Issues in the application of the average reference: Review, critiques, and recommendations

JOSEPH DIEN
University of Illinois at Urbana–Champaign, Urbana, Illinois

In standard event-related potential (ERP) recordings, activity at the reference site affects measurements at all the "active" electrode sites. Analyses of ERP data would be improved if reference site activity could be separated out. Apart from current source density methods, which can miss deep or distributed events, the major alternative to an "indifferent" electrode reference is one that combines all active electrodes, subtracting out the activity common to all of them. The average reference has merits (such as an insensitivity to scalp currents near any single electrode) and limitations (such as a dependence on the number and locations of all the electrodes in the average). This review compares the effects of different references on the scalp topography of the auditory N1 recorded with an array of 128 electrodes. Furthermore, it shows how assumptions associated with each reference affect inferential methods such as analysis of variance and correlation. Finally, it seeks to evaluate the efficacy of the average reference in estimating the true zero potential line.

The relentless advance of technology makes it necessary to revisit and reconsider many of the methodological choices in electroencephalographic (EEG) research. In particular, high-density recording montages (Tucker, 1993) make possible techniques not possible for sparse arrays (Srinivasan, Tucker, & Murias, 1998).

One issue that has engendered ongoing debate concerns reference technique. Ideally, the voltage readings should represent a pure measure of activity at the recording site. The difficulty is that voltage is a relative measure that necessarily compares the recording site with another (reference) site. If there is any activity at the reference site, this will contribute equally to the resulting voltage recording (Figure 1). This paper will discuss the issues involved in reference technique, with a focus on the average reference, followed by concrete illustrations using an example data set, and ending with recommendations for future research.

There are various methods for coping with reference site activity. One is to use bipolar electrode arrays in which the electrodes are arranged in closely spaced pairs that are referenced to each other. Since voltage readings represent the difference in potential between the two sites, such pairs detect local changes in potential fields. Because the two electrodes are closely spaced, there is little concern that they may be experiencing different electrical events. The drawback to such arrays is that they essentially compute the first derivative of the scalp fields. Information about absolute differences in potential between nonadjacent pairs is obscured in such a system. In addition, analysts

have preferred to use the second derivative, which produces an estimate of the current source density (CSD) (Perrin, Pernier, Bertrand, & Echallier, 1989). The second derivative is preferable because, unlike the first derivative, it is not directional and can be readily mapped topographically. The CSD is particularly useful because it produces a reference-free estimate of the source and sink locations (current flows to and from the scalp surface) that overlies generators. The CSD does so by highlighting sharp features in the scalp topography and is therefore not suitable for imaging potentials from deep or distributed sources, which will not produce such sharp differences in regional current flows. When such potentials—as from the N1 (deep) or P300 (distributed?)—are of interest, a different approach is required.

The most common technique (common reference) is to use a common reference site that is situated at a distance from the recording sites. In such arrangements, it is recognized that the reference site is contributing to the recordings, but since it will be doing so equally to all the recordings, their relative differences will remain informative. Typical reference sites include the nose and the mastoids. In high-density arrays, this approach becomes problematic because recording sites encroach on the reference site. The reference site can no longer be treated as equidistant to the recording sites. Similar (nearby) recording sites will have similar potentials to that of the reference and the difference (the recorded waveform) will be small. In essence, amplitudes of recordings are attenuated according to distance from the reference site. Although it is sometimes argued that the mastoids are relatively inactive, this has been persuasively shown to be false (Lehtonen & Koivikko, 1971).

One method for addressing this issue is to use a noncephalic reference, such as the sterno-vertebral sites. This arrangement has the advantage of being more distant than reference sites on the head. Such sites are not as advanta-
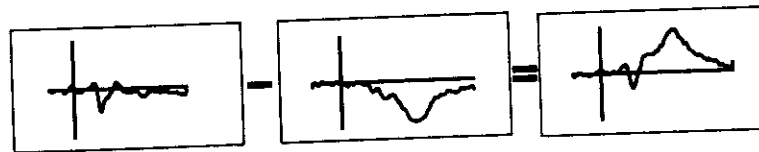
Reference Issue
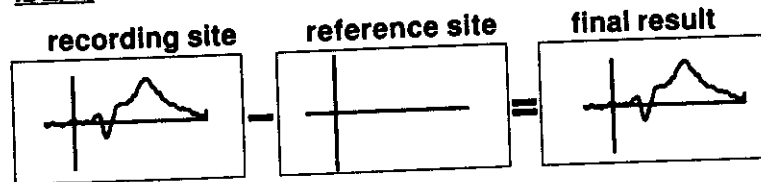
IDEAL

ACTUALLY...

Figure 1. Effect of reference site activity. This figure illustrates how a given recording (right panel) can arise due to either activity at the recording site alone (the ideal case, left panel) or a combination of activity at both the recording site and the reference site (the actual case, middle panel).

geous as their physical distance might suggest, however. It turns out that all sites below the neck are isopotential with respect to one another (Katznelson, 1981); in other words, even a toe reference is equivalent to a neck reference (Wolpaw & Wood, 1982). Unfortunately, whether it is more accurate to state that the toe is as active as the neck (both subject to the same currents) or that the neck is as inactive as the toe (both relatively isolated from EEG current flows) is an empirical question that awaits resolution. In addition, sterno-vertebral references are more invasive than are cephalic references, and other noncephalic references are more susceptible to EKG artifact. They also produce more noise and other artifacts (Lehtonen & Koivikko, 1971). Safety concerns have precluded simply hooking subjects directly to earth ground, which would provide a reference point free of stimulus-locked activity.

The average reference has been proposed as a method for providing an inactive reference (Offner, 1950). The underlying principle is that electrical events produce both positive and negative poles. The integral of these potential fields in a conducting sphere sums to exactly zero (Bertrand, Perrin, & Pernier, 1985). If one can obtain an even and comprehensive sampling of the potential field over the surface, it should provide an estimate of the true zero surface, it should provide an estimate of the true zero (with respect to infinity). In an average reference solution, the average across all the electrodes is considered to be an estimate of the activity at the reference site that has been equally represented in all the recordings. Or to put it another way, a potential field cannot produce the same activity at sites across the entire head, so if such a feature is measured, it must be an artifact resulting from reference site activity. It is subtracted from all the sites, providing (in principle) a dereferenced solution.

