Psychophysiology, 39 (2002), 791–808. Cambridge University Press. Printed in the USA.

DOI: 10.1017/S0048577202010648

The intractability of scaling scalp distributions to infer neuroelectric sources

THOMAS P. URBACH[a] AND MARTA KUTAS[a,b]

[a]Department of Cognitive Science, University of California, San Diego, La Jolla, California, USA
[b]Department of Neurosciences, University of California, San Diego, La Jolla, California, USA

Abstract

ERP researchers use differences in scalp distributions to infer differences in spatial configurations of neuroelectric generators. Since McCarthy and Wood (1985) demonstrated that a spatially fixed current source varying only in strength can yield a significant Condition × Electrode interaction in ANOVA, the recommended approach has been to normalize ERP amplitudes, for example, by vector length, prior to testing for interactions. The assumptions of this procedure are examined and it is shown via simulations that this application of vector scaling is both conceptually flawed and unsound in experimental practice. Because different spatial configurations of neural generators cannot reliably be inferred from different scalp topographies even after amplitude normalization, it is recommended that the procedure no longer be used for this purpose.

Descriptors: Event-related potential, ERP, Amplitude normalization, Topography, Scalp distribution, Source configuration

The ability to detect differences in the spatial distribution of cortically generated scalp potentials is a cardinal virtue of multichannel EEG recordings. In addition, for experiments designed to engage different brain systems, the reliable identification of distinct spatial configurations of neural generators may be a central concern. Differences between distributions of scalp potentials are typically established by the finding of a statistically significant interaction between Experimental Condition and Electrode Position in a repeated measures analysis of variance (ANOVA). However, the inference from a reliable topographic difference in surface potentials to conclusions about the specific type of differences in the neural generators is problematic. In an influential paper, McCarthy and Wood (1985) showed that a Condition × Electrode interaction alone is not sufficient grounds for inferring that the spatial configurations of generators in the two conditions differ, because such an interaction can result when a dipolar generator in a fixed spatial location varies only in strength. To protect against drawing this unwarranted conclusion, McCarthy and Wood proposed a vector scaling procedure that normalizes the overall amplitude of the distribution while preserving its topographic shape. Amplitude normalization procedures have since come into wide use and are explicitly recommended in published guidelines on ERP research (Picton et al., 2000) for purposes of identifying distinct source configurations. This report reviews amplitude normalization and several key concepts related to these inferential issues and then argues that amplitude normalization is unreliable in its intended application in ERP research and that use of the procedure should therefore be discontinued.

This research was supported by NICHD grant HD22614 and NIA grant AG08313 to M.K. This work has benefited from discussions with Seana Coulson, Kara Federmeier, Steve Hillyard, Sabine Windmann, and Rina Schul with special thanks to Esmeralda De Ochoa for her help preparing the manuscript.

Address reprint requests to: Thomas P. Urbach, Department of Cognitive Science, 0515, University of California at San Diego, La Jolla, CA 92093-0515, USA. E-mail: turbach@cogsci.ucsd.edu.

Amplitude Normalization and Generator Distributions

Despite some recent debate about amplitude normalization (Haig, Gordon, & Hook, 1997; Ruchkin, Johnson, & Friedman, 1999), a thorough exposition of the procedure's motivation, justification, and consequences for EEG research has not appeared in the literature. A number of procedures are plausible candidates for amplitude normalization, some are equivalent, others are not (see Appendix). Furthermore, it is not entirely clear that terms like the "strength" and "spatial configuration" of neural generators and the "topography" and "topographic shape" of distributions of potentials are used consistently.

Generators, Strengths, and Spatial Configurations

Idealized distributions of neural generators may be construed as sets of point current sources and sinks of specified intensity (strength) and polarity sprinkled throughout a volume conductor. The electrical field associated with these generators propagates as a function of distance and the geometry and electrical properties of the media, for example, cerebrospinal fluid, skull, and scalp. Although the potential associated with each generator varies with distance in a nonlinear manner, the field at any point, including the surface of a bounded volume conductor, varies directly with the intensity of the source. The field for multiple sinks and sources is the sum of the fields associated with each generator individually; some simple examples of distributions of generators in a two-dimensional homogenous conductor and their associated fields are illustrated in the two left columns of Figure 1.

Although the difference between strength and spatial configuration may seem straightforward, some simple examples illustrate how matters can become murky, particularly when strength is contrasted with spatial configuration. For instance, it makes a difference whether terms like “generator” or “source” refer to point sources, or to dipolar source–sink pairs in particular, or to some other ensemble of current sources and sinks. In some cases, the intended usage may be clarified by context, for example, an assertion about the orientation of a generator presupposes an axis of orientation that a point source does not have. In other cases, however, the meaning may be less clear and, of course, whether or not generator locations do, in fact, differ may depend on what is meant by “generator.” If “generator” refers to a dipole, there is a sense in which merely rotating the dipole 90 degrees does not change the location of the generator. However, if the positive and negative poles are treated as a separate point source and point sink, rotating the dipole does entail changes in generator locations. Similar concerns arise in connection with what is meant by the “strength” of a source. It seems clear enough that multiplying the intensity of the positive and negative poles of a dipole by a factor of two is a change in strength alone and not a change in spatial configuration. However, when the poles are multiplied by a factor of -1, this change in strength is equivalent to rotating the dipole 180 degrees. If rotation counts as a change in the spatial configuration of sources, then so must the equivalent polarity reversal;

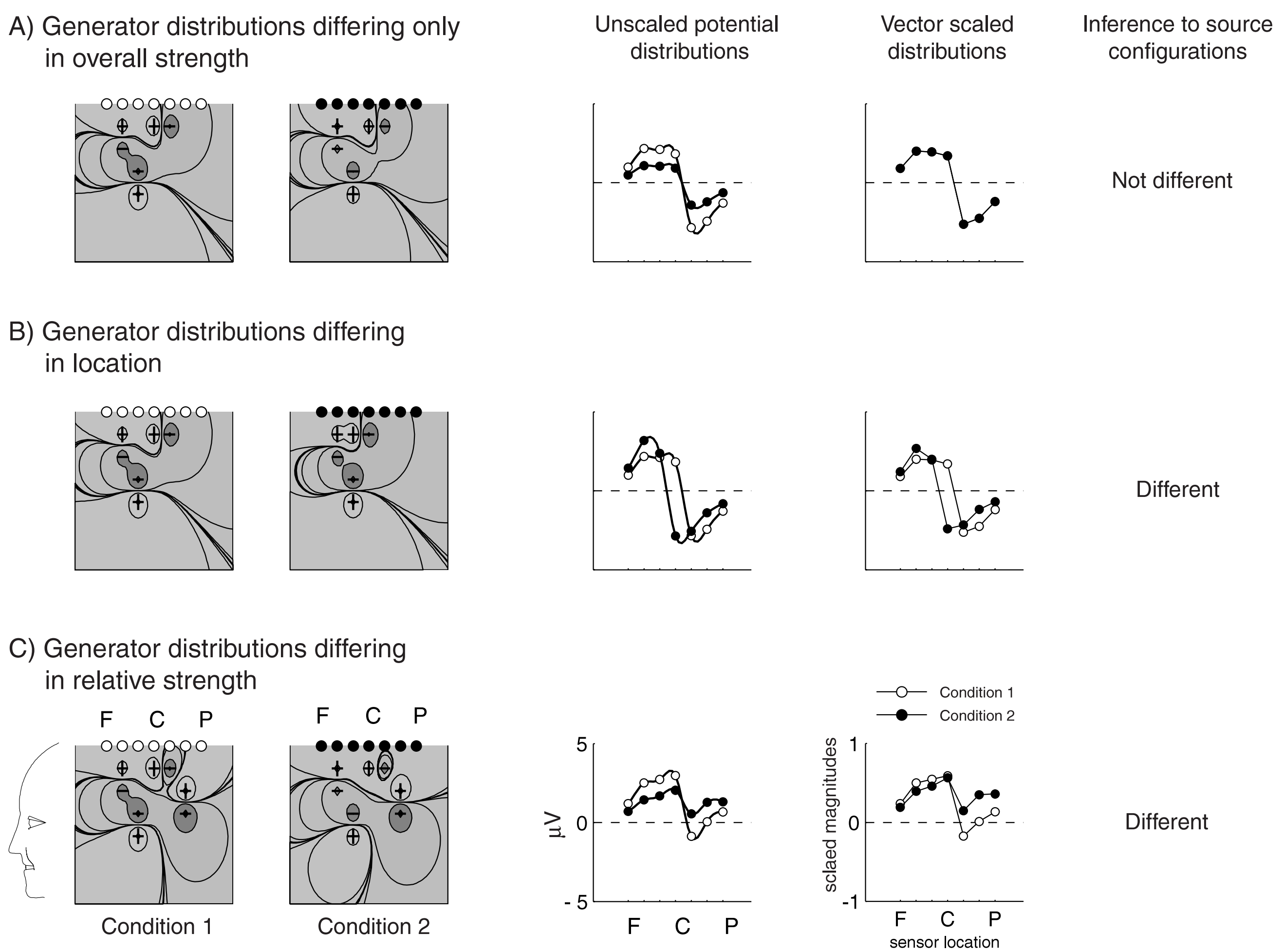


Figure 1. Schematic illustration of the ambiguity of Condition × Electrode interaction effects. The two leftmost columns show generator distributions in two experimental conditions. Positive potentials generated by current sources (+) are shaded lighter, negative potentials generated by sinks (−) are shaded darker, and isopotential contour lines mark orders of magnitude. For simplicity, potentials are assumed to vary with the inverse of the distance from the source. The rows of large open (Condition 1) and filled (Condition 2) circles indicate the “electrode” locations. The smaller open and filled circles in the line graphs indicate the values of the field at the corresponding electrode locations for the Condition 1 and Condition 2 generator distributions. A: Generator distributions where the locations and polarities are identical and each of the corresponding generators differ in strength between Conditions 1 and 2 by the same scalar multiple, in this case a factor of 2. B: Generator distributions where the polarities and strengths are identical, but the location of one of the dipoles is different. C: Generator distributions where the locations and polarities are identical but the generators differ in relative strength.

the result is both a difference in source strength and a difference in spatial configuration. Other examples similarly strain the putatively obvious dichotomy between strength and spatial configuration as well. Consider a strong dipole and a weak dipole some distance apart in a volume conductor and, without any change in location, suppose that the strength of the first dipole decrease while the strength of the second increases. The upshot of these changes in strength is that the strong and weak dipoles effectively trade places. Depending on what one means by "spatial configuration," it might seem reasonable to treat this as a different spatial configuration of generators. However, because neither the locations nor polarities of any of the four poles have changed whatsoever, this difference is also a difference in the strength alone. Here again, it seems that the distinction between differences in strength and differences in spatial configuration, if there is one, is not entirely clear and thus not especially informative. The following terms and definitions provide a framework for articulating exactly what is happening in these and other cases.

The term "generator" will refer to a point current source at a location with a polarity and a positive, nonzero intensity[1] and be represented as an ordered triple, that is, a generator $g = \langle L, p, i \rangle$ where the location vector L gives the $\langle x, y, z \rangle$ coordinates in space, polarity $p = +1$ or -1, and intensity i is a real number strictly >0. Requiring the intensity to be nonzero prevents phantom generators that do not generate a field. A distribution of generators G in a volume conductor is defined as a set of generators, and two such distributions G_1 and G_2 are identical if and only if every generator has the same location, polarity, and intensity in G_1 as it does in G_2. This regimentation segregates the three ways that distributions of generators may differ, that is, with respect to the location, polarity, or intensity of their constituent generators. When the intensity of the generators in G_1 differs from the intensity of the generators in G_2, there are two mutually exclusive cases: either the intensities in G_1 all differ by the same multiplicative factor from the corresponding intensities in G_2 or they do not. This differentiation can be incorporated into an explicit characterization of conditions under which generator distributions G_1 and G_2 differ as follows:

$G_1 \neq G_2$ iff

1. The locations of the generators are not all the same OR
2. The polarities of the generators are not all the same OR
3. The intensities of the generators differ
   a. in overall strength, that is, differ such that there is a single factor $m > 0$, $m \neq 1$, where, for each generator, $i_{G_1} = m i_{G_2}$

   OR (exclusive)

   b. in relative strength, that is, the strengths of the generators in G_1 differ by different factors for different generators such that there is no single factor m as defined in a.

