SPECIAL REPORT

Scaling is necessary when making comparisons between shapes of event-related potential topographies: A reply to Haig et al.

DANIEL S. RUCHKIN, a RAY JOHNSON, JR., b AND DAVID FRIEDMAN c

Abstract

A R. Haig, E. Gordon, and S. Hook (1997) disputed G. McCarthy and C. C. Wood’s (1985) contention that scaling should be used when assessing the statistical significance of between condition (or group) differences in the shapes of event-related potential (ERP) scalp topographies. Haig et al. based their contention upon the lack of empirical realism in McCarthy and Wood’s model of within-group ERP noise, claiming that McCarthy and Wood’s results could not be generalized to realistic ERP data. We argue, on both empirical and theoretical grounds, that Haig et al. do not make a compelling case against generalization of McCarthy and Wood’s results. Moreover, Haig et al.’s conclusion is based upon a misconception of how scaling should be used. We conclude that when a quantitative measure of differences between topographic shapes is needed, scaling is not an option—it is a requirement.

Descriptors: Event-related potential, Scaling, Topography

In recent years, researchers have relied increasingly on the topographic information inherent in event-related potentials (ERPs) to demonstrate when different patterns of brain activity occur in different conditions or groups of subjects. Differences in the shapes of scalp topographies signify differences in the patterns of brain activity. An essential part of the procedure of determining whether differences are significant between topographic shapes is the requirement to scale the data prior to testing, as described by McCarthy and Wood (1985). Recently, however, Haig, Gordon, and Hook (1997) disputed the McCarthy and Wood contention about the necessity of scaling. Given the importance of this procedure and the presence of both strengths and weaknesses in Haig et al.’s arguments, we believe it is necessary to reexamine McCarthy and Wood’s and Haig et. al’s reasoning and clarify when scaling is and is not necessary.

McCarthy and Wood (1985) demonstrated that, when an Electrode × Condition interaction in an analysis of variance (ANOVA) is used to test for differences in topographic shape, the interaction may be significant even when the topographic shapes are the same. They used a mathematical model that approximated empirical ERP data, and provided two examples of significant Electrode × Condition interactions occurring when the topographic shapes were the same across conditions. A key requirement for the interaction to be potentially misleading was that the between-condition difference in overall amplitudes had to be sufficiently large. McCarthy and Wood noted that an unambiguous interpretation of Electrode × Condition interactions requires an assessment of the contributions of both amplitude and shape differences to the interaction. Because amplitude differences only obscure the real object of interest (the topographic shape differences), the amplitude differences should be removed. The simplest way of removing the ambiguities about the basis of significant Electrode × Condition interactions is to remove the amplitude differences between the conditions by suitable scaling. McCarthy and Wood further provided an example demonstrating that scaling did not eliminate the capability to detect true differences between topographic shapes.

Haig et al. (1997) questioned the assumptions that McCarthy and Wood (1985) used in formulating their model of ERP variability. They correctly pointed out that McCarthy and Wood did not use a physiologically realistic model for the noise. Moreover, Haig et al. asserted that the problem raised by McCarthy and Wood cannot be extended to data that do not meet their assumptions, and that whether misleading Location × Group interactions occur is entirely dependent on the properties of the within-group variability. Haig et al. argued that amplitude scaling may cause misleading results. We are in limited agreement with some of Haig et al.’s views. However, with respect to Haig et al.’s first contention, we believe that they have overgeneralized about the limitations and applicability of the McCarthy and Wood model. With respect to their second contention, we believe that Haig et al. have a mis-
conception with respect to the use of scaling that has led them to “throw out the baby with the bath water.”

Critique

Haig et al. (1997) stated that “the really crucial assumption that has led to ANOVA yielding misleading results” in the McCarthy and Wood study is that McCarthy and Wood (1985) assumed that the standard deviations at each site were the same and there were no between-site correlations of the variability. Haig et al. asserted that within-group variability of real ERP data can vary across sites and there is likely to be between-site correlations in the variability. Our experience with ERP variability in cognitive experiments is consistent with Haig et al.’s assertion that the McCarthy and Wood model is not realistic. We have observed that standard deviations vary across sites, as do the ratios of the averages to the standard deviations, and that the variability is correlated across sites. However, we have also encountered numerous instances in which Location × Condition interactions were significant for unscaled data but not for scaled data.

Although McCarthy and Wood’s (1985) model of ERP variability does not have full empirical support, it does not necessarily follow that the use of such a simplified model of ERP variability leads to results that are misleading or cannot be generalized. Haig et al. (1997) argued against the generality of McCarthy and Wood’s results by devising an example in which all subjects’ topographies have the same shape, with only overall amplitude differing between subjects. For such a case, an attempt to evaluate a Location × Group interaction produces a singularity error, which is unusual for real ERP data.