Mathematically, the average reference is computed by subtracting the mean of all n electrodes (ΣE_n) from each channel. For data originally referenced to R, a given record-

ing is actually the difference in potential between that site and the reference ($E_n - R$). The average reference is therefore computed by subtracting the mean of all the other recordings: $(E_n - R) - \{[(E_1 - R) + (E_2 - R) + \ldots (E_n - R)]/n\} = (E_n - R) + \{R - [(E_1) + (E_2) + \ldots (E_n)]/n\} = (E_n - [E_1 + E_2 + \ldots E_n]/n)$. To the extent that the assumptions of the average reference have been met, the term ($[E_1 + E_2 + \ldots E_n]/n$) will equal zero, leaving ($E_n - 0$). The reference site itself can be estimated as starting as ($R - R$) (by definition) and then subtracting the average of the channels, resulting in $(R - R) - \{[(E_1 - R) + (E_2 - R) + \ldots (E_n - R)]/n\} = (R_n - R) + \{R - [(E_1) + (E_2) + \ldots (E_n)]/n\} = (R - [E_1 + E_2 + \ldots E_n]/n)$. Again, if the assumptions are met, the average term will equal zero, leaving just ($R - 0$).

It turns out that data can be freely rereferenced after collection algorithmically (Osselton, 1965). To define a site as the reference is simply to define it as indicating zero voltage at each time point. In this case, any divergence between its trace and any other site is by definition due to the other site, and the waveform for that site should incorporate this difference. This modification can be carried out by simply subtracting the reference site waveform from all the locations. The end result is for the reference site to be set to zero (a waveform subtracted from itself is a flat line), and the remaining sites to have incorporated the activity previously attributed to it. Likewise, the average reference is computed by taking the average of all the sites and then subtracting it from all of them.

The rereferencing procedure can be described mathematically. For an electrode at Pz referenced to the right mastoid (M2), the voltage difference can be expressed as (Pz − M2). To rereference to Cz, which is expressed as (Cz − M2), one simply subtracts the latter term from the former: (Pz − M2) − (Cz − M2) = (Pz − Cz).

The popular mean mastoids reference requires a special note. It can be carried out in two ways. The first is to phys-

ically link the two sites, forcing them to have the same voltage. As has been noted elsewhere (Katznelson, 1981), such a procedure is likely to distort the scalp potential fields by thus adding a low-resistance current path. Also, such a reference can cause problems for measuring lateralized potentials to the extent that the two mastoids have different impedances (Miller, Lutzenberger, & Elbert, 1991; Nunez, 1991), although at least one empirical examination of this problem did not reveal any noticeable effects (Andino et al., 1990). The other method is to take the arithmetic mean of the two mastoid sites after the fact and use it as the reference level. In this case there will be no distortion of the potential fields, but the mastoids will not be completely reduced to a flat line. After the arithmetic transformation, the two sites will split their difference (having opposite polarities).

Mathematically, a computed mean mastoid reference can be described as ($[M1 + M2]/2$). Given that a site such as Pz recorded originally referenced to the right mastoid starts out at ($Pz - M2$), it can be converted to mean mastoid by subtracting ($[M1 - M2]/2$) from it. This results in ($Pz - M2$) $-$ ($[M1 - M2]/2$) = ($Pz - M2 - M1/2 + M2/2$) = ($Pz - M1/2 - M2/2$) = ($Pz - [M1 + M2]/2$). The left mastoid will equal $M1 - ([M1 + M2]/2) = M1/2 - M2/2$. The right mastoid will equal $M2 - ([M1 + M2]/2) = M2/2 - M1/2$.

Critique of the Average Reference

The average reference has been sharply criticized by Desmedt, Chalklin, and Tomberg (1990). Using mathematical simulations, they argued that the average reference can be inaccurate because in practice it is not possible to sample the bottom half of the head in the same manner as the top half. This undersampling results in the neutral point being misestimated as being higher up on the head than it should be for vertically oriented fields, since the average reference places zero at the midpoint (average) of the *measured* field. More seriously, this effect will not simply alter the DC offset of the ERP since it depends on the orientation of the active generators at a given time point; the authors gave such effects the dramatic term of "ghost fields."

The term *ghost field* is somewhat misleading in that it implies that this effect creates artifactual potential fields. In their article, Desmedt et al. (1990) showed graphic simulations (p. 247) in which negative "ghost fields" were produced ringing the "true field." In truth, such distortions simply represent a shifting upward of the negative region of the dipolar potential field. They stated that these fields were "spurious" because "they do not exist when the potentials over the whole head are used" (p. 250), but this statement is misleading. Negative (and positive) potential fields are always present, but they may simply not be located in the sampled region of the scalp. The situation is much the same as for a mountain rising out of a valley. If a cartographer mistakenly positions sea level (zero potential) too high, he/she will mistakenly identify a portion of the mountain (positive region) as being part of the valley

(negative region), but this will not change the overall appearance of the geography. Likewise, this mispositioning of sea level will produce dramatic changes in the appearance of maps if the cartographer chooses to arbitrarily assign altitudes above sea level one color (e.g., white) and those below another color (e.g., black). Such dramatic changes are an artifact of the illustration method, not of the data. Such coloring schemes imply a qualitative difference between the "negative" and "positive" regions that are in truth part of the same continuous field.

This is not to say that the position of Desmedt et al. (1990) is wholly without merit. While the effect of these "ghost fields" is relatively trivial when examining the potential field across the head as a whole, it can certainly alter the appearance of the waveform at specific electrodes. A component will not be seen at an electrode sited at its zero potential line (it will have zero amplitude), but will be seen if the zero point is shifted away from it. If one is using a relatively sparse array or is focusing on isolated sites, such changes can be quite misleading.

