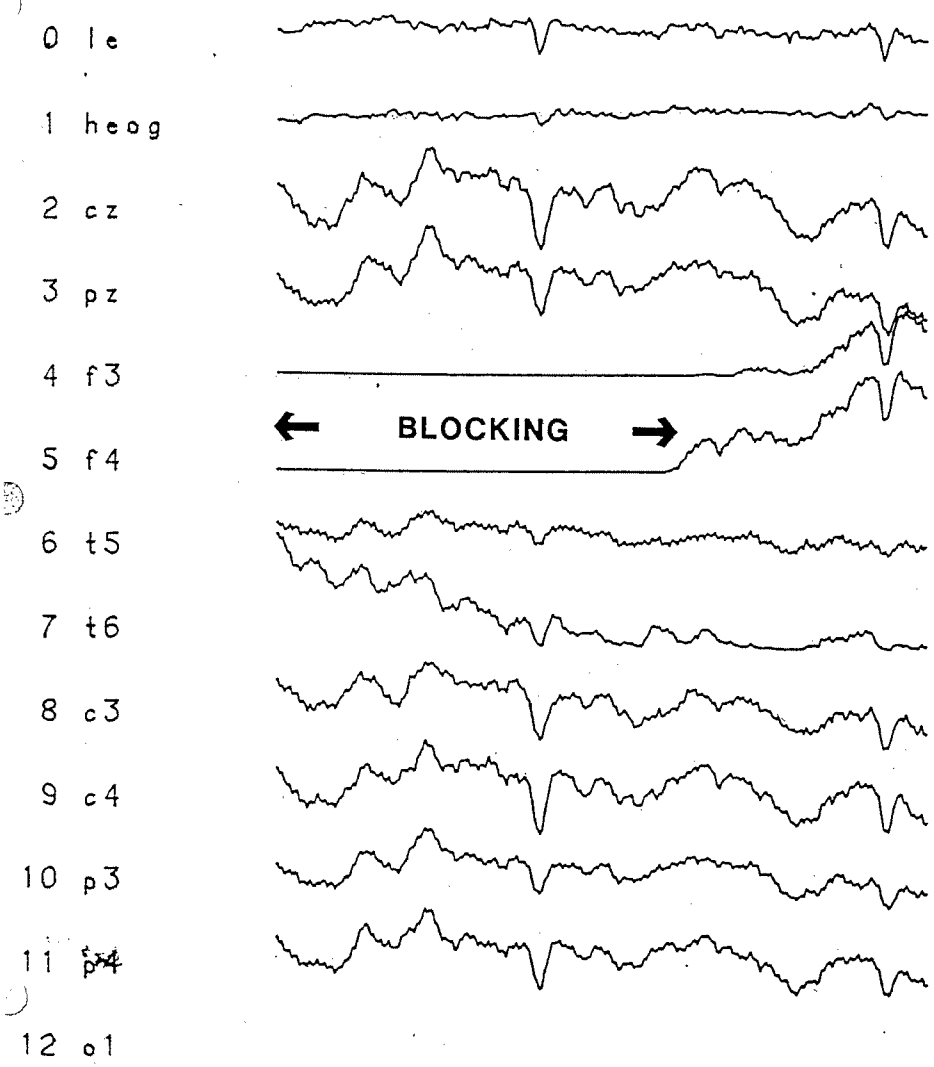
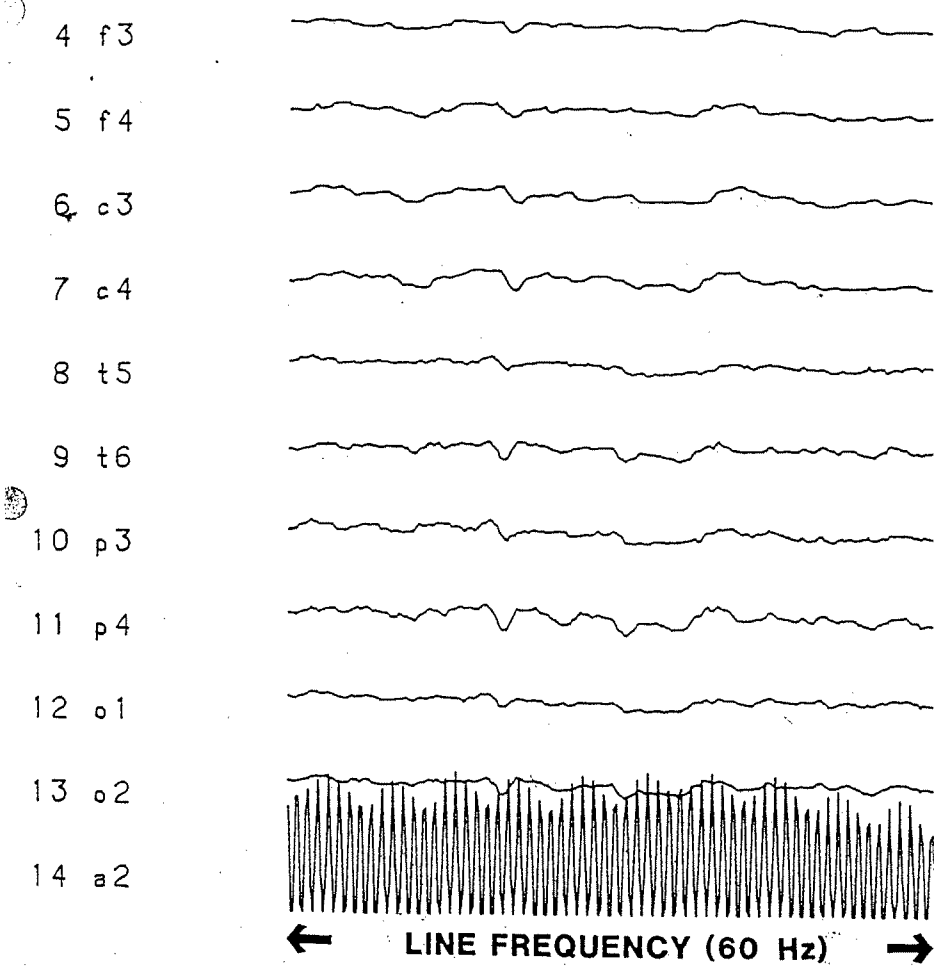