No clear, compelling argument is made for why this isolated example of a singularity rules out the possibility of generalizing McCarthy and Wood’s (1985) results to real ERP data. In our view, the magnitude of the variability is its crucial aspect. When topographic shapes are nearly the same but amplitudes differ between groups (or conditions), the topographies of the mean amplitudes will appear to be nonparallel, despite the shape similarity. For example, if Fz, Cz, and Pz means show a linear gradient in each of two conditions or groups, but the means are twice as large in one condition or group as in the other, the means will give the appearance of a statistical interaction. However, in such a case, the topography (shape) does not actually differ. Whether the nonparallelism leads to a significant interaction depends on the standard errors of the means, which decrease as the statistical power of an experiment increases. If the standard errors are sufficiently small, then a significant Location × Group (or Condition) interaction will emerge. Thus, although retaining high statistical power is desirable, reaching this goal will also increase the likelihood of obtaining potentially misleading Location × Condition (or Group) interactions. Therefore, it is important to have a method, such as scaling, that will eliminate the amplitude effects from any significant location interactions.

Violations of Assumptions Underlying ANOVA and Multivariate Analyses of Variance (MANOVA)

Complications can arise when applying scaling to between-group comparisons because the scaling operation may result in a gross imbalance between the scaled variances of the groups. In such cases, the assumptions of ANOVA and MANOVA will not be satisfied by the scaled data. The problem of imbalance of variances does not arise for within-group, between-conditions comparisons.

It is possible that an imbalance of scaled variances will not be a problem in all circumstances. Havlicek and Peterson (1974) used simulations to investigate how violations of the assumptions underlying the two-sample t test affect obtained distributions of t. Their results indicated that, if inequality of variances is the only violation of assumptions, then it is likely that the results of a t-test will be valid. Further investigations along these lines are needed to establish when one can and cannot rely on the test of a scaled Electrode × Group interaction.

How To Carry Out and Interpret Scaling

Haig et al. (1997) critiqued the two main methods of scaling proposed by McCarthy and Wood (1985). The first method was an ad hoc procedure in which values were normalized by first finding the maximum and minimum amplitudes over electrodes for each condition. Then, within each condition, the minimum amplitude is subtracted from each data point, and the result is divided by the difference between the maximum and minimum. Haig et al. constructed an example in which a genuine difference in topographic shape was eliminated by this procedure. They concluded that scaling by the ad hoc maximum/minimum procedure could produce misleading results. We are in complete agreement with their conclusion on this point.

The second method of scaling proposed by McCarthy and Wood (1985), vector scaling, is achieved by representing the ERPs as vectors in a multidimensional signal space. The ERP amplitude at each electrode corresponds to a coordinate of the vector. Contrary to Haig et al.’s (1997) assertion that this procedure is arbitrary and ad hoc, vector scaling has a firm theoretical foundation in the mathematics of vector analysis, signal space, and multidimensional analysis, as can be found in the numerous texts cited by McCarthy and Wood. There is a one-to-one correspondence between a topographic shape and the orientation of the vector that represents that shape in signal space. The amplitudes of the data are embodied in the length of the vector. When topographic shapes are different, their corresponding vector representations always have different orientations. When topographic shapes are the same, their corresponding vector representations always have the same orientations (the vectors then differ only in length). The Electrode × Condition analysis of an ANOVA on vector-scaled ERPs in effect tests whether the across-subjects average vectors have the same orientation.

Haig et al. (1997) provided an example of vector scaling. Consistent with McCarthy and Wood (1985), their example demonstrated that vector scaling preserved a genuine difference between topographic shapes. Haig et al. then used the scaled data to interpret the between-group differences in topography. This latter step was not recommended by McCarthy and Wood and should not be carried out, because the interpretation of the effect of condition or group on topography may be different for scaled and unscaled data (as it was in their example). That is, scaling is to be used only for assessing whether there is a statistically reliable difference between the shapes of two or more within-group topographies, not for interpreting results. The unscaled interaction is not inherently invalid, but it is an invalid basis for a conclusion about differences in the shape of the distributions.

Conclusion

Scaling should be used only to determine whether there is a difference between topographic shapes. That use was the only one that McCarthy and Wood (1985) recommended. An interpretation of
the meaning of a topographic difference should be based on unscaled data. The fact that unscaled data should be used to interpret topographic differences does not mean that scaling should not be used to determine the statistical significance of a topographic difference.

If differences in topographic shape are not an issue, then scaling is neither necessary nor desirable when testing for Electrode × Condition (or Group) interactions. The fundamental reason for scaling amplitude data is to make a quantitative determination of whether topographic shapes are different. Such tests require that amplitude differences be removed prior to the comparison, because failure to do so means that a meaningful quantitative comparison of shapes cannot be made.

The ability to determine the presence or absence of differences in topographic shapes is crucial to using the spatial information inherent in ERPs. If there is a significant difference in topographic shape between conditions, then it can be inferred with high confidence that the underlying patterns of brain activity are different. However, the detailed structure of the patterns of brain activity cannot be inferred with the same high degree of confidence. Whereas differences in the shapes of scalp topographies can occur only if the patterns of brain activity differ, the detailed nature of the brain activity differences cannot be determined uniquely from scalp ERP data alone. Nevertheless, in suitably designed experiments, the detection of topographic shape differences can be used to test theories and resolve issues regarding the underlying patterns of brain activity in different experimental conditions. The issue of misleading interactions comes about because ANOVA is a convenient way of assessing the statistical reliability of a quantitative comparison of shapes. However, whether ANOVA or other approaches are used (e.g., Hotelling’s $T^2$), there is always the fundamental necessity of scaling the data when topographic shapes are compared.

REFERENCES


(Received June 25, 1998; Accepted May 6, 1999)