One might ask why the exact location of zero potential is relatively trivial when examining scalp topography, but so important for individual waveforms that it can make a component disappear entirely. This issue is analogous to a measurement problem encountered in the psychometrics of attitude measures. It is common to measure attitudes on Likert scales with ratings ranging from *very positive* to *very negative*. The problem is that when a subject chooses the center point, it is unclear whether this action indicates a lack of an opinion or a combination of strong negative and positive opinions that balance out. Resolving the ambiguity requires changing the question format, such as by having separate rating scales for positive and negative opinions. Likewise, in voltage plots an amplitude of zero can either mean an absence of a component or can mean that the electrode site is simply midway between the positive and negative poles. Unfortunately, the ambiguity in this case is not so readily resolved. In any case, this point underscores that moving the zero line on the scalp topography is relatively trivial (since it is not qualitatively different), although it can be misleading for individual waveforms due to measurement ambiguities.

Desmedt et al. (1990) argued against using the average reference because of these "ghost fields"; this conclusion is premature. Common references avoid this drawback of average reference but have a complementary problem that is avoided by the average reference. As noted, activity at the reference site also produces distortions of the recorded potentials—in effect, "ghost fields." Since both approaches can potentially produce "ghost fields," it is an empirical question which will produce more useful results.

A similar situation applies to the other primary critique— that the average reference can produce distortions of focal features (Tomberg, Noel, Ozaki, & Desmedt, 1990). When a small region of cortex located near the scalp generates the ERP, the scalp potential field will decline rapidly from the site of maximum amplitude; hence the region where the ERP is readily discernible will cover a small area and appear

ove.... ap-
positioning
the appear-
bitrarily as-
white) and
h dramatic
thod, not of
litative dif-
regions that

smedt et al.
fect of these
ning the po-
an certainly
ecific elec-
ctrode sited
plitude), but
om it. If one
on isolated
—

ro potential
graphy, but
can make a
alogous to a
chometrics
attitudes on
v positive to
ject   ses
ion in...cates
ng negative
ving the am-
t, such as by
nd negative
tude of zero
or can mean
een the pos-
ambiguity in
se, this point
scalp topog-
itatively dif-
vidual wave-

the average
s conclusion
drawback of
problem that
d, activity at
the recorded
n approaches
an empirical
ults.
ry critique—
ions of focal
990). When
p generates
api   om
n w..re the
a and appear

focal. In effect, the problem is that the focal side of the dipolar field may be oversampled when the focus corresponds exactly to a recording site and undersampled when it does not, compared with the more diffuse opposite side of the dipolar field. The former causes it to appear to be attenuated (at a given electrode, as noted), the latter magnified. Again, however, a complementary problem applies to monopolar references. Potential fields will be distorted to the extent that they overlap the reference site. If the reference site and the recording site are on the same side of the field, the recorded field will be attenuated (since only the difference is recorded); if they are on opposite sides, the recorded field will be magnified (since the subtraction of opposites adds). Again, the question is therefore an empirical one about which produces more accurate results.

Effects of the Average Reference

Although one might argue that choice of reference makes little difference since references are readily interconverted, this choice can have dramatic effects on analysis. The initial analysis for any data set is simple description. Although lacking the rigor of statistical tests, it allows anomalies to be detected, unexpected effects to be recognized; it also guides the interpretive process. If the waveforms are distorted, this visual inspection may end up being misleading rather than informative. Some vigorous but misguided debates centered on interpreting the location of zero potential were carried out before the field came to understand the central role of the reference site.

For example, arguments were conducted over the generating source of the auditory potentials. An initial report with a coronal montage concluded that the auditory evoked potentials (the N1 and the P2) emanate from the superior temporal plane because polarity inversion (the zero potential line) was observed over this location (Vaughan & Ritter, 1970). In truth, this location was predetermined by the reference site because the authors used a nose reference and (for a vertical dipole that is somewhat tilted forward) the nose is at the same potential as the supratemporal plane. Another set of authors argued with this result because they did not observe any reversal using a sternovertebral reference (Kooi, Tipton, & Marshall, 1971); if they had used recording electrodes on the body, they might have observed that the zero point was below the brain entirely on level with the reference site! Today, due to converging evidence from intracranial and magnetic field recordings (Näätänen, 1992), there is little question that the bulk of the auditory N1 arises from the supratemporal plane.

As another example of how the reference site can affect interpretation of potentials, a face-sensitive P2 has been reported at midline electrodes that is not only larger to faces but also earlier (Jeffreys, 1989). The latency change has been proposed to be due to a slightly earlier N2 potential located at the reference site (at the mastoids; Bötzel, Schulze, & Stodieck, 1995); such a potential at the reference site would be superimposed on the P2 seen at the midline sites. According to this report, the P2 itself does not in fact show latency shifts. It should be noted that Jeffreys correctly commented on the ambiguities posed by reference site activity.

How Does Reference Affect Analyses?

The remainder of this paper will provide concrete illustrations of how reference issues can impact analyses, followed by recommendations for future research. Since this is a methodological paper, the potential impact of reference issues will be shown using two relatively straightforward components. In particular, this paper will focus on the auditory N1, a clearly defined component with a vertical orientation, and the slow wave, a distinctive component with an apparently horizontal orientation (N. K. Squires, K. C. Squires, & Hillyard, 1975).

Following the method section, this paper will be structured in discrete sections covering each point. The effects of reference on three of the most common analytical techniques (description, analysis of variance [ANOVA], and correlation) will be discussed. These illustrations will compare Cz, mean mastoids, and average reference. For vertically oriented fields, these place zero at the top, bottom, and middle of the head, respectively. Since the average reference is best used for high-density data, a 128-channel data set will be utilized. Finally, the average reference will be evaluated by examining how well it localizes the auditory N1, a component with a known source whose vertical orientation makes it a worst case scenario for the average reference (Desmedt et al., 1990).

METHOD

Subjects

Fourteen University of Illinois undergraduates were recruited from psychology classes and by flyers. Subjects received either credit toward meeting introductory psychology requirements or monetary compensation. One subject was unable to perform the task and 1 subject's head was too small for the electrode net. leaving 12 subjects (4 males and 8 females; mean age, 23 years).

Equipment

EEG data were collected from 128 recording sites, plus a Cz reference sensor, with the Geodesic Sensor Net (Tucker. 1993). The Geodesic Sensor Net is a tessellated icosahedron network of elastic bands holding in place a dense array of plastic tubes filled with sponge Ag/AgCl sensors. After amplification with a 0.1-Hz to 50-Hz (64 dB attenuation at 1.44 Fcutoff) bandpass, the signals were digitized with a 12-bit A/D converter at a rate of 250 samples/sec using a National Instruments NB-MIO16 analog to digital board. A Power Macintosh 8100/110 computer handled both data collection and stimulus display using EGIS (Electrophysiological Graphical Interface System).