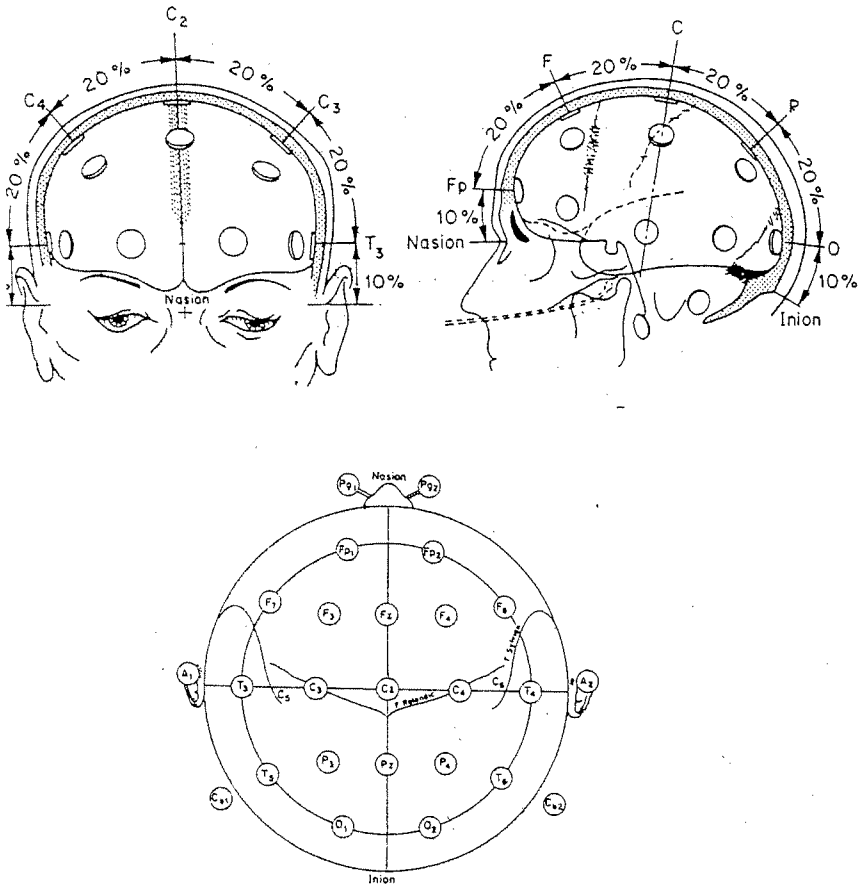