Differences in overall strength are analogous to having all generators ganged to a master gain control: Turning the gain up or down multiplies the strength of each generator in the distribution by the same factor. Conditions 3a and 3b are mutually exclusive, that is, any difference in generator intensity is a difference in overall strength or a difference in relative strength but not both. Conditions 1, 2, and 3 are not mutually exclusive and differences in location, polarity, and strength can co-occur.

The relation between differences in the strength and differences in the spatial configuration of generators may be articulated in this framework. The key relation involves two generator distributions that differ *only* in overall strength and this case will be termed "multiplicatively related," that is:

Two generator distributions G_1, G_2, are multiplicatively related iff

1. The locations of the generators are all the same AND
2. The polarities of the generators are all the same AND
3. The intensities of the generators differ in overall strength, that is, 3a above is satisfied.

In this special case, the only difference between G_1 and G_2 is a difference in the overall strength of the generators, that is, the locations and polarities are all identical and there are no relative differences in strength. The identity conditions for "spatial configurations of generators" may, at last, be given as follows:

Two generator distributions G_1, G_2 have the same spatial configuration of generators iff G_1 and G_2 are multiplicatively related.

This definition of spatial configuration may be illustrated in application to the examples introduced above. Rotating a dipole by 90 degrees counts as a change in spatial configuration. Since rotation changes the location of the poles, the two distributions are not multiplicatively related and, thus, the spatial configurations of the generators differ. For the polarity reversal case, changing the polarity of the poles is not a difference in strength as defined above, so the generator distributions are not multiplicatively related. This case also counts as a difference in spatial configuration. Finally, the generator distributions where the strong and weak dipoles trade locations (strengths?) are also different spatial configurations. Even though there is no difference in generator location or polarity, the intensity of the first dipole changes by some factor <1 and the intensity of the second dipole changes by some factor >1. Because this is a difference in relative strength and not overall strength, the distributions are not multiplicatively related, and consequently, the two spatial configurations of generators differ.

It may not be immediately obvious that being multiplicatively related and having the same spatial configuration are or should be treated as equivalent, because it might be argued that a spatial configuration of generators should be defined by the location of the generators alone, regardless of generator strength. . Of course, "spatial configuration of generators" has been defined by stipulation, and other definitions are possible, but the definitions above have two salient virtues. First, they give a principled way to characterize exactly what is going on in the troublesome cases described at the outset. Second and more importantly, these definitions explicitly characterize the only sense of "different spatial configurations of generators" in which it is true that distributional differences that remain after normalizing the amplitude of scalp potentials entails that the spatial configurations of the correspond-

[1]It is possible to fold polarity in with intensity by allowing intensity to range over positive and negative nonzero numbers. In principle, the substantive points can be developed either way, but treating polarity and intensity separately will make for a tidy separation of "strength" and "spatial configuration."

ing generators are different. Thus, although other definitions might be imagined, the definition of "spatial configuration of generators" articulated herein must be the one assumed in all previous papers that use vector scaling, if amplitude normalization is to establish that spatial configurations of generators differs.

Scalp Potential Distributions and Topographic Shape

The potential field at the surface of a bounded volume conductor may be spatially sampled, for example, by electrodes, at discrete locations. The values at these locations define a distribution of potentials that may be represented as a vector $S = \langle v_1 \ldots v_j \ldots v_a \rangle$, $j = 1, 2, \ldots, a$, where v_j is the potential at location j, and a is the number of electrodes. Assuming the same sensor locations, identity for two such distributions S_1 and S_2 is simply identity of the potentials at each location; distributions differ if they are not the same at every location.

The distribution of surface potentials in this sense must be clearly distinguished from the topographic shape and overall amplitude of such distributions. Whereas the distribution is given by the numerical magnitudes of the potentials at each location, the topographic shape of a distribution is determined by the relative magnitudes of the potentials across all locations. For instance, a distribution of surface potentials $S_1 = \langle 2, 3, 5 \rangle$ is different from $S_2 = \langle 4, 6, 10 \rangle$ at all three locations, but S_1 and S_2 have the same shape (topography) because their internal proportions are the same, that is, for S_1 and S_2, $v_1/v_2 = 2/3 = 4/6$, $v_1/v_3 = 2/5 = 4/10$, and $v_2/v_3 = 3/5 = 6/10$. Topographic shape and overall size behave like familiar notions of geometric shape and size. If the sides of a triangle with lengths 2, 3, and 5 are multiplied by a factor of 2, the overall size of the triangle changes but its shape is the same because its original proportions are preserved. If the sides are not all multiplied by the same factor, the original proportions are not preserved and the geometric shape changes. For topographic shape, identity of internal proportions may conveniently be expressed in terms of scalar multiplication of the distributions (vectors):

> The topographic shape of distributions of surface potentials S_1 and S_2 is the same iff there is some scalar multiple m such that $S_1 = mS_2$.

In what follows, "topographic shape" and "topography" are used to refer exclusively to the shape of a distribution rather than the distribution itself. In this usage, the distributions S_1 and S_2 immediately above differ, but have the same topographic shape, that is, the same topography.

Generator Distributions and Surface Potential Distributions

For present purposes, there are two key inferential relations between distributions of generators and distributions of scalp potentials. The first may be expressed as follows:

> If two generator distributions G_1 and G_2 are the same, the corresponding distributions of surface potentials S_1 and S_2 are the same.

Thus, when two distributions of surface potentials differ, it follows that the corresponding generator distributions differ. The relation between spatial configurations of generators and topographic shape is as follows:

> If two generator distributions G_1 and G_2 have the same spatial configuration, the corresponding distributions of surface potentials S_1 and S_2 have the same topographic shape.

This relation follows from the definitions of spatial configuration, topographic shape and the fact that the potentials at any point vary directly with source strength. If the location and polarity of the generators are held constant and the strength differs by the same factor for all generators, that is, $i_{G_1} = mi_{G_2}$, the scalp distributions also differ at each point by this factor m. Thus, if proportional strengths of the generators are preserved, so are the proportional amplitudes at the scalp, that is, the scalp distributions have the same topographic shape. The form of this relation relevant for inferences from topographic shape to spatial configurations of generators follows immediately:

> If two distributions of surface potentials S_1 and S_2 do *not* have the same topographic shape, then the corresponding generator distributions G_1 and G_2 do *not* have the same spatial configuration.

Thus, in the theoretical ideal, differences in topographic shape permit valid inference to the conclusion that the spatial configurations of the generators differ *provided* that spatial configuration is defined as above. Whether or not the remaining explanations should all be classified as differences in the spatial configuration of generators is a semantic issue. This paper is agnostic on whether the spatial configuration nomenclature is appropriate and the aim here is just to clarify what does and does not follow about generator distributions from differences in surface potential distributions.

The Motivation for Amplitude Scaling of Scalp Potential Distributions

Conventional recordings of human scalp potentials with macroelectrodes have a temporal resolution on the order of a millisecond, a spatial resolution on the order of a centimeter at the scalp, and decades of research have consistently demonstrated their sensitivity to differences in perceptual and cognitive tasks. Any experimental measure with these properties would thus be of tremendous experimental value regardless of the physiological processes responsible for the effects. The fact that scalp potentials are generated by the electrochemical activity of the neural tissues that are actually doing the perception and cognition as they are doing it is an added bonus. Thus, in addition to their intrinsic value as a sensitive, noninterruptive, multidimensional, real-time measure, it is very tempting to draw inferences from scalp potentials to active brain areas, that is, neural generators. It is uncontroversial that in the absence of artifacts, a statistically significant difference between scalp distributions established by a Condition × Electrode interaction suffices to show that the distributions of neural generators differ: that is, as above, if $S_1 \neq S_2$, then $G_1 \neq G_2$. However, the fact that two generator distributions differ may involve some combination of differences in location, polarity, and strength; distributional differences in surface potentials show nothing more specific than this. Distributional differences in surface potentials may, for example, be the result of differences only in the overall strength of generators and there may be no difference whatsoever in their spatial configuration. This issue was first addressed by Hansen and Hillyard (1980) and, subsequently, by McCarthy and Wood (1985). The inferential problem is summarized in Figure 1, which schematically illustrates three ways in which distributional differences between conditions might be generated. In all three comparisons (Figure 1A, 1B, and 1C), the difference between the two conditions is larger at some "electrodes" than others and, with sufficient statistical power, all would yield significant Condition × Electrode interactions in an ANOVA. For the comparison in Figure 1A, the location and polarity of the generators is identical in both conditions and the two distributions differ only in overall strength. For the comparison in Figure 1B, the locations, hence spatial configuration, of the generators differs between conditions: The dipole closest to sensor C is shifted toward sensor F in Condition 2. In Figure 1C, a case that will be important for subsequent discussion, the location and polarity of the generators is the same in both conditions, but the generators differ in relative strength because some of them (both poles of the dipole farthest to the right) have the same strength in both conditions whereas the others change in strength. Because of this difference in relative strength, the spatial configurations of the generators differ even though their locations and polarity do not.

The difficulty inferring different spatial configurations of generators from distributional differences in scalp potentials arises because a Condition × Electrode interaction effect may be found in cases like those in Figure 1A where the generators differ only in overall strength, so clearly, this test alone cannot reliably distinguish such a case from cases where the spatial configurations actually do differ. Thus, where the research question requires that differences in overall generator strength be ruled out as an explanation of differences in scalp distributions, ANOVA conducted on scalp potentials is inadequate. To rule out differences in overall generator strength and thereby sharpen the conclusions drawn from Condition × Electrode interactions, McCarthy and Wood (1985) considered three analytic procedures and as a general solution, recommended normalizing the amplitude of the distribution of potentials in each experimental condition by vector length.

Before introducing the details of the vector scaling procedure, an important distinction should be emphasized. When comparing distributions of scalp potentials in different experimental conditions, two types of question must be clearly distinguished. First, do the distributions differ between the experimental conditions? Second, in what way or ways do the neural generators differ between the experimental conditions? For many research purposes, it suffices to determine whether there are reliable distributional differences at the scalp, and if so where they occur, that is, at which electrode locations. For instance, on most accounts, the distribution of an ERP effect over the scalp is criterial for component individuation. In oddball paradigms, for example, the P3b elicited by low probability targets and the P3a elicited by novel stimuli are distinguished in part by the centro-parietal maximum of the P3b and the frontal maximum of the P3a. Differentiating these components depends on establishing that there are reliable distributional differences, for example, by conducting a repeated-measures ANOVA with Conditions and Electrodes as factors. If the Condition × Electrode interaction effect is significant, it may be inferred that the difference between the conditions varies by electrode location, that is, the potentials are distributed over the scalp differently in the different conditions. The specific electrode locations responsible for this interaction can be identified by post hoc pairwise comparisons between the conditions at each electrode using a procedure that appropriately controls the Type I error rate, for example, with the Bonferroni *t* procedure or Tukey (1953) test. These standard analytic procedures suffice to secure the conclusion that scalp distributions between experimental conditions differ and to identify the electrode locations where the difference is reliable. The validity of these inferences then, does not depend in any way on normalizing or scaling the amplitude of the distributions of the potentials in the two conditions. Indeed, if the actual distribution of the effect is of interest, the distributions should not be scaled, because doing so can attenuate or even shift the spatial locus of the effect (Picton et al., 2000, and see also our Figure 1C and Figure 7).[2] Thus amplitude normalization is not a follow-up procedure that must be conducted to ensure that distributional differences responsible for significant Condition × Electrode interaction effects are statistically reliable. For some important types of inference in ERP research, amplitude scaling distributions of scalp potentials is neither required nor appropriate. Indeed, inappropriate application of the procedure can lead to errors when determining the locations where experimental effects occur.