Experimental Design and Procedures

This experiment (gathered for other purposes) was chosen for this exposition since it provides a 128-channel data set in which an N1 frequency effect and a late positive complex target effect are present. The experiment consisted of eight blocks of 250 (1.024-msec) trials each with an interstimulus interval of 1.3 sec (276-msec pause for disk writing). The stimuli were kindly provided by Clara Fitzgerald and Len Trejo. Stimuli were presented binaurally at 250, 1000, 2000, and 4000 Hz. The stimuli consisted of 71.2% 70-dB standards
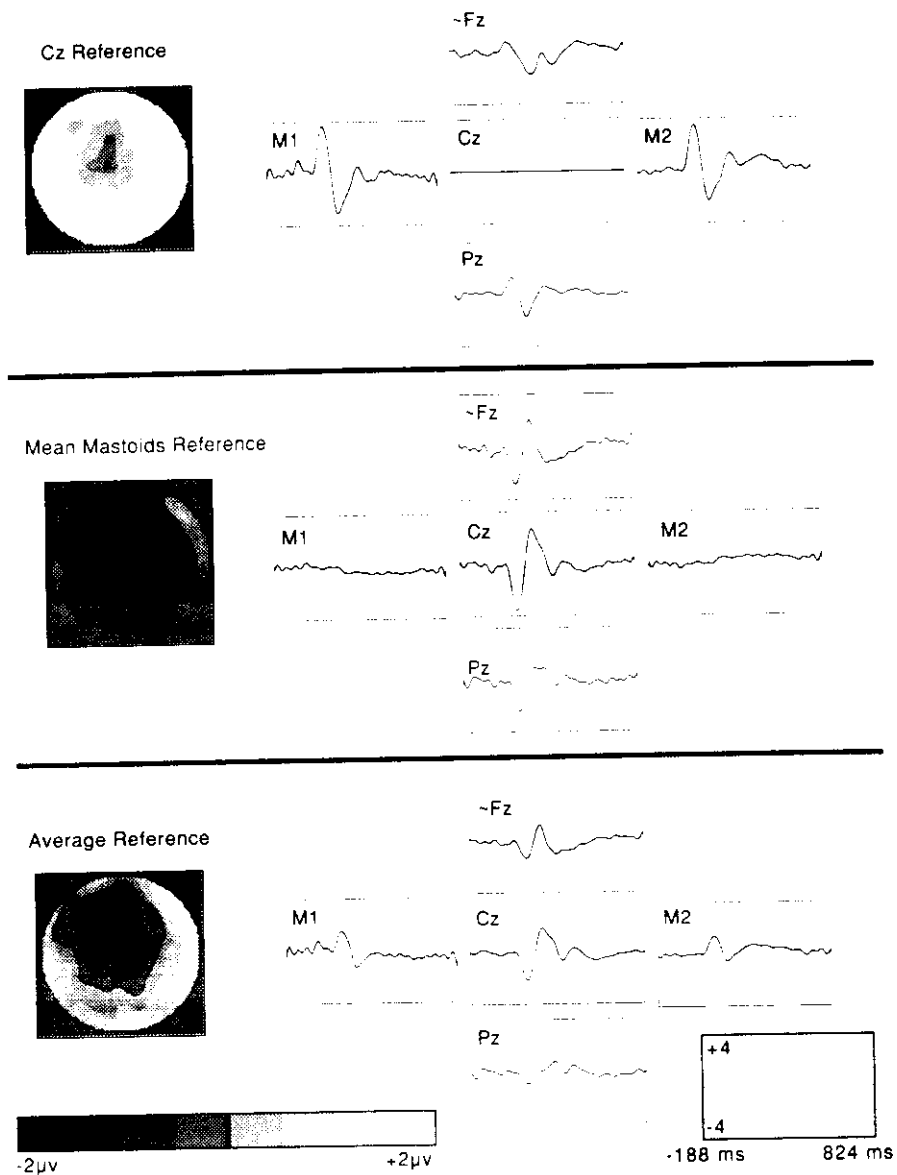
Figure 2. Effect of reference on N1 scalp topography. Grand average of 250-Hz standards, displayed using average reference, Cz reference, and mean mastoids reference. The 2-D figures on the left are at the height of the N1 at 136 msec and are plotted with front facing top. Black is most negative and white is most positive. In the waveforms at right, negative is plotted down.

and two 14.4% deviants, one at 62 dB and one at 78 dB (counterbalanced across subjects). Loudness of the stimuli was equated using Type A weighting, which takes nonlinearities in perceived loudness into account. The stimuli had a duration of 60 msec (18 msec rising and falling, 24-msec level). The frequencies were blocked and presentation order counterbalanced within the two task sections of the experiment.

The first four blocks were devoted to a passive auditory oddball task in which subjects were instructed to read materials of their own choice and to ignore the sounds. The second four blocks consisted of an active auditory oddball task in which subjects were told to press a button to infrequent target tones as quickly as possible while still being accurate. They were further told to ignore the standards

and nontarget infrequents. Subjects kept their gaze fixated upon a dot on a computer screen during the active task (114-cm distance).

