Applying Principal Components Analysis to Event Related Potentials: A Tutorial

Running Head: PCA Tutorial

Dien, J. (2012). Applying Principal Components Analysis to Event Related Potentials: A Tutorial. Developmental Neuropsychology, 37(6) 497-517.

This is an Accepted Manuscript of an article published by Taylor & Francis in Developmental Neuropsychology on 8/14/2012 available online: https://doi.org/10.1080/87565641.2012.697503.

Address for correspondence: Joseph Dien, Maryland Neuroimaging Center, University of Maryland, 8077 Greenmead, College Park, MD 20742, USA.

Phone: 202-297-8117. E-mail: jdien07@mac.com. URL: http://joedien.com.

ABSTRACT

Principal components analysis (PCA) has attracted increasing interest as a tool for facilitating analysis of high-density event-related potential (ERP) data. While every researcher is exposed to this statistical procedure in graduate school, its complexities are rarely covered in depth and hence researchers are often not conversant with its subtleties. Furthermore, application to ERP datasets involves unique aspects that would not be covered in a general statistics course. This tutorial seeks to provide guidance on the decisions involved in applying PCA to ERPs and their consequences, using the ERP PCA Toolkit to illustrate the analysis process on a novelty oddball dataset.

Key Words

Event Related Potentials, Principal Components Analysis

INTRODUCTION

Event-related potentials (ERPs) is a well-established method for measuring the electrical signals from the brain. While it has very high time resolution, on the order of milliseconds, it has very poor spatial resolution. Thus, voltage fields from a small part of the brain may volume conduct through the head and be detectable at many or most of the recording electrodes on the head. Not only does it become challenging to determine which part of the brain a signal arises from, the signals from multiple parts of the brain can end up overlapping in ways that makes it difficult for the human eye to distinguish them.

Consider, for example, the waveforms presented in Figure 1. How many ERP components (bursts of voltage) contributed to these recorded waveforms? While one can clearly distinguish at least three peaks at about 200, 300, and 600 ms, past that determination it becomes somewhat ambiguous. Further complicating matters is that electrical fields are by their nature dipolar, producing a positive voltage on one side of the head and a negative voltage on the other side of the head. It is therefore unclear whether the negativity at 600 ms in the top channel is a different ERP component from the positive peak seen in the other two channels or whether it might just be the negative side of the same ERP component.

A method for dealing with these issues is principal components analysis or PCA (for reviews, see Gorsuch, 1983; Harman, 1976) which for many years has been used to statistically decompose the ERP waveforms into their constituent building blocks (Dien & Frishkoff, 2005; Donchin & Heffley, 1979; Glaser & Ruchkin, 1976; Möcks & Verleger, 1991). With the increasing use of high-density recording montages, the resulting increase in the volume of data has made the need for this procedure even more pressing.

Thus, the primary utility of PCA with ERP datasets is to: 1) help identify the constituent components of the ERP, 2) provide a dependent measure of these components for inferential testing, and more recently 3) help improve the localization of the ERP sources (Carretie et al., 2004; Dien, 2010a; Dien, Tucker, Potts, & Hartry, 1997; Dien, Frishkoff, Cerbone, & Tucker, 2003; Dien, Spencer, & Donchin, 2003; Dien & O'Hare, 2008; O'Hare, Dien, Waterson, & Savage, 2008; Pourtois, Delplanque, Michel, & Vuilleumier, 2008). The first two roles are especially useful for developmental populations where the data can be especially noisy due to movement artifacts and the third role can be especially helpful for neuropsychological populations where the locus of neurological deficits can be a primary concern.

The usage of PCA for ERP datasets has remained largely static over the four decades since it was first introduced (Donchin, 1966; Ruchkin, Villegas, & John, 1964). For the most part it has been limited to applying the standard procedures according to common conventions, such as the use of the Varimax rotation (Kaiser, 1958). This state of affairs is unfortunate for two reasons: First, the common conventions were largely developed for questionnaire datasets and therefore do not take the unique characteristics of ERP datasets into account. Second, statistical science has progressed greatly since those early days and it would be desirable to take full advantage of these advances.

The chief obstacle to making full use of PCA has been the need to rely on commercial statistics packages. Since these packages were not written with ERP datasets in mind, even basic analyses are cumbersome to implement and more advanced procedures have largely been neglected. The ERP PCA Toolkit (Dien, 2010b) was developed to address these difficulties[1]. By implementing the statistical algorithms within a widely used programming environment (Matlab), it has been possible to package the procedures in a fashion that makes use of current statistical advances, including some not yet available in commercial packages, and to optimize them for the special needs of ERP datasets.

This tutorial will provide a guided tour through the PCA process, relying on the ERP PCA Toolkit (version 2.14). This paper is not, however, meant to be a tutorial on how to use the Toolkit per se (see instead, Dien, 2010b). This tutorial is also not intended to be an introduction to ERPs for novices but rather a guide for more experienced ERP researchers who are interested in learning how to utilize PCA. This tutorial will focus on the decisions that need to be made and their consequences. For information about the statistical mechanics of the procedure and how they relate to ERPs, see (Dien & Frishkoff, 2005).