Vector Scaling

Distributions of scalp potentials were defined above as the magnitudes at a set of scalp locations and represented as a vector, $S = \langle v_1 \ldots v_a \rangle$. As detailed in the Appendix, amplitude normalization by vector scaling is a two step process that first projects potentials measured at these *a* scalp electrodes onto the axes in an *a*-dimensional vector space. The vector representation of the distribution of potentials affords a perfectly general and mathematically precise characterization of topographic shape and overall amplitude: Shape corresponds to vector orientation and amplitude corresponds to vector length. The length of a vector and its orientation can vary independently and orientation can be held constant under transformations that change vector length.

[2] Recognizing how these distributional distortions can arise is a special case of understanding how amplitude normalization transforms scalp distributions in general. For instance, it appears to be widely accepted that a popular type of vector scaling eliminates the main effect of condition. Although this is true when distributions have the same topographic shapes, when the topographic shapes differ, there may or may not be a residual main effect of Condition (and for that matter there may or may not be a Condition × Electrode interaction effect). The point can be illustrated by a simple example with two scalp distributions S_1 and S_2 for two conditions and three electrode locations:

$$S_1 = \langle -2.323, -1.333, 1.267 \rangle$$

$$S_2 = \langle -1.463, 0.517, 5.717 \rangle.$$

The vector lengths $|S_1|$ and $|S_2|$ for S_1 and S_2, respectively, are given by

$$|S_1| = \sqrt{(-2.323)^2 + (-1.333)^2 + (1.267)^2} = 2.963$$

$$|S_2| = \sqrt{(-1.463)^2 + (0.517)^2 + (5.717)^2} = 5.924.$$

Using these vector lengths to scale the corresponding distributions gives

$$S_{1 Vector\ Scaled} = \langle -2.323/2.963, -1.333/2.963, 1.267/2.963 \rangle$$

$$= \langle -0.784, -0.450, 0.428 \rangle$$

$$S_{2 Vector\ Scaled} = \langle -1.463/5.924, 0.517/5.924, 5.717/5.924 \rangle$$

$$= \langle -0.247, 0.087, 0.965 \rangle.$$

The difference between the two scaled distributions is approximately the same at each of the three electrodes (about 0.54), which means there is a main effect of Condition for the vector-scaled distributions. This example is illustrated in Figure 7, third row from the top (rounding error is responsible for the discrepancies in the third decimal place between the scaled magnitudes calculated above and the values in Figure 7). The possibility of main effects after normalizing amplitudes is also suggested by the vector-scaled 36 electrode distributions illustrated in Figure 6.

Vector scaling is one such transformation, and the actual transformation is calculated by dividing each potential in the distribution by the overall vector length given by the square root of the sum of the squares of the potentials. After dividing the potentials in the distribution by the vector length, the vector representation of the scaled distribution has unit length, so any distributions scaled in this way have the same overall amplitude. Furthermore, because all the potentials are divided by the same factor, their relative proportions, hence vector orientation and topographic shape, are preserved (Figure 2 and Figure 3).

Vector scaling thus eliminates overall amplitude differences between distributions while preserving the topographic shapes, and, in doing so, is widely thought to solve the problem of inferring different spatial configurations of generators (McCarthy and Wood, 1985; Picton et al., 2000; Ruchkin et al., 1999). The supposition is that if the amplitudes of the measured potentials in two experimental conditions are vector scaled and ANOVA on these scaled distributions still yields a significant Condition × Electrode interaction, this interaction effect must be due to differences in topographic shape, and, as reviewed above, in the theoretical ideal, such differences in shape entail different spatial configurations of generators.

These inferences are illustrated in the two columns on the right of Figure 1. In an example where sources in the two conditions have the same location and orientation and differ only in strength (Figure 1A), the distribution of the potentials at the seven electrode locations shows a crossover interaction effect: The difference between the conditions is larger at some electrodes than others and reverses polarity. In this example, the distributions differ at each electrode by a factor of 2, so the vector representations have different lengths but the same orientation, that is, these two distributions of scalp potentials have the same topographic shape and differ only in overall amplitude. The plots of the two distributions after vector scaling show that the scaled distributions are identical and the crossover evident in the unscaled potentials has been eliminated entirely. An ANOVA conducted on these scaled distributions should be statistically unlikely to find a significant Condition × Electrode interaction effect, and, in the absence of such a finding, the inference to different spatial configurations of current sources is invalid. So in this case, vector scaling has succeeded in preventing the misattribution of the distributional difference in the unscaled potentials to different spatial configurations of generators.

In contrast, an example in which vector scaling would support an inference to different spatial configurations is illustrated in Figure 1B. Here, the distributions of the potentials at the sensors again show a crossover effect, but unlike the previous example, the potentials at each electrode do not differ by the same scalar multiple, so their topographic shapes differ. Scaling by vector length clearly does not eliminate distributional differences, and an ANOVA conducted on the scaled magnitudes is likely to find a significant Condition × Electrode interaction. Because the scaled distributions differ in shape even after overall amplitude differences are removed, the inference to different spatial configurations of generators is warranted. Finally, the third crossover effect illus-

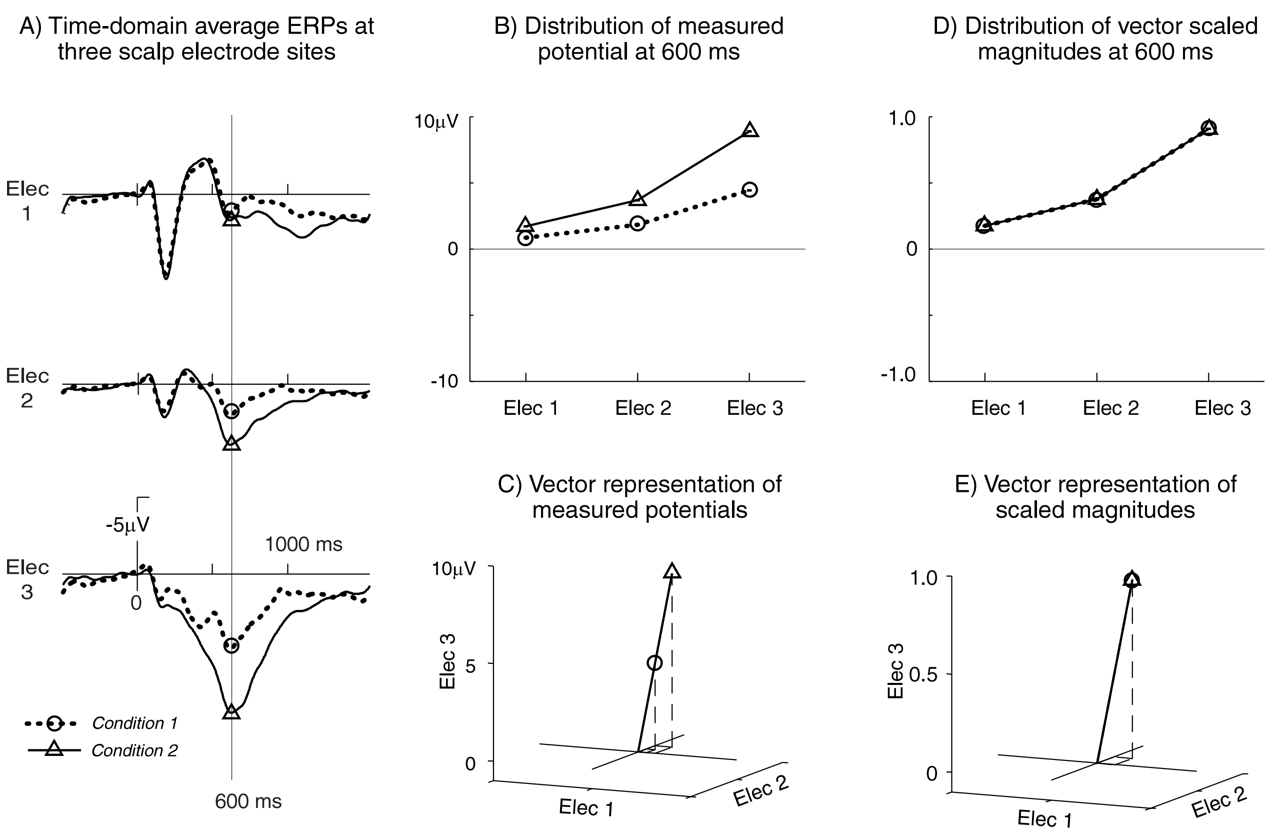


Figure 2. Vector representations of potential distributions that have the same topographic shape. A: Representative ERP waveforms. B: Distributions of unscaled peak amplitude measurements made at 600 ms poststimulus in Condition 1 and Condition 2. C: Vector representation of the potential distributions in B. Note the identical vector orientations. D: Distributions of the magnitudes in Condition 1 and Condition 2 after amplitude is normalized by vector scaling. E: Vector representation of the normalized magnitudes.

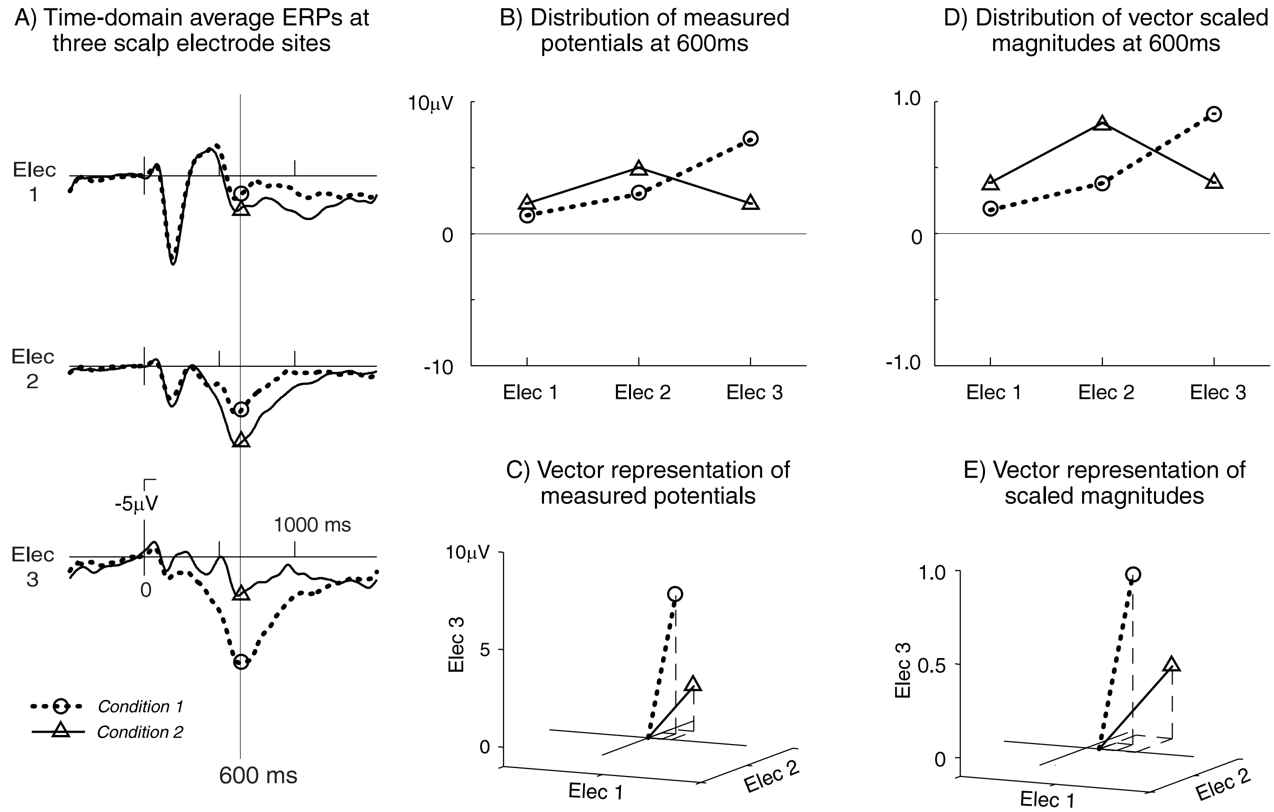


Figure 3. Vector representations of potential distributions that have different topographic shapes. A: Representative ERP waveforms. B: Distributions of unscaled peak amplitude measurements made at 600 ms poststimulus in Condition 1 and Condition 2. C: Vector representation of the distributions in B. Note the different vector orientations. D: Distributions of the magnitudes in Condition 1 and Condition 2 after amplitude is normalized by vector scaling. E: Vector representation of the normalized magnitudes.

trated in Figure 1C behaves in the same way as the example immediately above. Here again the amplitudes at each electrode do not differ by the same scalar factor and the distributional differences that remain after vector scaling are likely to lead to significant Condition × Electrode interaction effects, thereby licensing the inference that the spatial configurations of generators differs.