Data Processing

Trials with eye blinks (75 μv) and eye artifacts (75-μv transitions in eye channels) were rejected. In addition, individual channels with sharp transitions of over 75 μv or absolute amplitudes of over 250 μv were dropped, and the remaining channels were retained. Trials with more than 10 bad channels were also dropped. An average of 96 trials (54%) were retained for analysis per standard stimulus cell (passive and active). A 20-Hz lowpass filter was then applied to the resulting average. Bad channels were identified, using a criterion that channels must correlate at least |.6| with any other two channels.
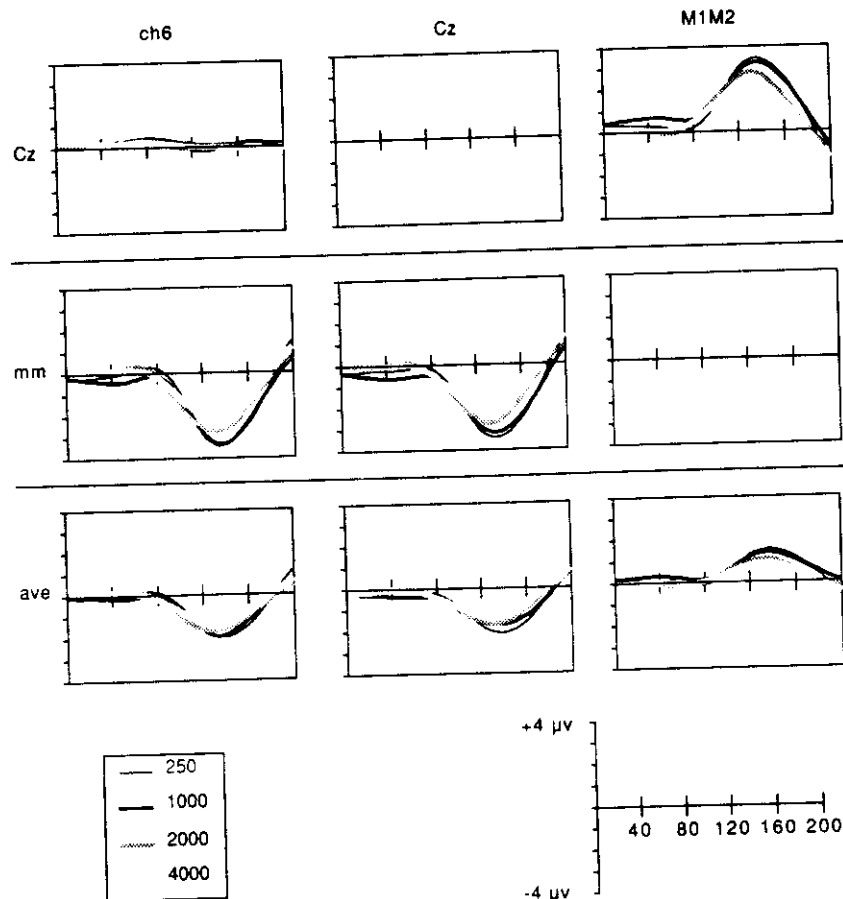
Figure 3. Effect of reference upon the N1 frequency effect. Waveforms show the N1 to the four different frequency standards (collapsed over task). Channels ~Fz (Channel 6), Cz, and the averaged mastoids (M1M2) are shown when referenced to Cz, mean mastoids, and average reference.

When a channel was marked bad in one cell, it was replaced in all cells of that subject to avoid confounds. In total, 104 (6.8%) channels were replaced.

Description

The start point of any analysis is simple description. The dramatic effects of the reference site on the resulting waveforms can be seen in Figure 2 (250-Hz standards arbitrarily chosen for illustration). Note how sites near the reference are reduced to near zero. Note also how meaningless the location is of the crossover point where a component goes from appearing to be negative to being positive. Only the average reference (to the extent that it is valid) has crossover points that correspond to "true" zero and can relate directly to localization.

Analysis of Variance

The most common inferential test for ERPs is the ANOVA, which can be strongly impacted by the choice of reference. By way of illustration, ANOVAs were conducted on an increase in the amplitude of the N1 to lower frequency stimuli (Figure 3). Parenthetically, it is not clear why this effect was found. Although the stimuli were equated for loudness, it is possible that the lower frequencies were still perceived as being subjectively louder. For the analysis, ~Fz (located in between Cz and Fz) was chosen as a site next to Cz where the N1 should be nearly maximal. Principal components analysis (PCA) (discussed later) was used to delimit the window (116–176 msec) and

the mean amplitude within the window was used as the dependent measure. The ANOVAs consisted of a frequency factor (250, 1000, 2000, and 4000 Hz).

As can be seen (Table 1), the significance level of the effect depended strongly on the choice of reference, with statistical significance being found only with the mean mastoid reference. While this might be seen as an endorsement for the use of mean mastoids reference, consideration of the underlying reasons suggests a more measured conclusion.

Table 1
N1 Frequency Effect at ~Fz

Reference	Effect	df	SS	MS	F	p	G-G	ε
MM	stim	3	1.137	.379	3.433	.0280	.0449	.732
		33	3.642	.110				
Cz	stim	3	.052	.017	.773	.517	.490	.778
		33	.739	.022				
Ave	stim	3	.217	.072	2.103	.119	.141	.721
		33	1.136	.034				

Note—The ANOVA tables provide degrees of freedom (df), sum of squares (SS), and mean sum of squares (MS) for the numerator (first row) and denominator (second row) for each F test. Also presented are the F ratios, the corresponding p value, the Greenhouse–Geiser corrected p value (G–G), and the Greenhouse–Geisser epsilon value (ϵ).

Table 2
N1 Frequency Effect at ~Fz and Mean
Mastoids Using Mean Mastoid Reference

Reference	Effect	df	SS	MS	F	p	G–G	ϵ
MM	stim	3	.568	.189	3.433	.0280	.0449	.732
		33	1.821	.055				
MM	chan	1	20.235	20.235	38.005	.0001	.0001	1.000
		11	5.857	.532				
MM	stim × chan	3	.568	.189	3.433	.0280	.0449	.732
		33	1.821	.055				

The effect of rereferencing is to reallocate potentials across the head (not to add or eliminate them). Changing the reference from the original Cz site to mean mastoids has the effect of adding the activity previously attributed to the mastoids to all the other sites. This can be expected to reduce the test statistic by adding error variance emanating from the mastoid sites to the ~Fz site. This reduction is counterbalanced since the error variance due to Cz (which is also present in the original Cz referenced mastoid recording) is canceled out during this procedure. Most importantly, the test statistic is modified to the extent that the reallocated potentials also contain a condition difference, thus adding to the between-group variance. The final effect on the test statistic of rereferencing therefore depends on the resultant of these three competing influences. The manner in which the condition differences are shuffled among the different channels can be seen in Figure 3.