THE DATA

The example dataset (Figure 1) is a P3 Novelty dataset that has been presented elsewhere (Dien et al., 2003; Spencer, Dien, & Donchin, 1999; Spencer, Dien, & Donchin, 2001) using an earlier version of the procedure. It contains fifteen participants. They participated in an oddball task in which they responded to a rare target tone. Periodically, a novel environmental sound such as a dog bark would be presented instead of a tone. The three cells of the dataset consist of standard tones, target tones, and novel sounds. The advantage of using this dataset is that it provides a simple task with well-understood ERP components and strong experimental effects, which allow the reader to easily ascertain the degree of success of the PCA procedure. Using the same dataset will also allow for some clear examples of how continuing to refine the procedures has had a meaningful impact on the results.

The first step in the PCA process (see Figure 2 for a list of the steps) is to decide what to include in the analysis. The typical ERP dataset has four sources of variance: time points, channels, subjects, and conditions. Decisions must be made about each of these.

Regarding time points, one can potentially restrict the time points to those of interest. An advantage to doing so is that irrelevant time points can only slow down the analysis and inject additional noise and thus imprecision into the process. It has been demonstrated, for example, that if one is only interested in the very early brain stem auditory potentials, one may obtain better PCA results if one leaves out the later time points as they contain much larger amplitude activity that drowns out the activity from the brain stem potentials (Curry et al., 1983).

What is not as obvious is what constitutes irrelevancy. A basic principle of PCA of ERP data is that it is as much about characterizing the ERP components of non-interest so that they can be separated out as it is about characterizing the ERP components of interest. Thus, if the P300 is the only ERP component of interest, one would nonetheless wish to also include all the time points containing the N200 since it tends to overlap the P300 and will need to be removed. Additionally, it can be helpful to include the prestimulus baseline period as a comparison point to the activity of interest, both to better determine what constitutes zero voltage and to provide an estimate for how much noise remains after the PCA process. In addition, including the baseline period can help reveal the presence of prestimulus activity, such as the contingent negative variation or CNV (Walter, Cooper, Aldridge, McCallum, & Winter, 1964) which could otherwise unwittingly be affecting the analysis (via application of the baseline correction procedure). In general, it is recommended that one retain the entire epoch, including the baseline period, unless there is reason to do otherwise.

Similar issues apply to channels. In principle, one could restrict the PCA to just the channels of primary interest. Again, however, one would also wish to include channels that characterize the overlapping ERP components of non-interest. As it turns out, because volume conduction of voltage fields ensures that ERP components will affect all the electrodes to at least some extent (except those along the zero voltage isopotential line), in the channel domain all ERP components overlap with all other ERP components. For this reason, the recommendation is to include all channels in the PCA process.

With regard to subject variance, in principle one could eliminate it by using a grand average instead of subject averages (and then applying the resulting factor loadings to the individual subject averages). Doing so would have the advantage of improving the signal-to-noise ratio by averaging out some of the residual noise in the subject averages. BESA's PCA option uses this very approach. On the other hand, individual difference variance can help the PCA distinguish between ERP components. If, for example, if the N2 and the P3 vary in their relative size across subjects, with some subjects having a large N2 and a small P3 and vice versa, then the PCA could use this difference to separate them. When put to the empirical test (Dien, Khoe, & Mangun, 2007), it was indeed found in a simulation dataset that using a grand average was less effective. It is therefore generally recommended to include all the subject averages, rather than using a grand average.

Finally, condition variance may require the most thought. It is generally best if the data contain the same mix of ERP components throughout. If, for example, the data contain both auditory ERPs and visual ERPs, the PCA will have a tendency to attempt to extract auditory ERP components from the visual trials and visual ERP components from the auditory trials. In general, this situation can yield degraded results. No universal rule can be provided on this point as every dataset is unique and requires a careful balancing of competing considerations. All that can be said is that if there is reason to think that different conditions contain distinct mixes of ERP components, it may be best to subject them to separate PCAs (which is to say not just somewhat different but very very different, as in visual versus auditory ERPs). In the example dataset, there is no reason to think that the different conditions (standard, target, and novel) will contain dramatically different ERP components, although it is certainly expected that their amplitudes may differ. Again, visual inspection can be used to make this determination.

INITIAL FACTOR EXTRACTION

The next step is to begin the PCA by choosing which source of variance to focus on. In a conventional PCA of ERP data, one source of variance is emphasized by using it as the variables and the others are used as the observations. For example, in a temporal PCA, the time points are the variables and the combinations of the other sources of variance are the observations (so, for example, one observation will be a particular channel from a particular subject from a particular condition). While it might seem odd to mix these different sources of variance like this, all of them are legitimate ways of dissociating two ERP components. All the sources of variance contribute to the solution but the one that is chosen to be the variables has the greatest influence (see Dien & Frishkoff, 2005). It normally would take some effort to arrange the data manually but the ERP PCA Toolkit does so automatically.

Temporal PCA generally produces better results than spatial PCA due to superior component separation in the temporal domain (Dien, 1998a). Essentially, whereas ERP components need not overlap in the temporal domain (they can occupy different time points), volume conduction ensures that they will always overlap in the spatial domain and hence be more difficult to separate.