These examples show that in the theoretical ideal, amplitude normalization can establish differences in the spatial configurations of generators. It should be noted, however, that this is true only if "spatial configuration" is defined in such a way that nothing about the differences in the location or number of generators follows from differences in their spatial configuration. Prior to amplitude normalization, differences in the distribution of surface potentials show that one or more of the three disjuncts is true:

1. generators differ in location OR
2. generators differ in polarity OR

3a. generators differ in overall strength or

3b. generators differ in relative strength (and not both 3a and 3b).

After amplitude normalization by vector scaling, differences between distributions of scaled magnitudes are differences in topographic shape. Such differences show that the spatial configurations of generators differ, that is, that one or more of the following is true:

1. generators differ in location OR
2. generators differ in polarity OR

3b. generators differ in relative strength.

Even in the theoretical ideal, nothing stronger than this disjunction may be validly inferred from differences in the topographic shape of surface potentials and it is important to note, in particular, that distributions where generators have the same locations and polarity but different relative strengths count as different spatial configurations. Thus, showing that spatial configurations of generators differ is *not* the same as showing that the location or number of the generators differ. The point that differences in relative source strengths count as differences in spatial configurations has been made before (Alain, Achim, & Woods, 1999; Picton et al., 2000), but the fact that this undermines the possibility of inferring differences in generator location from differences in spatial configuration deserves wider attention.

Critique of Vector Scaling

Vector scaling is intended to license inferences to the conclusion that the spatial configuration of neural current sources differs between experimental conditions. At this point, one might be tempted to conclude that although the distributional differences that remain after amplitude normalization do not provide a great deal more inferential traction than distributional differences in the unscaled surface potentials, they do rule out one possible explanation, that is, differences in overall strength for generators with fixed locations and polarities. Thus, there would seem to be no harm in vector scaling as long as one does not read more into the conclusion that spatial configurations differ than is warranted by the facts. However, the discussion so far has considered the inferential issues only in the theoretical ideal. It is argued next that measured ERP scalp distributions unavoidably depart from this ideal in ways that make amplitude normalization unsound even for its limited application in ruling out differences in overall generator strength. That is, in ERP practice, normalizing the amplitudes of measured distributions in two experimental conditions is liable to leave residual differences in topographic shape that are *not* the result of between-condition differences in the spatial configurations of generators, even when spatial configuration is understood in the circumscribed sense defined above.

The measured distributions of scalp potentials in experimental ERP research depart from ideal distributions of surface potentials generated by current sources and sinks in two fundamental ways. First, distributions of scalp potentials recorded during the poststimulus interval of interest are measured by subtracting some baseline potential distribution. Even in well-designed experiments, where these baselines do not differ between conditions, nothing ensures that they are numerically zero at all electrodes. When subtracted from the poststimulus distributions of interest, these nonzero baseline potentials can result in differences in the topographic shape of the *measured* distributions even when the poststimulus generators themselves do not differ in spatial configuration. Because topographic shape alone cannot distinguish genuine differences in the spatial configuration of poststimulus sources versus the contribution of the baseline potential, the amplitude normalization procedure does not allow valid inference to different spatial configurations of poststimulus generators. Indeed, nonzero baseline potentials will pose a problem for identifying generator configurations for any procedure that operates on the algebraic difference of poststimulus and baseline distributions. A second issue for amplitude normalization is noise, for although time-domain averaging across trials improves the signal-to-noise ratio, ERP measurements are never noise free. Setting aside technical artifacts, electrical interference, and noncortical potentials, which might at least be mitigated, variability resulting from differences between individual subjects is unavoidable. Noise is a problem for vector scaling because it contributes to the amplitude of a distribution and tends to increase vector length. Noise-induced overcorrection can result in residual differences in topographic shape after scaling, even when the spatial configurations of the generators are identical and the levels of noise are the same. Nonzero baseline potentials and noise variability are facts of life in ERP research and the following section details how they render the inference from differences in vector-scaled distributions to differences in source configurations invalid.

Consequences of Nonzero Baseline Potentials for Vector-Scaled Distributions

In ERP paradigms, experimental effects found in recorded poststimulus potentials cannot be unequivocally attributed to differences in stimulus processing because differences present before the stimuli could persist into the interval of experimental interest. The standard procedure to address this possible confound is to measure the recorded poststimulus potentials by subtracting a baseline potential distribution, typically recorded during a brief prestimulus interval. This baselining procedure is well motivated, because, in the absence of such an adjustment, a difference between the prestimulus potentials would invalidate inferences that attribute poststimulus effects to differences in stimulus-related processing. However, an additional assumption, perhaps implicit, is that *only* a difference between the baseline potential distributions invalidates such an inference, that is, that in the absence of a difference between conditions, the numerical magnitudes of the baseline are irrelevant for the attribution of poststimulus effects to stimulus processing. This second assumption is unproblematic for the measurement and analysis of unscaled potentials but fatal for vector scaling.

For ANOVAs conducted on the unscaled distributions, F ratios for the Condition $\times$ Electrode interaction effect are invariant under changes in the distribution of the baseline potentials as long as there is no between-conditions baseline difference. For example, in this type of analysis, it is immaterial whether a measured 5 μV effect is a difference between 8 and 3 μV or between 3 and -2 μV. The situation, however, is quite different for vector scaling. Baseline distributions with different numerical magnitudes result in measured distributions with different amplitudes. Because vector scaling is an amplitude scaling procedure and measured amplitude depends on the baseline potentials, it is not surprising that both the topographic shape given by the vector representation and the F ratios obtained by conducting Condition $\times$ Electrode ANOVAs on vector-scaled potentials vary as a function of the numerical magnitudes of the baseline potential distribution even when the baseline distributions in the two conditions are identical. In this respect, the analysis of potentials measured against a baseline distribution behaves in fundamentally different ways for unscaled and vector-scaled distributions.

The consequences of nonzero baseline potentials can be illustrated by extending the previous examples. Consider again a case in which poststimulus generators in two conditions differ only in overall strength (Figure 4A). These generator distributions are multiplicatively related so the scalp potential distributions have the same topographic shape and vector scaling eliminates distributional differences. Now suppose that identical (nonzero) generators are active during the baseline interval in both conditions (Figure 4B). When the baseline distributions (Figure 4B) are subtracted from the poststimulus scalp distributions (Figure 4A), the measured distributions of potentials that result (Figure 4C = Figure 4A $-$ Figure 4B) are no longer multiplicatively related. Thus, even though the baseline distributions are identical in both conditions and poststimulus sources differ only in overall strength, the *measured* distributions do *not* have the same topographic shape. The inferential problem arises because the surface distributions measured against the same nonzero baseline are indistinguishable from surface distributions associated with generators that differ only in relative strength, that is, Figure 1C and Figure 4C are identical. Baseline sources and poststimulus sources have the same effect on the topographic shape of the measured distributions; thus, amplitude normalization treats baseline generators that are the same in both conditions just like poststimulus generators with fixed strength, location, and polarity. When the actual poststimulus generators differ only in overall strength, the measured distribution appears to have some generators that change in overall strength (the poststimulus generators) and some that do not (the baseline generators). This combination looks like a difference in relative generator strength, hence a difference in the spatial configuration of generators. This example clearly demonstrates that distributions of potentials in two conditions derived by subtracting the same

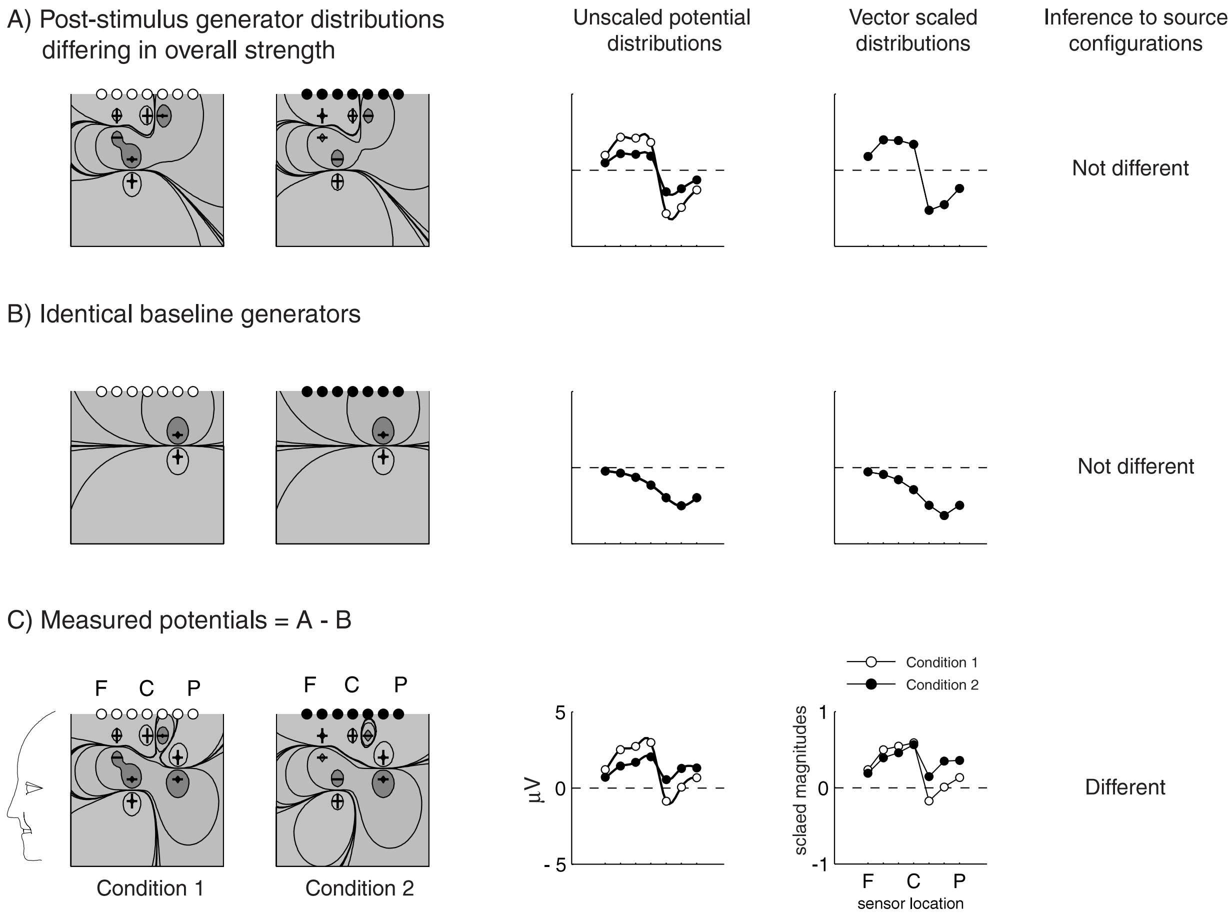


Figure 4. Nonzero baseline potentials can change the topographic shape of measured potential distributions. A: Poststimulus potentials associated with generator distributions that have the same spatial configuration and differ only in overall strength. B: Identical nonzero baseline potentials in both conditions. C: The measured potential distributions (A − B) do not have the same topographic shape, even though the poststimulus and baseline distributions, considered individually, do. The algebraic difference of these poststimulus and baseline potentials is indistinguishable from the potentials produced by generator distributions with different relative strengths (Figure 1C). Thus, measured potential distributions may have different topographic shapes even when there is no between-condition difference in the spatial configuration of the poststimulus or baseline generator distributions.

nonzero baseline potentials may have different topographic shapes even in the absence of any between-condition difference in the spatial configuration of the generators. So, unless baseline potential distributions are numerically zero (which we, of course, may never know), the inference from differences in vector scaled *measured* distributions to different spatial configurations of *poststimulus* sources is invalid.