FIGURE 1-F





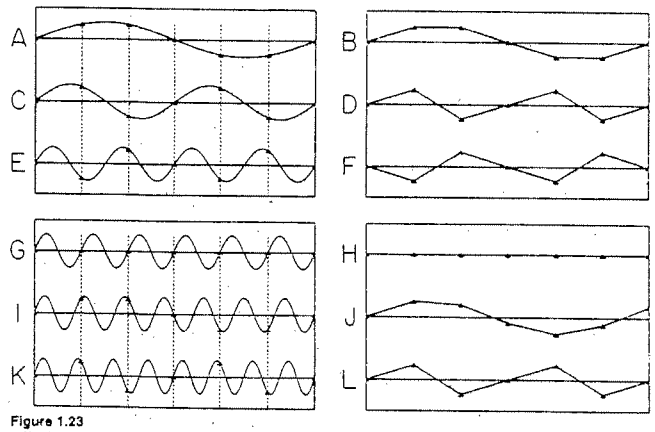


Figure 1.23
Aliasing
Six samples per cycle of a sine wave (A) give a rather faithful representation (B). Three samples per cycle (C) still correctly represent the sine wave's frequency (D). Fewer than two samples per cycle (E–K) incorrectly represent the sine wave's frequency (F–L) as a frequency that is lower than the true frequency or even (G, H) as a constant signal.