The major case where one might not wish to use temporal PCA is where the ERP component has a variable time course. In a temporal PCA, the factors are defined in terms of a fixed time course (as described by the factor loadings). The only way to allow for an ERP component to have a substantially variable latency in a temporal PCA is to assign multiple factors, each representing a different latency, complicating interpretation (Dien, Spencer, & Donchin, 2004; Möcks, 1986). For this reason, it is generally recommended to use temporal PCA, unless there is reason to believe that substantial latency jitter may be present. Simple inspection of the grand average waveforms is usually sufficient to make this determination. As seen in Figure 1, while the P300 (the positive peak at about 300 ms) can indeed have a quite variable latency, in the present case the task was quite simple and so its latency was reasonably constant across conditions. ERP components other than the P300 tend to generally have more subtle latency changes. In such a case, the factor waveform will reflect the central tendency of the latencies with the subtle differences being discarded. Thus, subtle latency changes are not an issue for temporal PCA as far as component identification goes but spatial PCA should be used if the intention is to test for latency changes.

Another choice to be made at this point is rather technical and has to do with the nature of the relationship matrix. Essentially, the initial step in a PCA is to generate a matrix summarizing the interrelationships between the variables (correlation, covariance, or sums-of-squares). Students in statistics courses are typically taught to use correlation matrices, which is appropriate for datasets where the metrics of the variables are incongruent (as in years of age and feet in height). Commercial statistics packages tend to make correlation matrices the default for similar reasons. For ERP data, where all the variables are in microvolts, the relative scales are meaningful (twice the value in one time point compared to another is indeed twice the quantity), a covariance matrix is the most logical choice and indeed yields better results (Curry et al., 1983; Dien, Beal, & Berg, 2005; Donchin & Heffley, 1979; Kayser & Tenke, 2003; Möcks & Verleger, 1991). This is a good example of why it is advisable to consider carefully the special characteristics of ERP data and how they may affect a PCA.

Yet another consideration is the reference scheme used for the data (for a discussion of reference issues, see Dien, 1998b). For a temporal PCA, the reference scheme will make little difference because both correlation and covariance matrices are both mean-corrected (the mean of the variable is subtracted, centering it on zero). In effect, when performing a temporal PCA, the data is average referenced. The reference scheme will affect spatial PCA however, contrary to statements otherwise (Pourtois et al., 2008), since there the variables are the channels and so the zero-centering is occurring along the timepoints (as well as the subjects and cells).

The effect of reference scheme upon spatial PCA can be thought of in this manner. PCA operates on the relationship matrix, such as a covariance matrix. In such a matrix, the diagonal contains the variances and the off-diagonal entries contain the covariances. For a 129-channel montage, there will be 129 diagonal elements and 16,512 off-diagonal elements, so the effects of the covariances far outweigh the effects of the variances. If we imagine, for didactic reasons, an extreme case where there is an ERP feature that is only present in one channel, it will appear in only that channel's variance entry (a single number in the diagonal of the covariance matrix). If we chose that channel to be the reference, then that channel would be subtracted from all the channels (including itself), with the result that all the other channels would now covary with each other to some extent whereas the reference channel would now have zero variance (be a flat line). This feature would now be represented in most of the off-diagonal elements and have a strong effect on the final PCA result.

Figure 3 illustrates how the reference choice can yield differing results. The present dataset was subjected to a four-factor PCA using Varimax rotation, with mean mastoids reference (bottom of the montage) and on the left a mean cz-pz reference (top of the montage) on the right. The figure shows the mapping of the factor loadings, rescaled to microvolts (see Visual Inspection section) by multiplying each by the variable's standard deviation. Note that since the goal was to dramatize how the reference site can make a difference, this example was purposely conducted in a manner intended to yield maximal instability of the results (too few factors, Varimax rotation, and extreme reference choices).

So which reference to choose, since reference choice is by its nature an arbitrary choice? If there is a whole-scalp sampling by the electrodes, then the most neutral choice is the one that best approximates the true distribution of the voltage fields, the average reference (Dien, 1998b). As such, it will minimize distortions caused by under or overrepresenting an ERP feature in the analysis. This choice is also preferable since temporal PCA also is effectively average referencing the data. Such considerations need to be balanced with a need to maintain consistency between the PCA and the rest of the analysis procedure.

FACTOR TRUNCATION

The next step has been the subject of some debate in the ERP literature. The goal of a conventional PCA is both data description and data reduction, that is to describe the structure of the independent variables and to summarize them with a smaller number of variables. In the case of ERP data, to determine how the measured voltages can be described in terms of a small set of ERP components and to obtain direct measures of them rather than having to analyze hundreds of measurements (at all the time points and at all the channels). The initial factor extraction serves the function of data description. The goal of data reduction is met by factor truncation, which is the discarding of the smaller factors that are deemed to be mostly noise or otherwise of non-interest.

There are a number of methods in common use for determining how many factors to retain when truncating. There has not yet been a systematic comparison of these methods in the ERP domain. The recommended method, implemented in the ERP PCA Toolkit, is the Parallel Test (Horn, 1965). The logic behind this test is that if one performs PCA on an entirely random dataset (in this case, a set of computer-generated random numbers with the same dimensions as the dataset of interest), one will obtain a set of randomly sized factors that when arranged in order of size will form a smooth slope. If there are real signals embedded in the noise, then they will emerge as a set of factors that are larger than one would obtain from noise alone. In the Parallel Test, one merely compares the results from the PCA of the ERP dataset to those obtained from a PCA of random data. Note that the present implementation differs from that the modified Parallel Test suggested in an earlier paper (Dien, 1998a). For the example dataset, the Parallel Test suggests that nine factors be retained, accounting for 95.43% of the variance (Figure 4).