It is important to realize that the problem posed by nonzero baseline potentials is not an idiosyncratic feature of this particular example, but follows from the mathematical properties of vector representations. Vector scaling trades on the insight that for an a-dimensional vector $V = \langle v_1, \ldots, v_a \rangle$, scalar multiples of V will have the same orientation, that is, whenever $V' = Vm$, for scalar m, the orientations of V and V' are the same. Geometrically, the tail of the vector is anchored at the origin of the axes and multiplying the projection along each axis by m changes the projections in proportion to one another, stretching the length of the vector without changing its orientation. However, the orientation of multiplicatively related vectors V and V' is not generally invariant under the addition or subtraction of a vector $B = \langle b_1, \ldots, b_j \ldots, b_a \rangle$ with nonzero b_j, because adding each b_j to the magnitudes projected along the jth axis moves the head of the vector, and, unlike multiplying by a scalar, nothing ensures that vector orientation is preserved. So except in a few special cases, whenever two vectors V' and V have the same orientation, that is, $V' = Vm$, the vector sums $(V' + B)$ and $(V + B)$ have different orientations.

In application to the analysis of scalp potentials, this means that if the poststimulus distributions represented by V' and V have the same topographic shape in the sense of vector orientation, the distributions measured against the nonzero baseline represented by B, that is, $V' - B$ and $V - B$, do not. The consequence of this unspectacular algebraic fact for ERP research is that when poststimulus distributions generated by sources differing only in strength are measured by subtracting nonzero baseline distributions, the vector-scaled measured distributions do not have the same topographic shape, and normalizing these measured distributions by vector length does not eliminate distributional differences. Nonzero baseline potentials thus pose a fundamental problem for the inference from vector-scaled distributions of measured potentials to different configurations of poststimulus sources and, in general, normalizing the amplitude of measured distributions tells us nothing about the spatial configuration of poststimulus generators.

To explore what this argument in principle might mean in experimental practice, simulated distributions of scalp potentials were constructed for two experimental conditions where sources differ in strength but not configuration. These two distributions were embedded in zero mean normally distributed noise with a standard deviation of 2 μV for 16 subjects and 36 electrode locations (Figure 5) and the simulation was run 1,000 times. The distributions were then normalized by the within- and across-subjects vector scaling procedures defined in the Appendix.

The contour and line plots give the overall average of the 1,000 grand means across the 16 subjects. The histograms give the distribution of the 1,000 F ratios for the Condition × Electrode

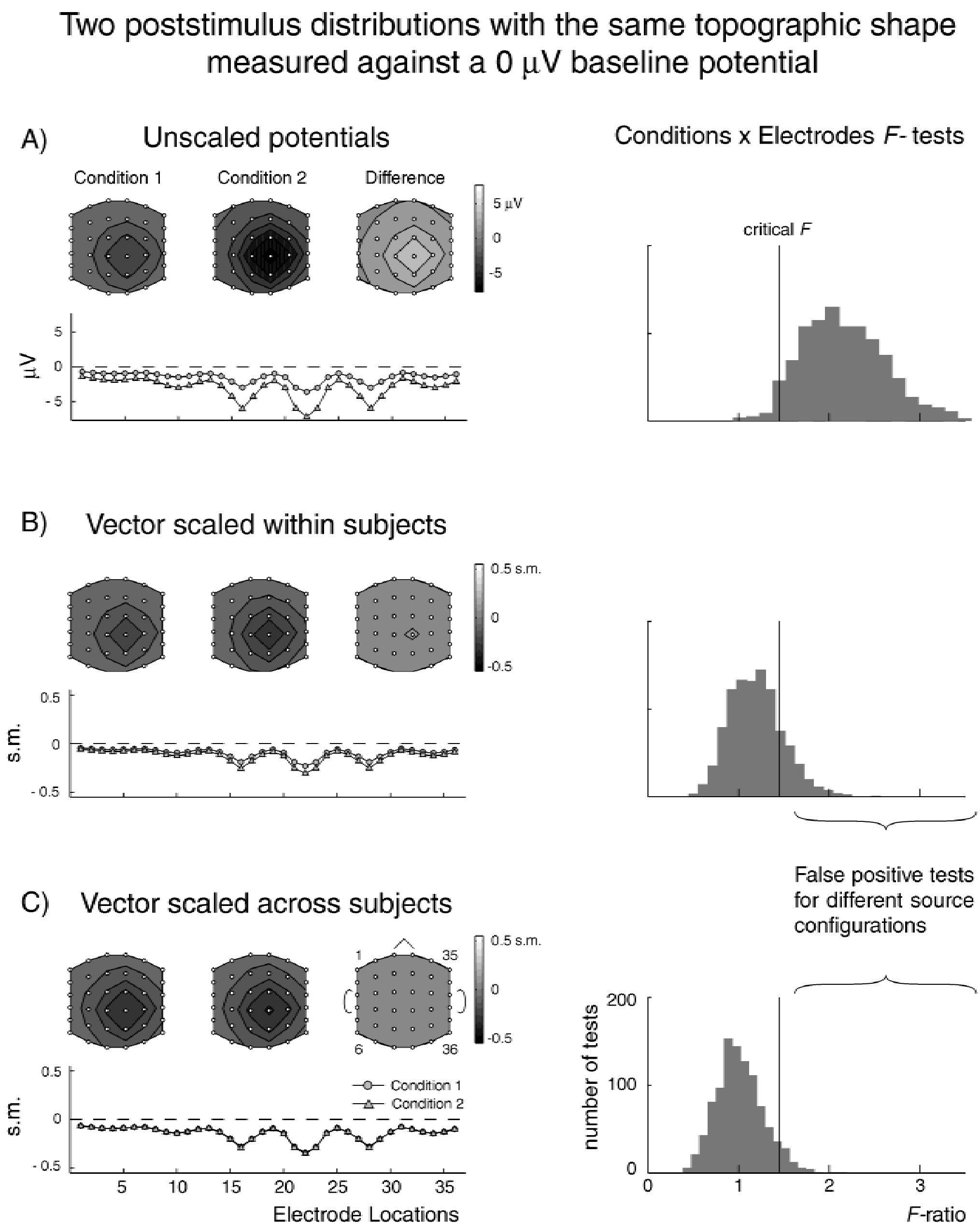


Figure 5. Simulated poststimulus distributions of scalp potentials with the same topographic shape in two conditions measured against an ideal baseline potential distribution of 0 μV at all electrodes. Contour and line plots of the distributions are averages of the grand averages across the 16 subjects for 1,000 runs for the unscaled potentials and scaled magnitudes (s.m.). Condition (2) × Electrode (36) repeated measures ANOVAs were conducted for each of the 1,000 runs. Histograms show the distribution of the Condition × Electrode interaction effect F ratios for the unscaled and both vector scaled distributions. The solid vertical bar at 1.46 in the histograms indicates critical $F(35,525)$ at $p = .05$ for the nominal degrees of freedom.

interaction effects obtained for the Condition (2) × Electrode (36) repeated measures ANOVAs. For all significance tests in this paper, critical F at $p = .05$ was calculated on degrees of freedom adjusted for violations of sphericity according to Huynh and Feldt (1976). The contour plots for the unscaled potentials show that this simulated negativity is a fairly large effect, with a slightly right lateralized parietal maximum of about 5 μV (Figure 5A). The plots for the vector-scaled distributions show that scaling within subjects greatly reduces the effect (Figure 5B), and scaling across subjects virtually eliminates it (Figure 5C). The distribution of F ratios confirms what would be expected for a large effect in moderate noise with 16 subjects. The vast majority of the tests (0.962) on the unscaled potentials exceeded critical F and because both types of vector scaling reduce or eliminate the distributional differences between the two conditions, far fewer of the effects were significant (0.174 for within-subject vector scaling and 0.059 for across-subject vector scaling). The proportions of significant tests for the scaled distribution are restricted to those cases where a significant Condition × Electrode interaction effect was found for the unscaled distributions. This example shows that both procedures, and vector scaling across subjects in particular, can do a good job of eliminating distributional differences under the idealization that the baseline potential distribution is numerically zero at all electrodes.

To illustrate the consequences of nonzero baseline potentials, a second set of simulations was conducted using the same parameters as above, except that the poststimulus distribution was measured by subtracting a baseline potential distribution that was a constant -2.5 μV at all electrodes in both conditions. There is no difference between the distribution of the baseline potentials themselves, and measuring the poststimulus distribution against this baseline simply shifts the effect by 2.5 μV. The plots for the unscaled distributions again suggest a right parietal source that differs only in strength between the conditions (Figure 6).

The magnitude of the experimental effect prior to vector scaling in this set of simulations is the same as in the first set (Figure 5A), and, as expected, the proportion of significant F tests for the Condition × Electrode interaction effect is nearly identical (0.963). From the algebraic argument above, the distributions measured by subtracting the same nonzero baseline should have different topographic shapes even though the poststimulus sources differ only in strength. The plots in Figure 6B and Figure 6C bear this out and show that, on average, neither vector scaling within nor across subjects eliminates the differences in topographic shape. The between-condition effect that remains for the vector-scaled distributions is of sufficient magnitude to result in Condition × Electrode interaction effects above critical F far more often than the 5% expected by chance. For vector scaling within subjects, the proportion of significant F tests was 0.879, and for vector scaling across subjects, the proportion was 0.486. Thus, where the null hypothesis is that the poststimulus source configurations do not differ, hypothesis testing with ANOVAs conducted on scaled distributions measured by subtracting nonzero baseline potentials can lead to seriously inflated Type I error rates. In experimental practice, a baseline potential distribution would be expected to vary from electrode to electrode and there would be noise variability in the baseline as well, but the flat noise-free baseline serves to illustrate the general point.

Thus, these simulations show that when the prestimulus baseline is subtracted from the analyzed waveforms, (1) Condition × Electrode interactions for vector-scaled distributions can be sensitive to the numerical magnitude of the baseline potentials even when the baseline distributions are identical in both conditions, and (2) that nonzero baselines can inflate error rates when testing for different poststimulus source configurations.

Vector scaling procedures fail to take into account the fact that the topographic shapes of distributions of potentials in two conditions may differ merely because some nonzero baseline distribution was subtracted from the distributions of experimental interest. This means that the standard and unavoidable practice of subtracting baseline potentials can lead to differences in the topographic shape after vector scaling even when there is no between-condition difference whatsoever in the spatial configuration of the generators either in the baseline or in the interval of experimental interest. Although the baseline-to-mean amplitude measure has been used to illustrate the problem, the arguments applies equally to any measure derived from the algebraic combination of two scalp distributions, including baseline-to-peak and peak-to-peak measures. The important consequence for the interpretation of ERP data is that even when between-condition differences in topographic shape remain after amplitude normalization by vector length or root mean square (r.m.s.) amplitude, it does not necessarily follow that the spatial configuration of the neural generators differs between the experimental conditions.

Two general comments are in order. First, subtracting nonzero baseline potentials poses an interpretive problem for unscaled potentials as well, because here, too, the resulting distribution reflects the combination of pre- and poststimulus generator activity. However, as long as there is no between-condition difference in the baseline potential distributions, F ratios for Condition × Electrode interaction effects for unscaled distributions are invariant under measurement against different baselines. Subtracting nonzero baseline potentials will, naturally, change the distribution of the unscaled potentials. However, unlike the unscaled distributions, for vector-scaled distributions, the experimental effect, that is, the difference between the conditions after scaling, *does* vary with the numerical magnitudes of baseline distribution itself. In this respect, the comparison of vector-scaled distributions is sensitive to the specific numerical magnitudes of the baseline potentials in a way that comparison of unscaled distributions is not. This point is illustrated for a variety of nonzero baseline potentials in Figure 7.