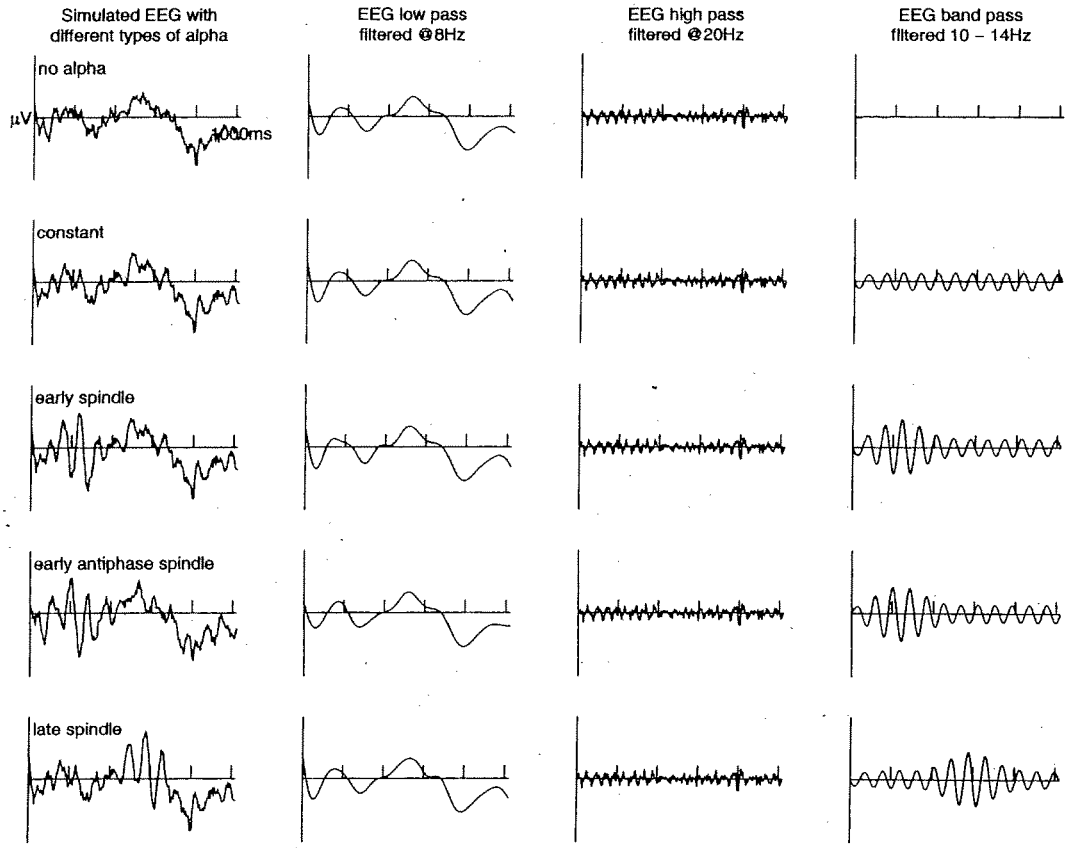


FIGURE 4-A

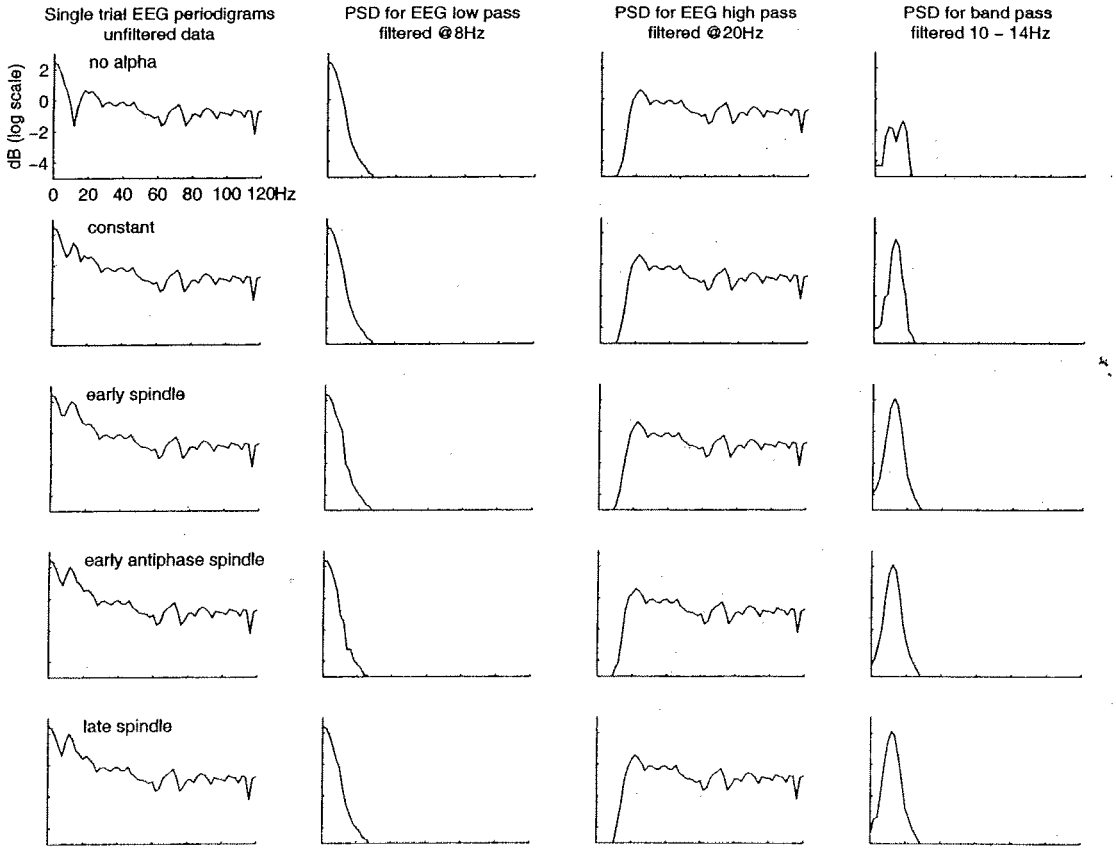


FIGURE 4-B