It should be noted that it has been argued that truncation should be avoided, retaining the entire set of factors (an "unrestricted" solution), on the basis that one obtains more accurate results (Kayser & Tenke, 2003; Kayser & Tenke, 2006). While it may be true that doing so improves the data description (this author has not yet had the time to examine this issue in depth), doing so does not address the data reduction goal. It is argued that the resulting plethora of factors poses serious multiple comparison risks (Dien, 2006). It may be that this concern can be addressed with an appropriate procedure. Certainly the analyses put forward in support of unrestricted solutions are quite intriguing. For now, this author considers the matter to be up for debate and in need of further examination and discussion.

FACTOR ROTATION

The next step is to rotate the factor structure to an interpretable simple structure. The essential problem with the initial results of the PCA is that the factors are calculated sequentially, with each one accounting for the most possible variance that hasn't already been accounted for by a prior factor. Since the best way to accomplish this is for the factor to be correlated with as many variables as possible (each variable contributes a portion of the total variance and so the more variables, the more variance), the resulting factors will generally not have a simple relationship with the ERP components (e.g., a factor might be a portion of the P1 plus a portion of the P2 plus a portion of the N2 plus a portion of the P3 and so forth).

The purpose of rotation is to rearrange the factors such that they have a simpler, more interpretable relationship with the underlying patterns (ideally, one factor per ERP component). This is possible because while the number of factors is fixed, their relationship to the underlying ERP components is not. For example, given two ERP components, one might have one factor per ERP component, one might have one factor representing what the two ERP components share in common and another representing how they differ, or anything in between those two extremes. Rotations seek to find the simplest combination, ideally one factor per ERP component, using various rules of thumb. Note, for example, how the first factor of the unrotated solution involves most of the time points and how the third factor of the unrotated solution is combining three apparently distinct ERP peaks whereas the rotated solution has peaks that are more uniphasic and discrete (Figure 5). A uniphasic pattern is a more plausible solution since the general consensus of the ERP community is that most ERP components are uniphasic, which is to say they have only a single peak or dip as the case may be.

The most important distinction between rotations is whether they are orthogonal or oblique, which is to say whether the factors are allowed to be correlated or not. The most common rotation, Varimax, is an orthogonal rotation, which therefore forces the factors to be uncorrelated. While it is often stated that this characteristic of orthogonal rotations is an advantage as it simplifies analysis, this author would argue that this is simply wishful thinking. If the two ERP components (e.g., N2 and P3) are correlated in actuality, the resulting Varimax factors will indeed be uncorrelated, but only at the cost of no longer being N2 and P3 factors (assuming the PCA was otherwise successful). For example, the way in which Varimax forces a positively correlated N2 and P3 to be uncorrelated is by literally stitching together hybrid factors (e.g., one might have an N2 factor glued to enough of the P3 to account for the shared variation and/or vice versa). Stating that the resulting factors as being an N2 and a P3 factor would then be misrepresenting the results when in fact they are most likely an N2-P3 factor and a P3-N2 factor, which is to say questionable hybrids of uncertain provenance. Thus, one might find that a factor reflects an experimental effect and yet be uncertain as to whether it was the N2 or the P3 that should properly receive credit as both factors would contain both ERP components, defeating the entire point of conducting the PCA in the first place.

An example of this kind of hybrid ERP factor situation can be seen in Figure 6, where, at least according to the Promax Rotation, the ERP activity described by Factors 1 and 2 are correlated r=0.582. The Promax rotation then goes on to suggest that Varimax has forced them to be uncorrelated by grafting a portion of ERP Component 2 onto ERP Component 1 (the ascending arm of the waveform) and a portion of ERP Component 1 onto ERP Component 2 (the descending arm of the waveform), thus producing hybrid factors. Only allowing the factors to be correlated has allowed this artificial constraint to be removed, resulting in factors that correspond to the underlying ERP components more purely. Although it is not possible to state with certainty whether the Promax solution for this ERP data is in fact more accurate barring omniscience, testing of simulation datasets (where the true answer is known) bears out this reasoning (Dien, 2010a; Dien, 1998a; Dien et al., 2005; Dien et al., 2007).

A counter-argument sometimes encountered by the author is that perhaps ERP components are not often correlated, in which case this concern would not be a practical obstacle. In truth, ERP components are nearly always correlated in PCA. For example, in a temporal PCA, the ERP components are correlated not just by condition variance but also by spatial and subject variance (as all three types of variance are represented in the observations). Spatial variance in particular nearly ensures that ERP components will be correlated in a temporal PCA, because the only way for them not to be correlated is if they are located at nearly right angles on the head, such that an equal number of electrodes have the same polarity voltage from the two ERP components as have opposite polarity voltage (Dien, 2010a), so that the polarity of one ERP component at an electrode does not predict the polarity of the other ERP component, all things being equal. Likewise, temporal variance nearly ensures that ERP components will be correlated in a spatial PCA (if the two ERP components occupy different time points, for example, then they will be negatively correlated since an observation, which is to say a time point, that contains one ERP component will not contain the other ERP component; zero correlation would require there to be as many time points where they overlap as time points where they do not, all things being equal, so that the presence of one does not predict the presence of the other). Indeed, simulation studies (Dien, 2010a; Dien, 1998a; Dien et al., 2005; Dien et al., 2007) and real data studies (Dien et al., 2003) clearly indicate that Varimax yields inferior results compared to oblique rotations. It is therefore somewhat difficult to defend continued use of Varimax rotations for ERP data.