Second, the specific failure of amplitude normalization illustrated by subtracting baseline potentials is a symptom of a deeper underlying problem that can be traced back to the mapping principle that was used to justify the procedure in the first place (repeated here for convenience):

> If two generator distributions G_1 and G_2 have the same spatial configuration, then the corresponding distributions of surface potentials have the same topographic shape.

This principle is true in theory, but the surface potentials referred to are never measured in experimental practice. In ERP research, it is well understood that talk about "the potential at an electrode" is a convenient fiction and shorthand for "the difference between the potential at an electrode and some reference potential, subsequently measured against a suitable baseline." Thus, the quantities typically being compared in ERP research are two distributions of potentials measured relative to some reference and baseline potentials and these are *not* measurements of surface potentials in the sense required for the mapping principle to be true. Instead, when amplitude normalization is used to infer differences in the spatial configuration of

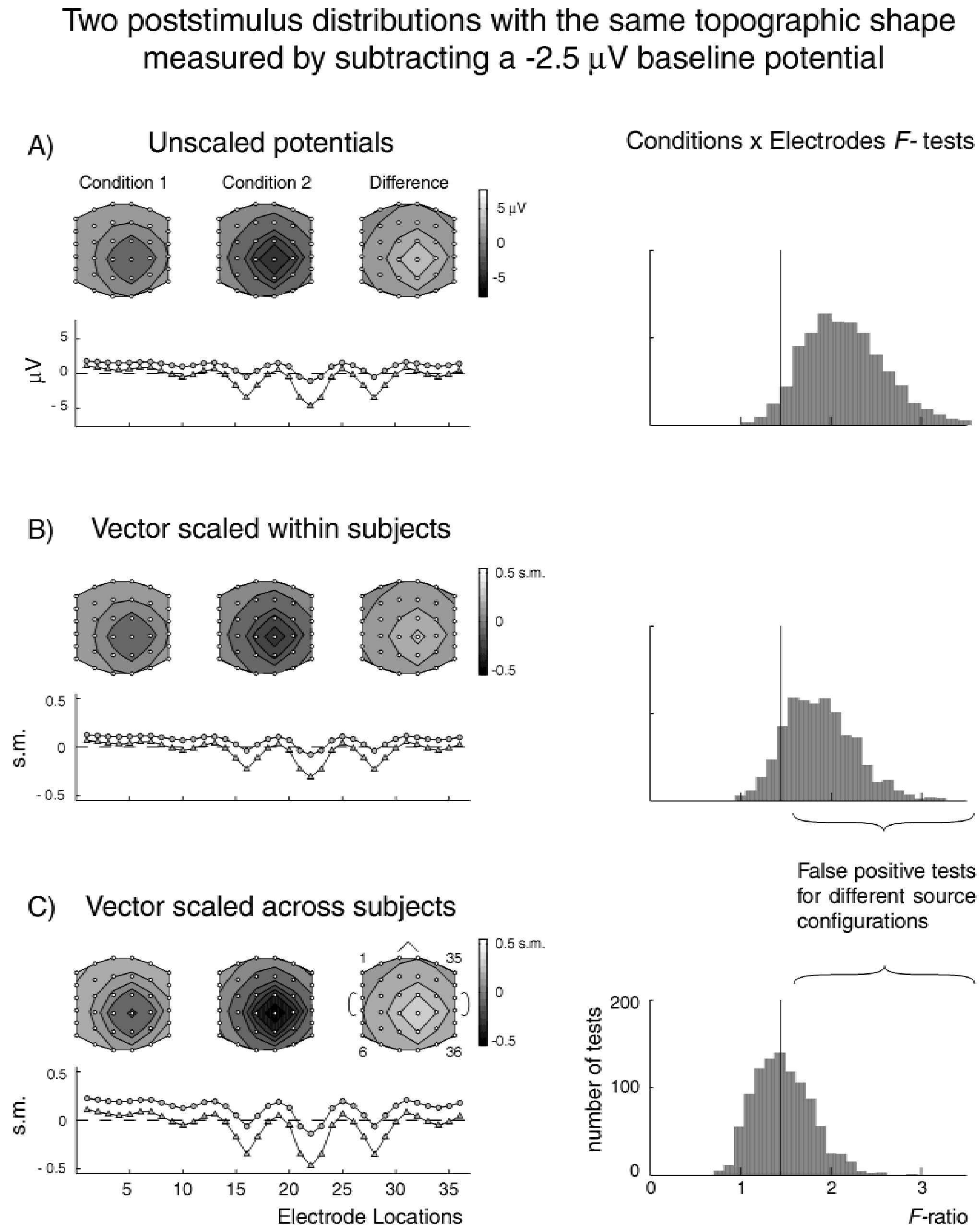


Figure 6. Simulated poststimulus distributions of scalp potentials with the same topographic shape in two conditions (compare Figure 5) measured against a noise-free flat baseline potential of $-2.5\ \mu$V at all electrodes. All other simulation parameters are identical to the simulations in Figure 5.

generators, an entirely different and incorrect "experimental measure" mapping principle is tacitly presupposed:

> If two generator distributions G_1 and G_2 have the same spatial configuration, then the corresponding distributions of surface potentials plus or minus the referenced baseline potentials have the same topographic shape.

The algebraic considerations rehearsed above show why this experimental mapping principle does not hold in general, and the simulations and Figure 7 give specific cases where it fails. Both the general argument and the counterexamples demonstrate that the inference from between-condition differences in the topographic shape of measured potentials to between-condition differences in the spatial configuration of neural generators is not valid.

Vector scaling is intended to address a serious inferential ambiguity, that is, that Condition × Electrode interaction effects on unscaled potentials can result from differences in the spatial configuration of neural generators or simply from differences in generator strength alone. However, vector scaling and equivalent r.m.s. amplitude normalization procedures require assumptions about the distributions being scaled that are not satisfied by experimentally measured ERP data, and, as a result, these procedures are liable to an equally serious inferential ambiguity of their own. Condition × Electrode interaction effects in vector-scaled distributions, that is, differences in topographic shape, can result from differences in the spatial configuration of generators or simply from anything that displaces the distributions relative to zero, such as subtracting a nonzero baseline potential. Thus, vector scaling replaces one inferential ambiguity with another, and, in the end, goes no further toward securing the conclusion that spatial configurations of neural generators differ between experimental conditions.

In addition to demonstrating the confounding effect of baseline potentials for vector scaling, these simulations also show that vector scaling within and across subjects may give different results. Further consideration indicates that noise variability presents a general problem for vector scaling that is independent of the baseline issue.

Consequences of Noise on Vector Scaling Procedures

In the absence of noise, vector scaling within and across subjects gives identical results; when there is noise variability, the two procedures may differ, and, in some cases where the spatial configuration of sources is in fact the same, neither procedure reliably eliminates differences in topographic shape. The consequences of variability for vector scaling have already received some attention. Haig et al. (1997) argue that if the covariance matrices of the scaled potentials satisfy the assumption of homogeneity, then, of mathematical necessity, this assumption is violated for the scaled potentials (see Ruchkin et al., 1999, for a reply). The concern here is different. With unscaled potentials, the expected values of the distribution do not change as the variability of zero mean noise increases, because positive and negative noise components, even if large, are equiprobable and tend to cancel out over the long run. However, vector length is a function of squared amplitude, and as noise variability increases, instead of canceling out, positive and negative noise components tend to increase vector length. Thus, as noise variability increases, so does vector length and this results in overcorrecting the amplitude. The consequences of this overcorrection are illustrated for simulated potential distributions at seven midline electrodes in two experimental conditions (Figure 8).

The salient feature is a 2-μV effect at the parietal electrode Pz, with smaller differences elsewhere including a 0.5-μV crossover at the frontal electrode Fz (Figure 8A, left). This sort of Condition $\times$ Electrode interaction can be produced by neural current sources differing in strength by a factor of three and is a canonical case in which normalizing the overall amplitude by vector scaling should eliminate the Condition $\times$ Electrode interaction. In the absence of noise, both vector scaling methods eliminate the Condition $\times$ Electrode difference between the conditions (Figure 8A, center and right) but neither method does when these same distributions are embedded in a moderate level of noise. On average, over 1,000 simulations with 16 subjects and zero mean noise with standard deviation = 2 μV, the unscaled distributions approximate the noise-free distributions (Figure 8B, left). However, for these simulation parameters, the results of the two vector scaling procedures differ from one another and neither converges on values that eliminate the Condition $\times$ Electrode interaction (Figure 8B, center and right). Scaling fails to eliminate distributional differences in this case, because the amplitude over correction at moderate noise levels has a greater impact on the distribution with the smaller overall amplitude (compare Condition 1 and Condition 2 in Figure 9).

The extent to which these residual noise-related differences inflate Type I error rates depends on several factors including the choice of scaling procedure, the variance of the noise, the specifics of the two distributions, and the overall amplitude difference between them. To illustrate the combined effect of noise variability and the contribution of nonzero baseline potentials, a series of simulations were conducted in which both factors were varied independently.

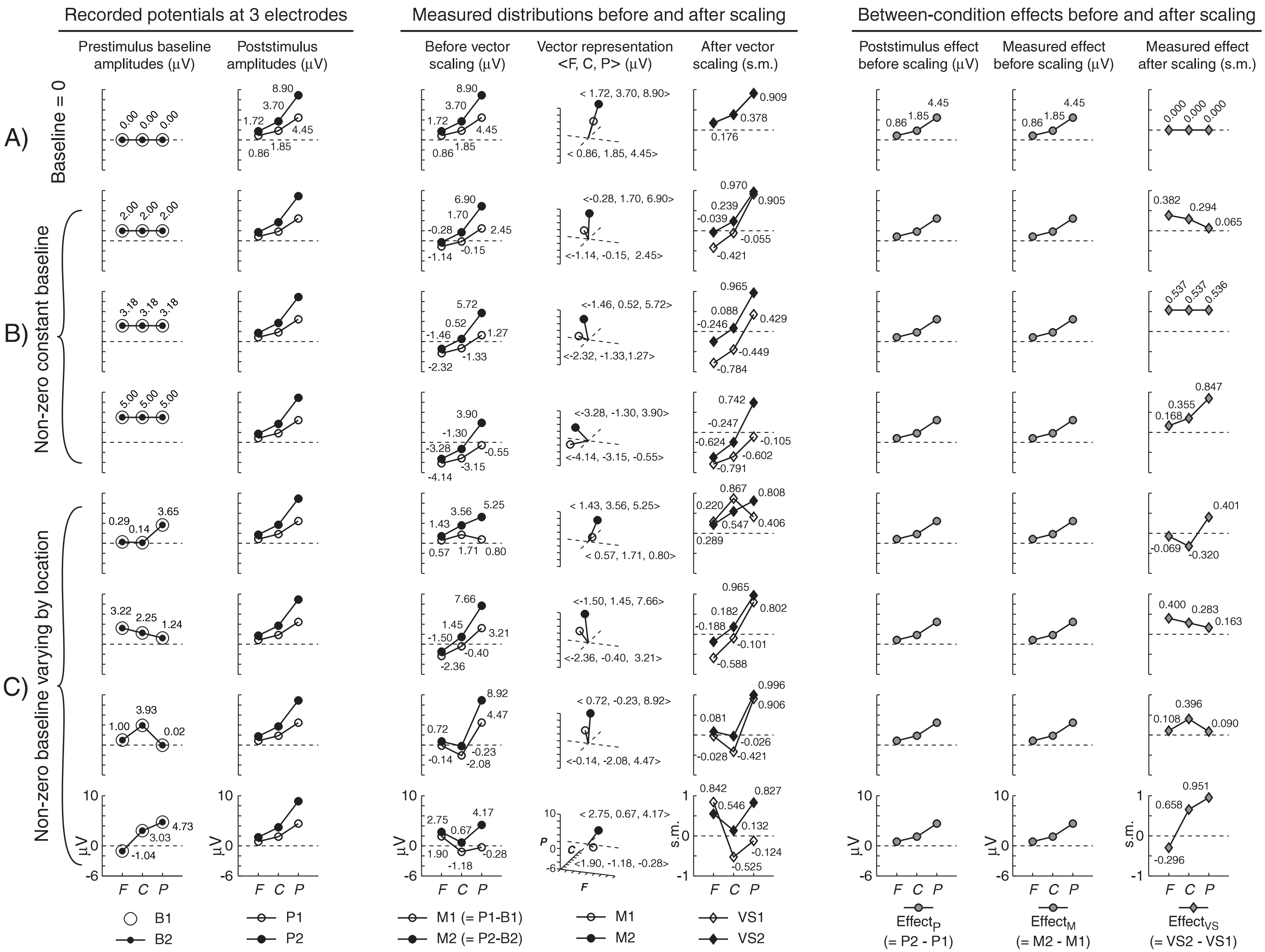
Error Rates as a Function of Scaling Procedure, Baseline, and Noise

The distributions of unscaled potentials were the same as in the first two sets of simulations and consistent with a right parietal source differing only in strength by a factor of two (as in Figure 5A). The two distributions were embedded in zero mean normally distributed noise for 16 subjects and spatially sampled at 36 electrodes. The standard deviation of the noise was varied from 1 to 10 μV in steps of 1 μV and the baseline potential (noise free and constant across electrodes) ranged from -5 to 5 μV in steps of 1 μV. The simulations were run 200 times for each combination of noise and baseline. For each run, Condition $\times$ Electrode ANOVAs were conducted as specified above. The proportion of the 200 tests of the Condition $\times$ Electrode interaction effect that exceeded critical F at each level of noise and baseline are summarized in Figure 10.