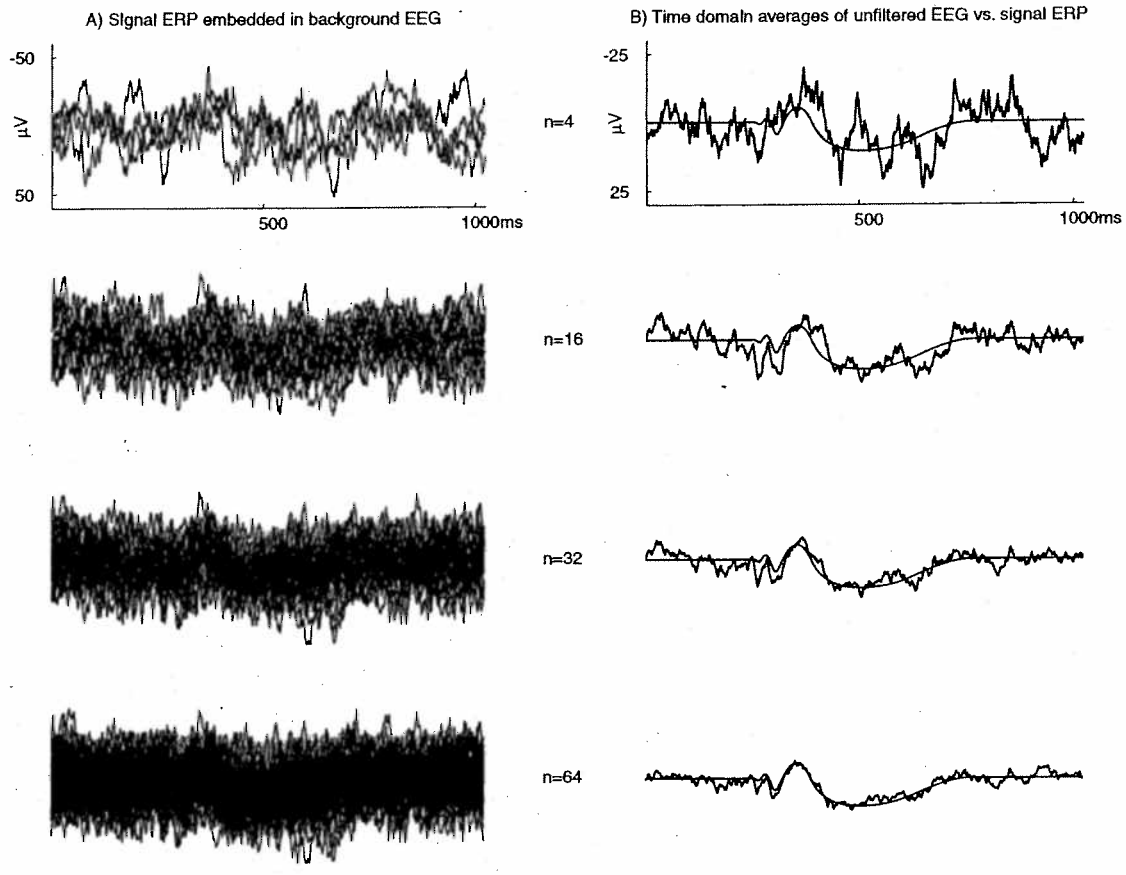


FIGURE 5

8

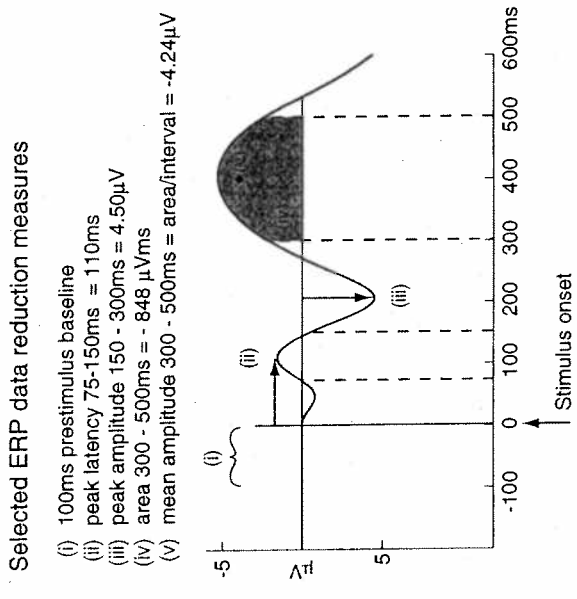


FIGURE 6

81