In this case, the power of the statistical test was increased because the mastoids lie at the other end of the dipolar field from ~Fz. In effect, this test constituted a planned comparison between the mastoid site and the vertex, which for the auditory N1 is the optimal comparison since it is a vertically oriented field. The reason such a comparison is optimal is because the absolute condition difference increases with the magnitude of the measured component. If a measure of ~Fz alone (against the hypothetical true zero point) provides amplitudes of $-2\ \mu v$ in Condition 1 and $-3\ \mu v$ in Condition 2, the condition difference is 1 μv; if the mastoid site has amplitudes of $+2\ \mu v$ and $+3\ \mu v$, the contrast between ~Fz and the mastoids is 4 and 6 μv, respectively, resulting in a condition difference of 2 μv. The between-group variance will have been doubled (and squared to four times) whereas the error variance will not necessarily increase as much.

That the advantage of the mean mastoids test is a function of the implicit comparison rather than the reference per se can be readily demonstrated. If the mean of the mastoids (zero, since they serve as the reference) is included in the ANOVA via a channel factor, it can be seen that the stim effect is unchanged by this zero variance addition (Table 2). Moreover, the new stim × chan interaction is identical to the stim main effect. On the other hand, if the mean mastoids are added to the average reference ANOVA (where they are decidedly nonzero), the result of the stim × chan interaction is essentially the same as that of the mean mastoids ANOVA (Table 3). There is no main effect for stim since the two sites go in opposite directions (~Fz goes negative and the mastoids go positive), canceling each other out. Similar results are seen for the Cz reference when the mastoids are included (Table 4).

It can be seen that the effect of the reference choice is wholly dependent on what channels are included in the ANOVA. For sparse arrays (such as Fz, Cz, and Pz), in which all the sites are usually included in the ANOVA, reference choice should make no difference. Reference issues become notable only when too many channels are used to include all of them in the analysis. For optimal statistical power, an ANOVA should include both sides of the dipolar field, either by explicitly including channels from both sides or by implicitly including the far side via the reference site. Sites that contain the component to a smaller extent should be excluded since they will only dilute the condition difference and complicate the analysis. Such channel selection would have to be made in a principled fashion, of course, to avoid capitalizing on chance variation. Following this logic, if mean mastoids reference is used for visual components (which are often posterior and horizontally oriented), a frontal comparison site might improve the statistical power of an ANOVA.

At any rate, this exposition makes it clear that reference site must be taken into account when attempting to replicate an effect. Even a perfect replication of the data may yield a different pattern of statistical significance if a different reference scheme is used without consideration of the channels included in the analysis.

Correlation

Correlations are also sometimes used as an adjunct measure to ANOVAs. An example of how reference issues can impact correlation analysis concerns a late positivity called the slow wave (N. K. Squires et al., 1975), behavior of which is quite similar to that of the P300 in that it is larger to rare targets, although there are differences; for example, uncertainty about the stimulus causes the P300 to be weaker but the slow wave to be stronger (Ruchkin & Sutton, 1983). This component has a frontal negative, posterior positive distribution suggesting a nearly horizontal orientation for the generator(s). An influential study reported (with 5 subjects) that the frontal negative and posterior positive regions are substantially uncorrelated at the single-trial level (Loveless, Simpson, & Näätänen, 1987), using a mean mastoid reference, concluding that these regions may reflect separate components. Although this finding would appear to be definitive, use of a different reference might have produced a different result.

Although the present data set is not well suited for eliciting a large slow wave, Subject 10 displayed a slow wave that is reasonably distinct from the P300, particularly in the 250-Hz condition (Figure 4). Following the same procedure as Loveless et al. (1987), the slow

Table 3
N1 Frequency Effect at ~Fz and Mean
Mastoids Using Average Reference

Reference	Effect	df	SS	MS	F	p	G–G	ϵ
Ave	stim	3	.031	.010	.614	.6108	.5932	.890
		33	.556	.017				
Ave	chan	1	20.217	20.217	37.794	.0001	.0001	1.000
		11	5.884	.535				
Ave	stim × chan	3	.561	.187	3.383	.0296	.0468	.732
		33	1.826	.055				

Table 4
N1 Frequency Effect at ~Fz and
Mean Mastoids Using Cz Reference

Reference	Effect	df	SS	MS	F	p	G-G	ε
Cz	stim	3	1.889	.630	5.598	.0032	.0056	.847
		33	3.711	.112				
Cz	chan	1	40.489	40.489	37.991	.0001	.0001	1.000
		11	11.723	1.066				
Cz	stim × chan	3	1.119	.373	3.372	.0299	.0471	.733
		33	3.651	.111				

wave was measured using the baseline-corrected mean amplitude over its duration (in this case, 500–800 msec). The correlation was computed between Pz (Channel 62) and ~Fz (Channel 6). Using the Cz reference, the result is $R(34) = -.427, p < .01$. Contrary to their conclusion, this would appear to indicate that these two regions are largely inversely correlated, as one might expect from the opposite sides of the same dipole. However, with mean mastoid reference (the same as that used by Loveless et al.), the result is $R(34) = .755, p < .0001$. With an average reference the result is $R(34) = -.0920, p = .594$.

What is going on? Essentially, the problem is that the reference can artificially inflate or deflate the correlation. Activity at the reference site unrelated to these two channels is attributed to them both, adding an artificial source of positive correlation (in this case, pushing the correlation toward unity). Contrariwise, to the extent that the two sites share activity with the reference (such as the neighboring Cz site), this activity will be removed, attenuating the positive correlation (resulting in a negative correlation with the Cz reference, albeit involving small residual amplitudes).

Although the ambiguity introduced by the reference issue clouds the Loveless et al. (1987) analysis, some interpretation is possible. In principle, since it has been shown that the mastoids are far from inactive (Lehtonen & Koivikko, 1971), it is safe to assume that some amount of positive correlation between the two sites in Loveless et al.'s data was due to reference activity. Given that they obtained correlations of about −.3 even with this added positive correlation, it seems likely that the two regions are indeed substantially inversely

correlated, presumably due to the slow wave. Of course, it would be rash to deny that other components might also be present that could dissociate these measures at least to some degree.