For the unscaled distributions (Figure 10A), the different baseline potentials have no effect on the proportion of significant F tests. At very low levels of noise, 100% of the tests exceeded critical F, and at very high noise levels, the proportion is close to the 5% expected by chance. Between these noise extrema, the proportion of significant tests falls off as noise increases, and, overall, the effect is likely to be detected as long as the standard deviation of the noise is below about 3 μV, and not very likely to be detected when the standard deviation of the noise is above 5 μV. These unsurprising results are included to establish that the basic distributional effect is of an experimentally plausible magnitude—neither so small as to be lost in typical levels of noise nor so large as to be detectable at atypically high levels of noise. For vector scaling within subjects (Figure 10B) and vector scaling across subjects (Figure 10C), Type I error rates can be as high as 100%. It has been argued that vector scaling across subjects is more conservative than vector scaling within subjects because the former preserves between-subject variability (Ruchkin et al., 1999), and, with a few exceptions at low levels of noise, this appears to be the case in the present example. Nevertheless, it is also clear that being relatively more conservative than within-subject vector scaling does not make across-subject vector scaling a conservative test across the board. In addition to the inflated Type I error rates, this example also shows that vector scaling across subjects can be sensitive to small changes in both baseline potentials and noise: As the baseline potential ranges between -2 and 2 μV and the standard deviation of the noise ranges between 1 and 3 μV, error rates vary from below 5% to over 80%. These simulations are greatly oversimplified in several respects, and do not attempt to determine the error rates for the two vector scaling procedures that might be expected in practice. Nevertheless, having shown that nonzero baseline potentials and noise variability are problems in principle for the intended application of vector scaling in ERP research, the simulations indicate these two factors can substantially and systematically inflate Type I errors in the analysis of simulated data that are not altogether unlike a typical ERP effect.

Conclusion

It is uncontroversial that reliable between-condition topographic differentiation in the distribution of scalp potentials show that corresponding neural generators differ in some respect or other. It


Recorded potentials at 3 electrodes
Measured distributions before and after scaling
Between-condition effects before and after scaling
Prestimulus baseline amplitudes (μV)
Poststimulus amplitudes (μV)
Before vector scaling (μV)
Vector representation <F, C, P> (μV)
After vector scaling (s.m.)
Poststimulus effect before scaling (μV)
Measured effect before scaling (μV)
Measured effect after scaling (s.m.)
A)
Baseline = 0
B)
Non-zero constant baseline
C)
Non-zero baseline varying by location
F C P
B1
B2
P1
P2
M1 (= P1-B1)
M2 (= P2-B2)
M1
M2
VS1
VS2
Effect$_P$ (= P2 - P1)
Effect$_M$ (= M2 - M1)
Effect$_{VS}$ (= VS2 - VS1)

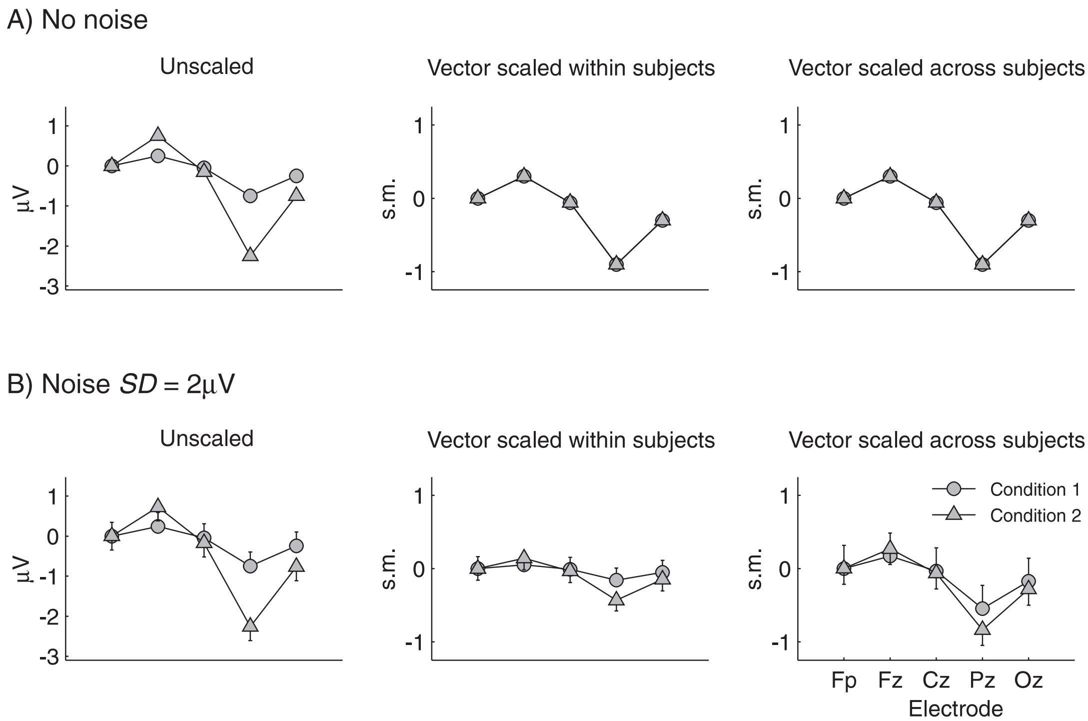


Figure 8. Noise variability causes amplitude over correction in vector scaling. A: Distributions of unscaled potentials and vector-scaled magnitudes (s.m.) with the same topographic shape at seven midline electrodes are properly scaled in the absence of noise. B: The overall mean across 1,000 simulations where the unscaled distributions were embedded in zero mean normally distributed noise with standard deviation = 2 μV for 16 subjects. Error bars indicate mean standard error over the 1,000 simulations. Vector scaling tends to overcorrect the amplitude of Condition 1 and does not eliminate distributional differences.

is also widely believed that amplitude normalization, for example, by vector scaling, can sharpen this conclusion by showing that spatial configurations of generators differ. This paper has reviewed why, under a suitable definition of "spatial configuration," this proposition is indeed true for ideal distributions of generators and surface potentials. However, it was also argued that under this definition, establishing different spatial configurations is of limited interest in principle and it was shown, in addition, that the proposition fails in application to ERP data. Three main points emerged.

First, the distinction between distributions of scalp potentials and distributions of neural generators must be clearly maintained. The only motivation for normalizing the amplitude of distributions of scalp potentials is to draw inferences about distributions of generators, and normalization is entirely irrelevant to the issue of whether there are differences between distributions of scalp potentials. Even if the procedure were otherwise sound, it was never intended to be nor should it be treated as a post hoc test to ensure the reliability of distributional differences in surface potentials. Amplitude normalization aims to improve upon the general conclusion that distributions of neural generators are somehow different by replacing it with a more specific conclusion about the way the generators differ. For some research purposes, it is important to

Figure 7. Scalp distributions and between-condition experimental effects before and after vector scaling. The two leftmost columns illustrate scalp distributions for recorded baseline amplitudes (B_1, B_2) and recorded poststimulus amplitudes (P_1, P_2) at three electrodes (*F*rontal, *C*entral, *P*arietal) in microvolts. The same two recorded poststimulus distributions are used in all examples; the baseline potential distributions vary from example to example, but in each comparison, $B_1 = B_2$, that is, there are no between-condition baseline differences. The three middle columns illustrate, in order, the distribution of the unscaled measured amplitudes ($M_1 = P_1 - B_1$, $M_2 = P_2 - B_2$), the vector representation of these measured distributions, and the distributions after vector scaling (VS_1, VS_2) in scaled magnitudes (s.m.). The three rightmost columns illustrate the distribution of the between-condition experimental effects for the poststimulus potentials ($P_2 - P_1$), the unscaled measured potentials ($M_2 - M_1$), and the vector scaled magnitudes ($VS_2 - VS_1$). In each plot, circles are unscaled potentials (in microvolts), diamonds represent s.m. A: Ideal baseline potential distributions of 0 μV at all electrodes. B: Nonzero baseline potentials with the same value at each electrode. C: Nonzero baseline potentials that vary by electrode location. For the unscaled potentials, the between-condition effect ($M_2 - M_1$) does not change when different baselines are subtracted from the poststimulus potentials. For the vector-scaled distributions, the between-condition effect for the scaled magnitudes ($VS_2 - VS_1$) varies widely when the different baselines are subtracted from the poststimulus distributions, and, depending on the distribution of the baseline potentials, these effects may increase monotonically from *F* to *P*, decrease monotonically, cross over, or appear as a main effect.