Turning to oblique rotations, where the factors are allowed to be correlated, there are two chief options. The first is Promax, which is a Varimax rotation where the orthogonality of the initial solution is then relaxed in a second step (Hendrickson & White, 1964). While there is a parameter called kappa that controls the degree of relaxation, a simulation study of ERP data suggests that it has little practical effect (Dien, 2010a). The second is Infomax (Bell & Sejnowski, 1995), a type of Independent Components Analysis (Hyvärinen, Karhunen, & Oja, 2001), where the rotation is performed on the basis of higher moments like kurtosis in addition to correlations (although in practice the correlations are usually removed prior to the rotation so that only the higher moments determine the rotation). Although ICA comes from a very different tradition than PCA, in ERP research it is generally applied after a PCA and is therefore in effect being used as a PCA rotation, albeit with distinct properties. ICA is generally not available in commercial statistics packages but has been made widely available to EEG researchers via the EEGlab software package (Delorme & Makeig, 2004).

Simulation studies suggest that Promax is most effective for temporal PCA and Infomax is most effective for spatial PCA (Dien, 2010a; Dien et al., 2007). The reason is that Promax rotates such that the factor loadings are as extreme as possible, either zero or very large. Such a rotation is appropriate in the temporal domain, wherein the time course for an ERP component typically starts at zero, then abruptly rises to a large amplitude, and then drops back to zero for the rest of the epoch. In contrast, due to volume conduction, ERP components tend to be non-zero at virtually every electrode, except along the zero isopotential line. Infomax rotates in order to maximize non-normality of the factor scores (following the logic that the Central Limit Theorem indicates that factor scores that reflect a mixture of ERP components will be more normally distributed than those that do not) and to minimize the ability for the higher exponents of each factor score to predict the other factor scores (independence), a criterion that lacks the drawbacks of the Promax approach in the spatial domain.

VISUAL INSPECTION

The next step is to visually inspect the results. While it has been common to display the raw factor loadings, this is again a practice that has been inappropriately carried over from other areas of psychology. Factor loadings are correlations and hence unitless whereas ERP waveforms are in microvolts, and so they cannot be directly compared to each other. To put it in other terms, since computing correlations involves standardizing the variables, the minor time points have been exaggerated and the major time points have been diminished. See Figure 7 for a sample grand average waveform that has similarly been standardized in a manner consistent with a temporal PCA (by dividing the microvolt values by the standard deviation of the time point across the entire dataset, meaning all the waveforms from all the subjects and conditions lined up as the observations) to see how viewing it in this form has distorted it. In order to make factor loadings comparable to the original waveforms, they should be rescaled to microvolts by converting them into covariance loadings (Dien, 2006), which is done by multiplying the correlation factor loadings (unitless) with the standard deviations of the variables (microvolt units) to produce covariance loadings (microvolt units).

More generally, one can conceptualize the PCA process as splitting the original rectangular matrix of voltage measurements into two separate matrices, the factor loadings and the factor scores. One could reverse the process by matrix multiplying the factor loadings and the factor scores to reproduce exactly the original data. One could also selectively reverse the process for just a single factor, thus generating the portion of the data accounted for by a single factor (Factor 6). This is what the ERP PCA Toolkit does, automatically regenerating the waveforms and the scalp topographies for each factor. This facility is very helpful for interpreting factor results, as in determining whether the scalp topography differs between conditions, a sign of multiple ERP components being agglomerated into a single factor (see Figure 8a for an example).

TWO-STEP PCA

Even after the rotation, the PCA may not be finished. The essential issue is that temporal PCAs have difficulty separating ERP components that have similar time courses and spatial PCAs have difficulty separating ERP components with similar scalp topographies. Thus, the temporal PCA of the example data did not separate the P3a and the P3b components (Figure 8a). In a two-step PCA (Dien et al., 2003; Spencer et al., 1999), the factor scores from the first step are subjected to the complementary type of PCA (e.g., spatial after an initial temporal PCA) in order to separate out these confounded ERP components. It is generally recommended to do a separate second step PCA for each of the initial first step factors. Thus, in the present example one would conduct nine separate spatial PCAs, one for each of the initial temporal factors. Of course, the caution given earlier that temporal PCA will tend to handle substantial temporal jitter by splitting the ERP components into multiple factors still holds (they can be diagnosed by their identical scalp topographies).

An important question at this point is how to decide whether a factor represents a distinct ERP component. Let us say, for example, that according to the PCA there is voltage activity that appears to have a similar scalp topography at two different latencies - could they be two different ERP components or just a single one that has been split into two due to something like the presence of substantial temporal jitter. The answer is that there is no cut and dry answer to this. The question of componentry is a basic challenge for ERP researchers using any method of inspection, not restricted to PCA (for a discussion of this issue, see Sutton & Ruchkin, 1984). PCA merely highlights potential ambiguity. That said, in the experience of this author, this situation is relatively infrequent.