EVALUATING THE AVERAGE REFERENCE

Given the potential utility of a reference solution that avoids reference site artifacts, the next question is whether the average reference provides an answer. Although the simulations presented by Desmedt et al. (1990) were helpful in clarifying the issues, such simulations cannot ultimately evaluate the average reference until a head model can be developed that takes the neck and lower face into account (along with appropriate conduction values). For now, this paper will provide a simple empirical test of the average reference by evaluating whether the average reference places the zero potential of the auditory N1 at a level compatible with its known generator (Näätänen, 1992).

N1 Localization

As noted earlier, Desmedt and colleagues (1990) have argued that since the underside of the brain is insufficiently sampled, it will cause potential fields to be shifted
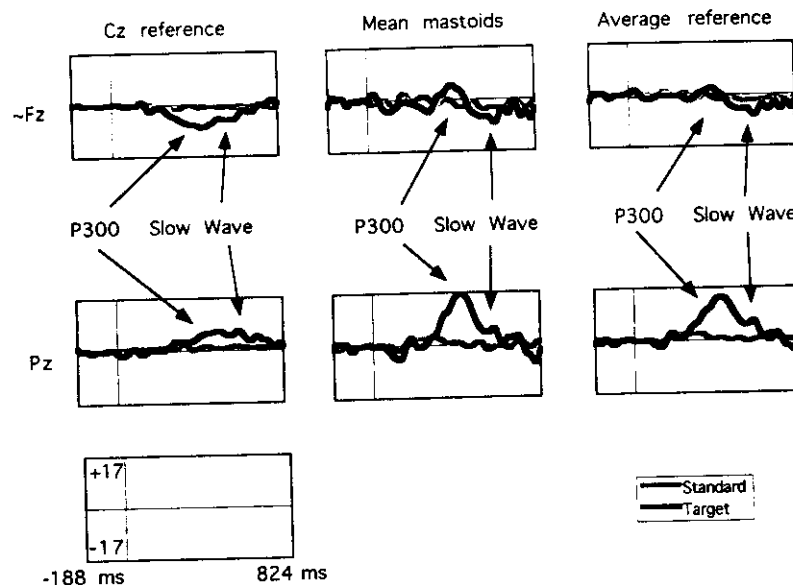

Figure 4. Effects of reference upon the slow wave topography. Recordings at ~Fz and Pz of Subject 10 under Cz, mean mastoids, and average reference. Negative down. Dark line is response to targets; light line is response to standards.

T3/T4

C3/C4

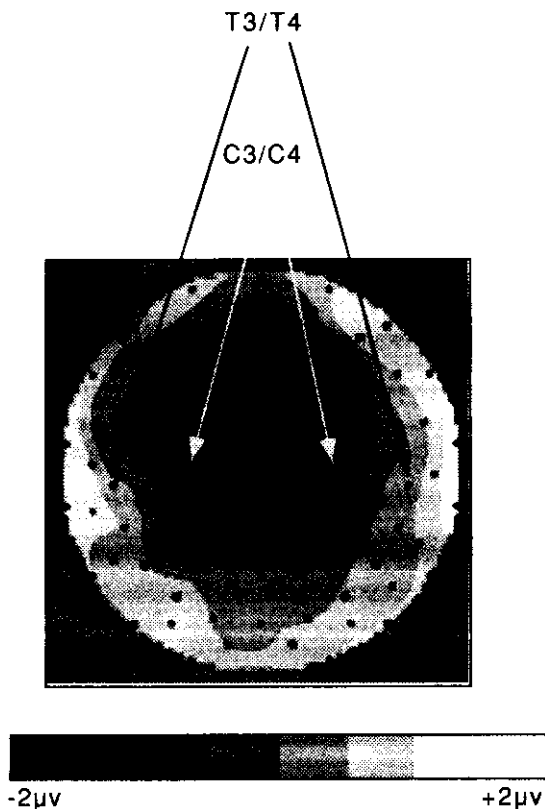-2µv   +2µv

Figure 5. Scalp topography of the N1 factor. Locations of the 129 electrode sites are indicated by dots. The locations of T3/T4 and C3/C4 are indicated by arrows. The zero potential line is indicated by the thin black line.

upward, producing "ghost fields." This effect should be especially pronounced for dipoles that are oriented vertically, like that of the auditory N1. To the extent that the N1 generators are oriented upright, it is expected that the zero potential region with respect to infinity should lie over the generating site in the supratemporal plane. The average reference can be evaluated by how well it approximates this topography.

In order to properly characterize the scalp topography of the N1, it is necessary to first separate it from overlapping potentials such as the processing negativity. A PCA was conducted on the first 200 msec using a covariance matrix and promax rotation. Following a scree test (Cattell, 1966; Cattell & Jaspers, 1967), 10 factors representing 99.99% of the variance were retained. Only the first factor had the peak latency and scalp topography of the N1. If one multiplies the factor loadings by the factor scores and the standard deviations of the time points, one reproduces the portion of the data set that is accounted for by the factor (Dien, 1997).

CT scans of 10–20 locations in a dozen subjects (Homan, Herman, & Purdy, 1987) have suggested that the supratemporal planes lie just dorsal to T3 and T4, perhaps a third

of the way to C3 and C4. Examination of the factor score plots indicates that the N1 zero potential line does indeed invert in this general region (Figure 5). Although there is some evidence of upward shifting (it inverts about halfway to C3/C4 rather than a third of the way), it appears that the effect is modest. Although this test is necessarily rather crude, it suggests that even for a worst case vertically oriented dipole, the mischaracterization of the potential field will be modest when a comprehensive electrode montage is used. Note that this conclusion is limited to this particular electrode montage and that this test would have to be repeated for other electrode montages.

RECOMMENDATIONS

The optimal choice of reference site depends on the study and the purpose of the analysis. In some cases the use of average reference is clearly not indicated. To be reasonably reliable and accurate, it requires a comprehensive high-density montage; in the absence of such a montage, a conventional reference technique is necessary. Standard reference sites also facilitate comparisons with previous studies and across differing recording montage. In such cases, it may be best to use a standard common reference for ANOVAs (with appropriate pairs of electrode sites representing opposite sides of the potential field, with perhaps additional electrodes to detect laterality effects).

