Projection onto Centroids Difference Vectors: A New Approach to Determine Between Group Topographical Differences, Applied to P3 Amplitude in Schizophrenia

Albert R. Haig* and Evian Gordon*

Summary: A common problem in electrophysiological research concerns between group analysis of EEG and ERP topography. This paper proposes a new technique for determining whether or not a significant difference exists between multiple scalp site measurements from two groups. The method requires no a priori assumptions about the data and is thus ideal for exploratory data analysis, and it also requires that only one statistical test need be performed (significantly reducing the possibility of type I errors). The Projection onto Centroids Difference Vectors (PCDV) method involves deriving a measure from each individual of whether their measurements across sites are closer overall to the mean measurements of the rest of their experimental group, or to the other group. These measures from each individual are then compared between groups using a Student's t test, which indicates whether one group's data is significantly spatially different from the other. In this study we describe the method in detail and apply it to both simulated data and to real auditory P3 data in unmedicated, medicated schizophrenics and matched normal controls. The PCDV method was also compared with statistical probability mapping (SPM). The PCDV method revealed the differences between the normal and patient groups more unambiguously than SPM, and the simulated data revealed that it was not liable to type I errors. PCDV provides an appropriate method for testing any between group EEG and ERP topographical differences.

Key words: ERP; Statistical analysis; SPM; PCA; P3; Schizophrenia.

Introduction

Statistical analysis of data from multiple scalp sites

One of the most common approaches in clinical EEG and ERP research involves statistical comparison of data from different subject groups, where measures have been obtained from multiple scalp sites. The most common method of preliminary data analysis, statistical probability mapping or SPM (Duffy et al. 1981), involves performing separate statistical tests for a difference in means at each electrode site, and then displaying the resultant p-values in an interpolated topographic map. Whilst this is a convenient and useful exploratory method, a limitation with this approach is the problem of multiple comparisons (Oken et al. 1986; Duffy et al. 1986; Duffy 1988), which requires caution in interpretation of the results, because of the likelihood of a type I error. In addition, a real difference between the groups may not necessarily show up as significant at any individual electrode site (type II error). It is often desirable to know whether overall there is a difference in the scalp measurements between two groups, and it is difficult to answer this question from SPM.

There have been a number of alternative approaches to this problem, some of which (from averaging electrode sites within regions to planned comparisons using MANOVA) require a priori hypotheses about spatial differences between the groups being compared. However, studies of EEG and ERP are still generally exploratory (Duffy et al. 1986), and it may therefore not be possible to make any firm hypotheses about the spatial distribution of the differences. In addition, only examining certain
spatial hypotheses determined *a priori* by the researcher, risks missing important real unexpected differences.

One approach which is not based on *a priori* hypotheses about the spatial distribution of differences which may exist, is principal components analysis or PCA (Hunt 1985; McGillem and Aunon 1987; Duffy et al. 1992), which provides a means of reducing the dimensionality of the data. This paper proposes a new technique, Projection onto Centroids Difference Vectors (PCDV), which is simpler than PCA and has advantages over SPM. PCDV provides a means of reducing the number of variables being tested from \( n \) (where \( n \) is the number of electrode sites) to just one (reducing the likelihood of a type I error), with this one test evaluating whether or not the members of one group are significantly spatially shifted relative to the other. It is helpful to consider the problem in terms of vectors.

A vector representation of the problem

If there are \( n \) electrode sites from which measures are derived in each subject, then the \( n \) measures from each subject can be considered as a vector in an \( n \)-dimensional space. These spatial vectors, which will be referred to as the signal vectors, each belong to one of two groups. Thus there are two clusters of vectors. All of the techniques used in analysis are concerned with determining whether there is a significant directional shift in one cluster of vectors compared to the other. To put this in non-vector language, the interest in all the statistical methods is whether the mean of the measures in one group are increased or decreased at various electrode sites compared to the other, and by how much. Of course, the clusters could differ from each other in more complex ways than in terms of just a directional shift, but none of the statistical methods commonly employed can detect this.

PCA

PCA provides a means of reducing the dimensionality of the data and therefore of reducing the number of variables to be tested. PCA is performed on the signal vectors from the normal subjects (Duffy et al. 1992). Often, most of the variance of the signal vectors will be accounted for by only a relatively few factors. The signal vectors can be projected onto just those factors which have been determined to be 'significant' by one method or another (Duffy et al. 1992). That is, these factors span a subspace which contains most of the 'relevant' signal structure. Differences between the groups can then be tested within this subspace.

Unfortunately PCA still has some drawbacks. Since there will generally be a number of 'significant' factors, the number of variables, while less than \( n \), may still not be insignificant. Furthermore, it is unlikely that a spatial difference between normal and clinical groups will be represented by just one factor. Therefore testing along each factor independently will be sub-optimal and may risk missing a real difference. Even if such a difference is detected by means of PCA, disentangling its true spatial distribution may turn out to be difficult. It is also possible, although unlikely, that the discarded subspace (that spanned by the remaining factors which have been determined not to be 'significant') may be that in which a subtle but significant difference between the clinical groups exists.

More complicated techniques, such as linear discriminant function analysis (Duffy et al. 1992), can be employed to avoid some of these problems with PCA. However the alternative approach of Projection onto Centroids Difference Vectors, tackles the issue of an overall spatial difference between the groups directly.

Projection onto Centroids Difference Vectors

Often, the mean measurements across sites from one group will differ in some way from the mean measurements across sites in the other. In vector terms, the mean measurements of one group are represented by the centroid vector of the individual vectors from that group. Thus the centroid vector from one group will most likely not be identical to the centroid vector from the other. That is, the mean measurements of one group will be slightly spatially shifted relative to the other. However, such a shift may not be significant, and may be due to chance. Consider now one of the individual vectors from one group, and the centroid vector of the remaining vectors in this group, excluding this one vector under consideration. We can now ask whether the single vector under consideration is closer to the centroid of the remaining vectors in the group, or closer to the centroid of the other group. If there is a real spatial difference between the groups then we would expect on average the single vector under consideration to be closer to the centroid of the remaining vectors in its group, than to the centroid of the other group. If there is no real spatial difference between the groups, we would not expect on average the single vector to be any closer to the centroid of the remaining vectors in its group, than to the centroid of the other group. If there is no difference between the groups, or in other words if the vectors in both groups are drawn from the same underlying vector probability distribution, then the other vectors in the group will not be any closer on average to the individual vector in question than the vectors from the other group, and hence the centroid of the other vectors in the group will not be any closer on average to the individual vector in question than the centroid of the other group. On the other hand,