Figure 9 displays the results for the example analysis. The parallel test suggested three factors to be retained for the second, spatial Infomax step. The P3 factor has been split into separate P3a and P3b factors, as suggested by the scalp topographies in Figure 8a. The scalp topography of the P3b factor contrasts with a later latency factor that seems to correspond to the Positive Slow Wave (Ruchkin, Munson, & Sutton, 1982; Ruchkin & Sutton, 1983; Sutton & Ruchkin, 1984) that has previously been separated from the P3b by PCA (Squires, Donchin, Herning, & McCarthy, 1977). Overall, the results are consistent with the original reports (Spencer et al., 1999; Spencer et al., 2001) with some differences attributable to the ongoing refinements in the procedure. The most notable difference is that the Positive Slow Wave scalp topography now clearly differs from the P3b scalp topography whereas in the original report the same spatial factor (SF1) accounted for both of them. This change is likely due to the use of separate spatial PCAs for each temporal PCA that freed them to be different rather than using a single spatial PCA across the entire epoch that biased the solution towards a single spatial factor that accounted for both features. This case therefore serves as an example of how continuing to refine the procedure can yield meaningful benefits.

ANOVAs

The final step is to subject the factors scores to inferential statistics to determine which effects are reliable. A chief issue is that PCA can generate hundreds of factors, especially if one uses unrestricted solutions or even restricted two-step PCAs. Given that a p-value of .05 can result in a false alarm for one out of twenty of them, this can be a serious issue. The following procedure is recommended. First set a threshold for factor size, below which it will not even be considered (e.g., .5% of total variance). This will screen out the great majority of junk factors that reflect uninteresting noise. The second step is to set aside factors that can be identified as being of *a priori* interest (such as a factor from a language study that has the same scalp topography and latency as the N400). Everything else should be deemed exploratory and controlled for multiple comparisons, as by a Bonferroni test. When approached in this manner, experience of the author has shown that PCA can retain good statistical power for identifying potential new ERP components while maintaining statistical rigor. In the present case, this procedure (if all the factors are treated as exploratory) resulted in 18 factors to be tested with a Bonferroni corrected alpha threshold of 0.0028 for which 2 factors achieved significance, with a further five factors potentially being significant if a case can be made to treat them as being of *a priori* interest (as in clearly corresponding to the P3a or the P3b based on the prior literature).

The utility of converting the PCA results to microvolt scaling can again be seen when one proceeds to interpret the significant ANOVA results with bar charts because, unlike standardized scores, the microvolt-scaled factor scores can then be directly interpreted. Consider, for example, that ERP components are electrical dipoles and hence will manifest as positive voltages over one side of the head and negative voltages over the other side of the head, such that they sum to zero over the entire body surface (Nunez, 1990). Whether an ERP component is considered to be a negativity or a positivity by researchers is determined by which side of the voltage fields falls over the traditional scalp recording sites. So how can a positive factor score simultaneously represent both the positive voltage sites and the negative voltage sites? In actuality, the sign of a factor score is arbitrary. It is only the product of the appropriate factor loading and the factor score that has a meaningful relationship to the sign of the original voltage data. In this case, the factor loadings for the negative voltage sites would be negative and the positive voltage sites would be positive. In order to produce a meaningful bar chart, one would need to choose a specific moment in time at a specific electrode and then multiply the corresponding factor scores and factor loadings, all appropriately scaled into microvolts. Note that converting the factor scores to microvolt scaling prior to the ANOVA would have no effect on the final results since the units in the numerator and the denominator of the F-ratio cancel out, leaving the same number regardless of the scaling.

Note also that if a spatial PCA (or a two-step PCA) was conducted, then it makes no difference to the result which channel is chosen; choosing a channel merely allows one to express the dependent measure in a readily interpretable manner. This is the case because the dependent measure (the factor scores) does not include channel variance; the channels are expressed as a set of factor loadings which all relate in a linear manner to the same set of factor scores. Thus, assuming identical standard deviations across the channels for the sake of simplicity, if Channel 1 has a factor loading of .6 and Channel 2 has a factor loading of .3, then the microvolt scaled dependent measure for

Channel 1 will be twice as large as if Channel 2 was selected (because converting to microvolt scaling involves multiplying the factor scores by the factor loading) but otherwise the ANOVA results will be identical (because both the numerator and the denominator of the F statistic will be twice as large and therefore this difference in magnitude will cancel out, leaving the identical F statistic). A similar logic applies to time points when a temporal PCA (or two-step PCA) has been conducted.

OTHER METHODS

It may be of some interest to consider how PCA relates to other methods. One related technique that has become popular is global field power or GFP (Lehmann & Skrandies, 1980). In this method, the voltage measurements at a given timepoint at every channel is first squared (to eliminate negative signs) and then added together. This provides an overall measure of voltage activity across the entire head. Discontinuities in the resulting graph over time can be interpreted as evidence for a change in the ERP componentry and used to segment the epoch into segments. This approach is somewhat equivalent to a temporal PCA in that both segment the time points. Compared to temporal PCA, it is simpler to use. In principle, it has the drawback that it could miss a shift between the dominance of one ERP component and another if they temporally overlap such that there is not a marked dip in the GFP between them. This issue can be addressed using methods like TANOVA (Pourtois et al., 2008) and single-trial clustering methods (De Lucia, Michel, & Murray, 2010), which detect changes in scalp topography. This collection of methods share the drawback of temporal PCA in that they may be blind to ERP components that have largely the same time course (Pourtois et al., 2008), such as the P3a and the P3b in the example dataset. Every method has strengths and weaknesses and so it will require systematic quantitative comparison studies to determine relative utility.