Recommendations for analyses on high-density montages are less clear-cut. Direct examinations of topography should utilize average reference, such as when plotting interpolated scalp fields, to reduce the substantial distortions produced by a common reference. Given the qualitative nature of such interpolations, variations due to montage differences should be relatively minor. Use of the average reference will also facilitate comparison with dipole localization packages like BESA, which automatically performs an average reference transform on data sets, since it is not biophysically plausible to define a reference site on the head as being inactive.

Both average reference or a common reference may be appropriate for the remainder of such an analysis. Correlation-based analyses should use the average reference if possible to avoid artificially inflated correlations due to reference site activity. When comparison with prior data sets is a primary concern, waveforms should be provided using the same reference type (such as mean mastoids). If comparison with topographical maps is of primary concern, then average reference should be used for waveforms. It may be necessary to provide waveforms using both references in some cases. ANOVAs should have the same reference as used in the plotted waveforms.

CONCLUSION

This paper has demonstrated that choice of reference has substantial effects on analysis and interpretation. Analyses of sparse montages should consider possible effects of the

:tor score
es   d
h the.e is
t halfway
:s that the
ily rather
:cally ori-
itial field
montage
iis partic-
iave to be

ds on the
cases the
To be rea-
rehensive
montage,
Standard
previous
. In such
reference
sites rep-
with per-
ects).
iity mon-
pograohy
ott   i-
al distor-
e qualita-
montage
e average
ole local-
performs
e it is not
i the head

iay be ap-
rrelation-
if possi-
s due to
rior data
provided
stoids). If
iary con-
iveforms.
both ref-
the same

rence has
Analyses
cts of the

reference site and should always indicate and justify reference site choice. Localization analyses of high-density montages (based on dipole modeling and scalp topography) should utilize average reference where possible.

REFERENCES

ANDINO, S. L. G., MARQUI, R. D. P., SOSA, P. A. V., LIRIO, R. B., MACHADO, C., DIAZ, G., RODRIGUEZ, P. F., & TORREZ, C. C. (1990). Brain electrical field measurements unaffected by linked earlobes reference. *Electroencephalography & Clinical Neurophysiology*, 75, 155-160.

BERTRAND, O., PERRIN, F., & PERNIER, J. (1985). A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalography & Clinical Neurophysiology*, 62, 462-464.

BÖTZEL, K., SCHULZE, S., & STODIECK, S. R. G. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental Brain Research*, 104, 135-143.

CATTELL, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

CATTELL, R. B., & JASPERS, J. (1967). A general plasmode (No. 3010-5-2) for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, No. 67-3, pp. 1-212.

DESMEDT, J. E., CHALKLIN, V., & TOMBERG, C. (1990). Emulation of somatosensory evoked potential (SEP) components with the 3-shell head model and the problem of "ghost potential fields" when using an average reference in brain mapping. *Electroencephalography & Clinical Neurophysiology*, 77, 243-258.

DIEN, J. (1997). Addressing misallocation of variance in principal components analysis of evoked potentials. *Psychophysiology*, 34 (Suppl. 1), S31.

HOMAN, R. W., HERMAN, J., & PURDY, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalography & Clinical Neurophysiology*, 66, 376-382.

JEFFREYS, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, 78, 193-202.

KATZNELSON, R. D. (1981). EEG recording, electrode placement, and aspects of generator localization. In P. L. Nunez (Ed.), *Electric fields of the brain: The neurophysics of EEG* (pp. 176-213). New York: Oxford University Press.

KOOI, K. A., TIPTON, A. C., & MARSHALL, R. E. (1971). Polarities and field configurations of the vertex components of the human auditory evoked response: A reinterpretation. *Electroencephalography & Clinical Neurophysiology*, 31, 166-169.

LEHTONEN, J. B., & KOIVIKKO, M. J. (1971). The use of a non-cephalic reference electrode in recording cerebral evoked potentials in man. *Electroencephalography & Clinical Neurophysiology*, 31, 154-156.

LOVELESS, N. E., SIMPSON, M., & NÄÄTÄNEN, R. (1987). Frontal negative and parietal positive components of the slow wave dissociated. *Psychophysiology*, 24, 340-345.

MILLER, G. A., LUTZENBERGER, W., & ELBERT, T. (1991). The linked-reference issue in EEG and ERP recording. *Journal of Psychophysiology*, 5, 273-276.

NÄÄTÄNEN, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

NUNEZ, P. L. (1991). Comments on the paper by Miller, Lutzenberger and Elbert. *Journal of Psychophysiology*, 5, 279-280.

OFFNER, F. F. (1950). The EEG as potential mapping: The value of the average monopolar reference. *Electroencephalography & Clinical Neurophysiology*, 2, 215-216.

OSSELTON, J. W. (1965). Acquisition of EEG data by bipolar, unipolar and average reference methods: A theoretical comparison. *Electroencephalography & Clinical Neurophysiology*, 19, 527-528.

PERRIN, F., PERNIER, J., BERTRAND, D., & ECHALLIER, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography & Clinical Neurophysiology*, 72, 184-187.

RUCHKIN, D. S., & SUTTON, S. (1983). Positive slow wave and P300: Association and dissociation. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in ERP research: Endogenous components* (pp. 233-250). Amsterdam: North-Holland.

SQUIRES, N. K., SQUIRES, K. C., & HILLYARD, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography & Clinical Neurophysiology*, 38, 387-401.

SRINIVASAN, R., TUCKER, D. M., & MURIAS, M. (1998). Estimating the spatial Nyquist of the human EEG. *Behavior Research Methods, Instruments, & Computers*, 30, 8-19.

TOMBERG, C., NOEL, P., OZAKI, I., & DESMEDT, J. E. (1990). Inadequacy of the average reference for the topographic mapping of focal enhancements of brain potentials. *Electroencephalography & Clinical Neurophysiology*, 77, 259-265.

TUCKER, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography & Clinical Neurophysiology*, 87, 154-163.

VAUGHAN, H., & RITTER, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography & Clinical Neurophysiology*, 28, 360-367.

WOLPAW, J. R., & WOOD, C. W. (1982). Scalp distributions of human auditory evoked potentials: 1. Evaluation of reference electrode sites. *Electroencephalography & Clinical Neurophysiology*, 54, 15-24.