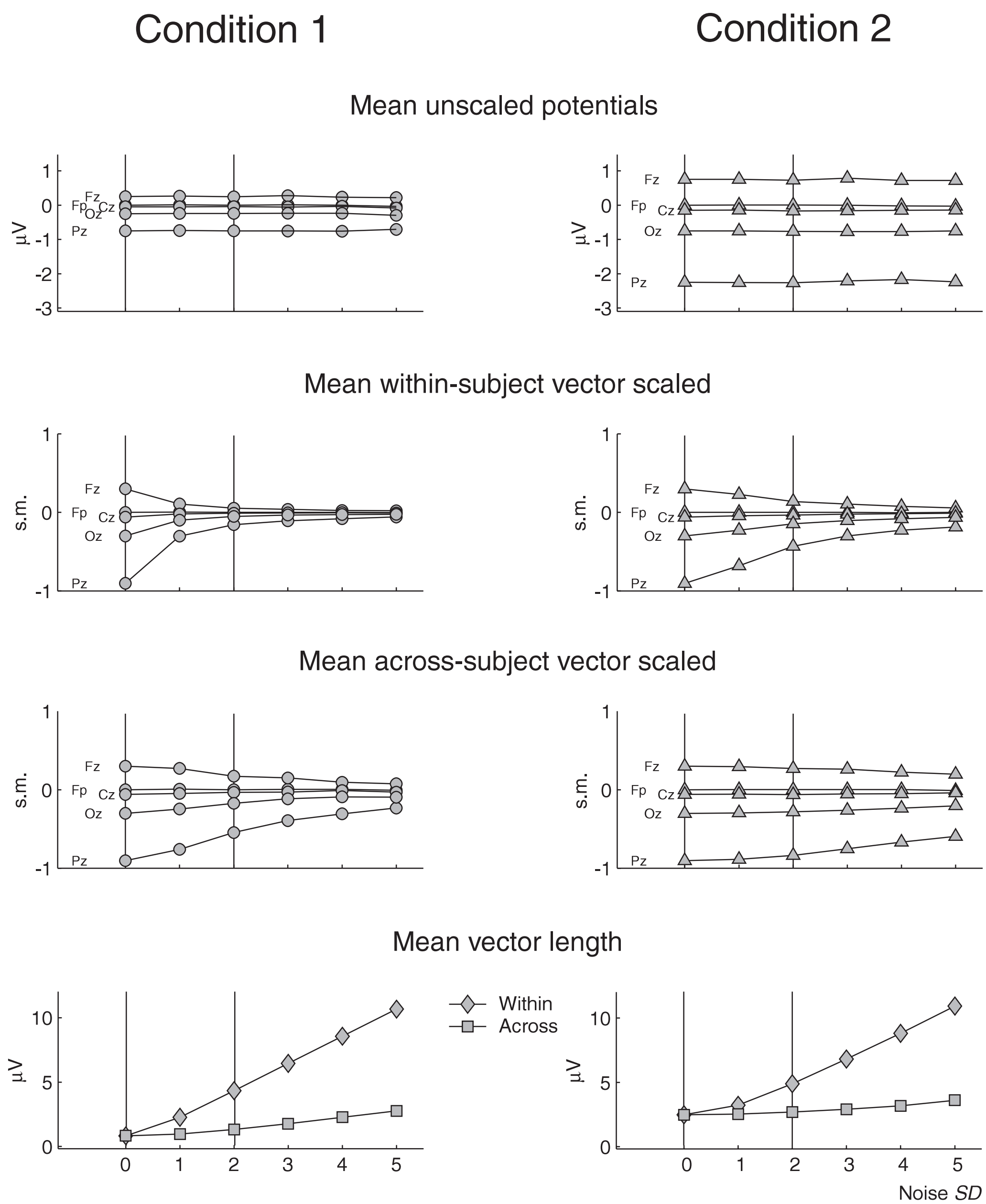


Figure 9. Vector length increases with noise variability. Plots give the mean across 1,000 simulations in each of two conditions for the distributions of unscaled potentials from Figure 8 over a range of noise levels. The unscaled potentials for 16 subjects at seven electrodes were embedded in noise with standard deviations ranging from 0 μV (noise free) to 5 μV. The distributions of the scaled magnitudes (s.m.) show that at moderate levels of noise, scaling by vector length fails to eliminate distributional differences even when ideal (noise-free) topographic shapes are identical.

draw such inferences; for other purposes it is not, and blanket insistence on amplitude normalization runs roughshod over this distinction.

Second, even for ideal distributions of generators and surface potentials, the extent to which amplitude normalization refines conclusions about generator distributions is rather limited. Prior to amplitude normalization, differences in scalp distributions show that neural generators differ in some combination of location, polarity, and relative strength or overall strength. After amplitude normalization, residual differences show that neural generators differ in some combination of location, polarity, or relative strength, that is, differ in spatial configuration. That is, of all the possible combinations of differences in generator locations, polarities, and strengths that could lead to different scalp distributions, amplitude normalization at best only rules out the one special case where the generators in the two conditions all have the same locations, have the same polarities, and differ in strength by the same multiplicative factor. If, after this case is ruled out, the remaining possibilities are lumped together as differences in the spatial configurations of the neural generators, then amplitude normalization can in theory establish that the spatial configurations of neural generators differs. However, for this to be true, different spatial configurations

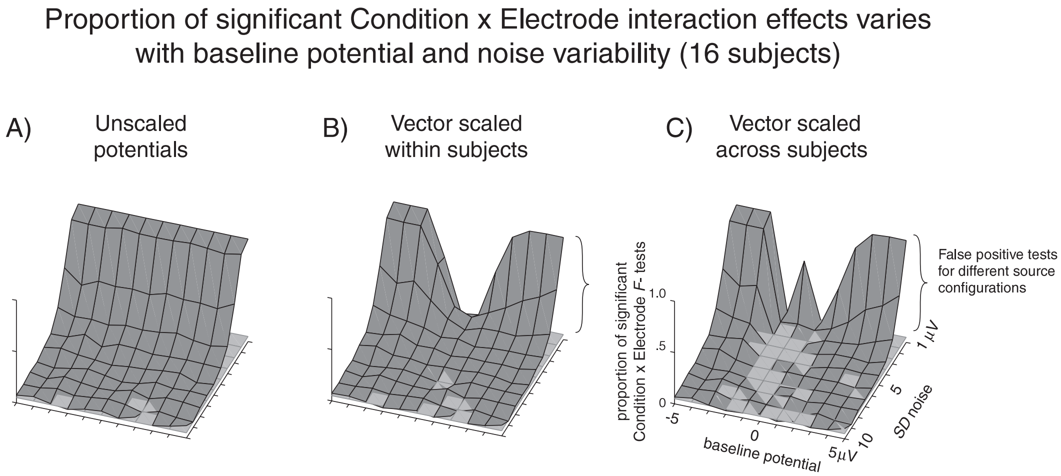


Figure 10. For testing the hypothesis that source configurations do not differ, the Type I error rates based on significant Condition × Electrode interaction effects vary as a function of vector scaling procedure, noise, and baseline potential. Two simulated poststimulus distributions with the same topographic shape (the unscaled potentials from Figure 5) were embedded in zero mean normally distributed noise for 16 subjects. The standard deviation of the noise ranged from 1 to 10 μV in steps of 1 μV. The distributions were measured against noise-free constant baseline potentials ranging from -5 to $+5$ μV in steps of 1 μV. Two hundred simulations were conducted at each combination of noise level and baseline. Repeated measures F ratios were computed for Condition (2) × Electrode (36) interaction effects for the unscaled potentials and for both types of vector-scaled magnitudes (s.m.) where $p < .05$ for the unscaled effect. The light gray cutting plane at .05 indicates ideal performance for the vector scaling procedures. A: Significant F tests are likely for unscaled potentials at all baselines if the standard deviation of the noise is below about 3 μV. B: Type I error rates for vector scaling vary with both noise and baseline and range up to 1.0. C: Type I error rates for vector scaling across subjects tend to be lower than for vector scaling within subjects, but vary with noise and baseline ranging up to 1.0.

must be defined so broadly that nothing follows about differences in the location or number of generators from differences in spatial configuration. On this definition of spatial configuration, for example, the same number of generators occurring in exactly the same locations may be different in spatial configurations, provided their relative strengths differ. Thus, although amplitude normalization can, in principle, establish that spatial configurations differ, this in fact does little to narrow down the range of possible explanations for the distributional differences in surface potentials.

Third, it was argued that quite apart from these limitations regarding what amplitude normalization can show in principle, there are fundamental problems in applying amplitude normalization procedures to measured distributions of scalp potentials in ERP research. Nonzero baseline potential distributions and noise are unavoidable and it was shown that both pose problems for the interpretation of differences between amplitude-normalized distributions in ways that they do not for unscaled potentials. Specifically, both nonzero baseline distributions and noise can lead to between-condition differences in topographic shape even when the spatial configurations of generators are identical. A series of simulations demonstrated how both nonzero baseline potentials and noise can lead to inflated Type I error rates when amplitude normalization is used to reject the null hypothesis of no difference between spatial configurations of generators. Thus, in principle, amplitude normalization is limited to ruling out the special case of an overall difference in generator strength as an explanation of distributional differences in scalp potentials: in ERP practice, it cannot do even this reliably.

Consideration of the consequences of the baseline potentials and noise shows that amplitude normalization is unreliable in its intended application in ERP research and the practice should be discontinued. However, because amplitude scaling is wholly irrelevant to establishing distributional differences in the surface potentials themselves, and even in principle provides only limited inferential traction in determining the specifics of how distributions of neural generators differ, it is not clear that much is lost by abandoning the procedure. Indeed, much is gained if experimental results that do not constitute good evidence for differences in the spatial configuration of neural generators are not treated as if they did.

REFERENCES

Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequencies and patterns. *Psychophysiology, 36*, 737–744.

Haig, A. R., Gordon, E., & Hook, S. (1997). To scale or not to scale: McCarthy and Wood revisited. *Electroencephalography and Clinical Neurophysiology, 103*, 323–325.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49*, 277–290.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, *1*, 69–82.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62*, 203–208.

Picton, T. W., Bentin, S., Berg, P., Donchin E., Hillyard, S. A., Johnson, R., Jr., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M.

J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*, 127–152.
Ruchkin, D. S., Johnson, R. Jr., & Friedman, D. (1999). Scaling is necessary when making comparisons between shapes of event-related potential topographies. *Psychophysiology, 36*, 832–834.
Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript, Princeton University.

(Received April 24, 2001; Accepted June 9, 2002)

APPENDIX

Vector scaling normalizes the overall amplitude of a distribution of scalp potentials, and in ERP practice, there are different candidate distributions for scaling: single-trial data, within-subject averages (across single trials), and grand means (across-subject averages). McCarthy and Wood (1985) caution against scaling single-trial data because of the effects of noise. The publication guidelines of Picton et al. (2000) recommend normalizing amplitude based on across-subject averages and a within-subject procedure is mentioned in a footnote. To vector scale within subjects, the distribution of each subject's mean across trials in a given condition is divided by the vector length of that same distribution. In this procedure, the scaling terms will generally vary from subject to subject. To vector scale across subjects, the vector length of the grand mean distribution across subjects is computed for a given condition and then each subject's condition mean across trials is divided by this scaling factor. In this across-subjects scaling procedure, the scaling term for each condition is the same for all subjects.

In a typical ERP experiment, EEG data are recorded for Subjects $i = 1, 2, \ldots, n$ at electrodes $j = 1, 2, \ldots, a$ in different experimental conditions. For each subject, the single-trial EEG data recorded at each electrode are typically time locked to a stimulus onset and averaged across trials in each condition. These time-domain waveforms are then reduced to the dependent measure of interest, for example, mean potential 300–500 ms poststimulus relative to the mean potential in a 100-ms prestimulus baseline at each electrode. Let X_{ij} denote the measured potential in Condition X for Subject i at Electrode j.

Vector Scaling within Subjects

First, the within-subjects vector length $|X_{i.}|$ for Subject i in Condition X is computed as

$$|X_{i.}| = \sqrt{\sum_{j=1}^{a} X_{ij}^2}$$

Then, the within-subject vector-scaled magnitude in Condition X for Subject i at Electrode j is computed as

$$\frac{X_{ij}}{|X_{i.}|}$$

In this procedure (Picton et al., 2000, p. 147, footnote), different vector lengths are obtained for each subject and each condition. Each of the original subject- and condition-specific distributions (X_{ij} for Subject i and Electrodes $j = 1, 2, \ldots a$) is scaled by its own vector length, which eliminates overall amplitude differences between subjects as well as conditions.

Vector Scaling across Subjects

Let $\bar{X}_{.j}$ denote the grand mean across subjects at Electrode j in Condition X, that is,

$$\bar{X}_{.j} = \frac{\sum_{i=1}^{n} X_{ij}}{n}$$

First, compute the across-subjects vector length, $|X_{..}|$, for Condition X as the vector length of the grand mean distribution across the n subjects, that is,

$$|X_{..}| = \sqrt{\sum_{j=1}^{a} \bar{X}_{.j}^2}$$

The across-subject vector-scaled magnitude in Condition X for Subject i and Electrode j is then computed as

$$\frac{X_{ij}}{|X_{..}|}$$

In this procedure, a single scaling factor, $|X_{..}|$, is used to scale the distributions for all the subjects in Condition X. By doing so, the grand mean amplitude is normalized between conditions without eliminating between-subjects amplitude variability. This formulation of the vector scaling procedure differs slightly from the procedure for scaling across subjects by r.m.s. amplitude recommended in Picton et al. (2000, p. 147). However, the F ratios computed in ANOVA are identical under both the r.m.s. amplitude and vector length scaling procedures, so for the intended purpose of normalizing overall amplitudes of scalp distributions, the two procedures are equivalent.