CONCLUSION

In conclusion, applied properly and cautiously, PCA can help interpret the structure of ERP datasets as well as facilitate in their source analyses (for examples where PCA was demonstrated to provide improved co-registration of ERP data with fMRI data, see Dien et al., 2003; O'Hare et al., 2008). They are not infallible, however, and so the results should always be taken cautiously and in light of prior studies. In general, there is also a cost to their use in that statistical power is often reduced compared to simple windowed measures, even without the need for multiple comparisons control. One must therefore weigh the merits of the increased complexity and reduced statistical power versus the improved characterization of the data. When that trade-off is deemed to be worthwhile, the ERP PCA Toolkit can be a helpful tool.

FIGURE LEGENDS

Figure 1. Example Novelty P3 Dataset. The figure displays the grand average waveforms at three representative electrode sites from a prior report of a novelty oddball paradigm (Spencer et al., 1999). "aFz" is just anterior to the channel Fz, on the front of the head. Cz is at the very top of the head. Pz is just behind it, over the parietal lobes. The "z" part of the channels names denote that they are located along the midline of the scalp running from front to back.

Figure 2. Flowchart of steps in PCA of ERP. This figure lists the different steps to be taken where decisions are required.

Figure 3. Effects of Reference on Spatial PCA. This figure shows the effect that the reference channel choice can make on spatial PCA results. On the left is the result of a four-factor Varimax rotation with mean mastoid references on the bottom of the head and on the right is the result with a mean of Cz and Pz channels on the top of the head.

Figure 4. Screenshot of the EP Toolkit's Parallel Test. The plot pits the size of the unrotated factors of the data against the size of the unrotated factors of a same-sized random dataset. The number of factors that are larger than what one obtains from a random dataset is the number that one should retain, in this case nine, as indicated by the arrow. The X-axis is the factors and the Y-axis is the variance accounted for by that factor.

Figure 5. Unrotated Versus Rotated Solution. This figure contrasts the unrotated factor loadings versus the corresponding rotated (Promax) factor loadings. Note how the rotation has caused the waveforms to become more ERP-like, simpler and uniphasic (the general consensus of the ERP community is that ERP components are usually uniphasic, which is to say they have only a single peak or dip as the case may be). The

original unrotated waveforms appear to be, in some cases, combinations of multiple ERP components.

Figure 6. Varimax Rotation Versus Promax Rotation. This figure contrasts the Varimax factor loadings versus the corresponding Promax factor loadings. Note how the Promax rotation has caused the waveforms to become even tighter and ERP-like, especially along the ascending slope of Factors 1 and 3 and the descending slope of Factor 2. Note also how Promax's relaxation of the arbitrary orthogonality constraint has allowed Factors 1 and 2 to become more distinct from each other.

Figure 7. Microvolt Scaling Versus Standardized Scaling. This figure contrasts the effects of presenting the grand average waveform at Cz and of the factor loadings in both microvolt scaling and standardized scaling (wherein the microvolt values are divided by the standard deviation of the time point across the entire dataset). Note how standardizing has caused the noise in the early time points for both the grand average and for the factor waveforms to become magnified. Also note how standardizing has distorted the P300 in the grand average waveform; while less obvious, the factor waveforms have been similarly distorted by the standardizing. Converting the factor waveforms to microvolt scaling removes these distortions and allows them to be directly compared to the grand average waveforms, which are already in microvolts.

Figure 8. Reconstructed P3 Factor. This figure compares the reconstructed P3 factor to the grand average waveform. These P3 factor waveforms represent the portion of the grand average accounted for by the P3 factor. The PCA Factor column illustrates how the scalp topography differs in the Target and the Novel conditions, suggestive that multiple ERP components were aggregated into this P3 factor (in this case, the P3a and the P3b). The grand average column is nearly identical to the PCA column, showing that at this time point (296 ms), this factor accounts for nearly all of the ERP data. The waveform column shows how at channel Cz (at the top of the head), this factor accounts

for only a portion of the ERP grand average waveform as one looks across the entire epoch.

Figure 9. Two-Step PCA Results. This figure displays how the Two-Step PCA procedure has split the initial P3 factor (illustrated in Figure 8) into three components, the P3a, the P3b, and the Frontal Negativity. See how it has clarified the nature of the condition effects, with the P3a and the Frontal Negativity being much larger in the Novel condition whereas the P3b is largely comparable in both the Target and the Novel conditions. In addition, the Positive Slow Wave factor is also displayed to show how its scalp topography is shown to differ from that of the P3b, confirming that it is a different ERP component, a conclusion that would have been difficult to make based on just the grand average data where these different ERP components were mixed together.

FOOTNOTES

1) The Toolkit can be downloaded for free (https://sourceforge.net/projects/erppcatoolkit/). Those interested in can also join the mailing list (https://lists.sourceforge.net/lists/listinfo/erppcatoolkit-support) to be alerted when new versions are posted.

REFERENCES

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129-1159.

Carretie, L., Tapia, M., Mercado, F., Albert, J., Lopez-Martin, S., & de la Serna, J. M. (2004). Voltage-based versus factor score-based source localization analyses of electrophysiological brain activity: A comparison. *Brain Topography*, *17*(2), 109-115.

Curry, S. H., Cooper, R., McCallum, W. C., Pocock, P. V., Papakostopoulos, D., Skidmore, S. et al. (1983). The principal components of auditory target detection. In A. W. K. Gaillard, & W. Ritter (Eds.), *Tutorials in ERP research: Endogenous components* (pp. 79-117). Amsterdam: North-Holland Publishing Company.

De Lucia, M., Michel, C. M., & Murray, M. M. (2010). Comparing ICA-based and single-trial topographic ERP analyses. *Brain Topogr*, *23*(2), 119-127.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9-21.

Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clin Neurophysiol*, *117*(3), 699-702; author reply 703-7.

Dien, J. (2010a). Evaluating Two-Step PCA Of ERP Data With Geomin, Infomax, Oblimin, Promax, And Varimax Rotations. *Psychophysiology*, *47*(1), 170-183.

Dien, J. (2010b). The ERP PCA Toolkit: An Open Source Program For Advanced Statistical Analysis of Event Related Potential Data. *Journal of Neuroscience Methods*, *187*(1), 138-145.

Dien, J. (1998a). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*(1), 43-55.

Dien, J. (1998b). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, and Computers*, *30*(1), 34-43.

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potential analysis: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*(8), 1808-1825.

Dien, J., & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-Related Potentials: A Methods Handbook.* Cambridge, Mass: MIT Press.

Dien, J., Frishkoff, G. A., Cerbone, A., & Tucker, D. M. (2003). Parametric analysis of event-related potentials in semantic comprehension: Evidence for parallel brain mechanisms. *Cognitive Brain Research*, *15*, 137-153.

Dien, J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax versus Infomax rotations. *Human Brain Mapping*, *28*(8), 742-763.

Dien, J., & O'Hare, A. J. (2008). Evidence for Automatic Sentence Priming in the Fusiform Semantic Area: Convergent ERP and fMRI findings. *Brain Research*, *1243*, 134-145.

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, *17*, 637-650.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the "Late Positive Complex": Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, *41*(5), 665-678.

Dien, J., Tucker, D. M., Potts, G., & Hartry, A. (1997). Localization of auditory evoked potentials related to selective intermodal attention. *Journal of Cognitive Neuroscience*, *9*(6), 799-823.

Donchin, E. (1966). A multivariate approach to the analysis of average evoked potentials. *IEEE Transactions on Bio-Medical Engineering*, *BME-13*, 131-139.

Donchin, E., & Heffley, E. (1979). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research (EPA 600/9-77-043)* (pp. 555-572). Washington, DC: U.S. Government Printing Office.

Glaser, E. M., & Ruchkin, D. S. (1976). *Principles of neurobiological signal analysis.* New York: Academic Press.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Harman, H. H. (1976). *Modern Factor Analysis* (3rd ed.). Chicago: University of Chicago Press.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *The British Journal of Statistical Psychology*, *17*, 65-70.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis.* New York: John Wiley & Sons.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187-200.

Kayser, J., & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*(12), 2307-2325.

Kayser, J., & Tenke, C. E. (2006). Consensus on PCA for ERP data, and sensibility of unrestricted solutions. *Clin Neurophysiol*, *117*, 695-707.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol*, *48*(6), 609-621.

Möcks, J. (1986). The influence of latency jitter in principal component analysis of event-related potentials. *Psychophysiology*, *23*(4), 480-484.

Möcks, J., & Verleger, R. (1991). Multivariate methods in biosignal analysis: application of principal component analysis to event-related potentials. In R. Weitkunat (Ed.), *Digital Biosignal Processing* (pp. 399-458). Amsterdam: Elsevier.

Nunez, P. L. (1990). Physical principles and neurophysiological mechanisms underlying event-related potentials. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 19-36). New York: Oxford University Press.

O'Hare, A. J., Dien, J., Waterson, L., & Savage, C. R. (2008). Activation of the Posterior Cingulate by Semantic Priming: A co-registered ERP/fMRI Study. *Brain Research*, *1189*(2), 97-114.

Pourtois, G., Delplanque, S., Michel, C., & Vuilleumier, P. (2008). Beyond Conventional Event-related Brain Potential (ERP): Exploring the Time-course of Visual Emotion Processing Using Topographic and Principal Component Analyses. *Brain Topogr*, *20*(4), 265-277.

Ruchkin, D. S., Munson, R., & Sutton, S. (1982). P300 and slow wave in a message consisting of two events. *Psychophysiology*, *19*(6), 629-642.

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

Ruchkin, D. S., Villegas, J., & John, E. R. (1964). An analysis of average evoked potentials making use of least mean square techniques. *Annals of the New York Academy of Sciences*, *115*(2), 799-826.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, *36*, 409-414.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal Analysis of the Late ERP Responses to Deviant Stimuli. *Psychophysiology*, *38*(2), 343-358.

Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. (1977). On the influence of task relevance and stimulus probability on event-related-potential components. *Electroencephalography and Clinical Neurophysiology*, *42*, 1-14.

Sutton, S., & Ruchkin, D. S. (1984). The late positive complex: Advances and new problems. *Annals of the New York Academy of Sciences*, *425*, 1-23.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380-384.