Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. Clinical Neurophysiology, 117(3):699-702. (https://doi.org/10.1016/j.clinph.2005.09.029)

This letter to the editor is addressed to the recent editorial (Kayser & Tenke, 2005) regarding my own paper (Dien, Beal, & Berg, 2005). I am delighted by the attention our contribution received and quite appreciate the comments made by the authors regarding the importance of the topic of PCA of ERPs, as well as the contribution of our paper regarding the utility of Promax rotations. As one in a series, our present paper continues to advance this topic using both real and simulated datasets (Curran & Dien, 2003; Dien, 1998; Dien, 1999; Dien, & Frishkoff, 2004; Dien, Frishkoff, Cerbonne, & Tucker, 2003; Dien, Frishkoff, & Tucker, 2000; Dien, Spencer, & Donchin, 2003; Dien, Spencer, & Donchin, 2004; Dien, Tucker, Potts, & Hartry, 1997; Spencer, Dien, & Donchin, 1999; Spencer, Dien, & Donchin, 2001). From the editorial, it sounds like we are making progress towards a consensus on the application of PCA to ERPs. However, Kayser and Tenke's editorial also expressed a number of continuing disagreements with the contents of our paper. In order to further develop this process of forging a consensus, this letter is addressed to these disagreements.

First of all, I quite regret their apparent discomfort our work evidently caused (e.g., " it seems almost reasonable to become upset", p. 1750). It was not our original intent to highlight our differences of opinion with Kayser and Tenke, much less to upset them; in fact, we were urged to address these issues by one of our anonymous reviewers who seemed quite insistent that we fully adopt the recommendations of the Kayser and Tenke (2003) report.

Overall, there are two separate areas of disagreement, concerning the scaling issue (correlation versus covariance) and the restriction (or truncation) of factor solutions. The latter involves analytic philosophy and is where Kayser and Tenke have contributed particularly useful insights. The readership may find our contrasting positions and reasoning to be illuminating. The former area involves mathematics and I fear the readership may find the Kayser and Tenke (2003) paper and their recent editorial to be misleading. I will first address the mathematical issues and then proceed to the analytical issue. I will then end on some remaining comments regarding the Promax rotation. To keep the discussion simpler, I will only refer to the case of temporal PCA wherein the variables are the time points.

Three Scaling Choices

As described in my paper, it is critical that researchers using this technique understand that the choice of scaling (correlation or covariance) is made three times during the PCA: extraction, rotation, and presentation. These are three separate decisions and so the same choice does not have to be made for each step. For example, a common procedure is to use the covariance matrix during the initial extraction but then to use correlations during the actual rotation and when presenting the loadings. This in fact is the default procedure in SAS and SPSS when a covariance matrix is being subjected to PCA (see below).

My main concern with Kayser and Tenke's discussion in both their 2003 paper and their editorial is that they are not clearly distinguishing between these three decisions. For example, their statement in the editorial that "Standard statistical packages (BMDP, SPSS, SAS) typically rotate the factor loadings as extracted and weighted" suggests they might think that the extraction choice and the rotation choice are typically treated as a single choice. The code in the appendix of their 2003 paper also seems to operate in

this fashion. For SAS and SPSS this statement is definitely incorrect. Not only have I verified my own Varimax rotation code (which does not work this way) against SAS output, I have directly obtained statements to this effect from the companies in question: "whether the COV option is on or not, all factor loading matrices are always applied to the standardized variables" (SAS Technical Support staff, personal communication, 2005). Likewise, "Rotation is based on the standardized (rescaled if the covariance matrix was used for extraction) loadings, regardless of whether the covariance or correlation matrix was analyzed" (David Matheson, SPSS statistical support, personal communication, 2005). Regarding BMDP, the BMDP technical support was not able to answer this question. Furthermore, their manual is rather unclear on this point and may in fact have contributed to our disagreements on this matter. Since I do not have access to BMDP software, I have not been able to test it directly.

In any case, the rotation choice is for the most part moot since, according to the mathematical proof I provided in my paper, Kaiser normalization cancels out the difference between correlation and covariance loadings. Kaiser normalization is the default in SAS, SPSS, and probably in most other statistics packages. "Kaiser normalization is the default" (SAS Technical Support staff, personal communication, 2005). "If Kaiser normalization was not suppressed by the NOKAISER keyword, it is applied to the standardized loadings" (David Matheson, SPSS statistical support, personal communication, 2005). It is again unclear what BMDP does in this regard.

As for the presentation choice, this is a matter of communication, not analysis. Saying that the choice of presentation scaling changes the analysis is like saying that deciding whether to graph the means or the standard deviations of the cells changes a dataset. It is just a choice of how to present the results once the analysis is finished. In other words, when Kayser and Tenke (2003) state in footnote 2 that "To clearly distinguish this extraction method from the regular use of the covariance matrix, we will refer to these two procedures as standardized and unstandardized covariance-based PCA solutions" and present different-looking waveforms for these two "options" in Figure 4, what they are really doing is presenting the same analysis twice, using two different scalings.

What are these two scalings? The results from the "standardized" covariance matrix is simply the factor loadings (correlations between the variables and the factor scores). The "unstandardized" covariance matrix is simply the identical factor loadings but with each one multiplied by the standard deviation of the corresponding variable (since each variable has a different standard deviation, the resulting waveform will have a somewhat different shape).

In all, I am in agreement with Kayser and Tenke on two points concerning the mathematical issues. The first is that they demonstrated that covariance relationship matrices are a better choice than correlation relationship matrices. While this has been previously reported with real data (Curry et al., 1983), it was well worth replicating. Thus, I compliment the authors on examining this issue, given contradictory recommendations in reviews on the subject (Chapman & McCrary, 1995; Donchin, & Heffley, 1979; Möcks, & Verleger, 1991; van Boxtel, 1998). Combined with my own simulation analysis (Dien, Beal, & Berg, 2005), I think we have established strong grounds for considering this point settled. The second is that they demonstrated that choosing covariance scaling at the presentation step yields waveforms that resemble the original ERPs more closely. This, too, is a point that appears to have attained a fair amount of consensus (cf. Dien, Tucker, Potts, & Hartry, 1997; Möcks & Verleger, 1991).

I therefore completely agree that "microvolt-scaled factor loadings are preferable when interpreting PCA solutions".

Disagreements Regarding Scaling Choices

We part ways when they express concern that using covariance (micro-volt) scaling at the presentation step would "necessarily requires an additional recalculation of the overall and explained variance contributions, and an additional reranking of the factors." Fortunately, covariance scaling can be used during the presentation step without these complications. The choice of presentation scaling is simply a matter of data presentation; because the actual data do not change, no change in rankings is necessary. Furthermore, the conversion from correlation to covariance scaling is fortunately readily performed in an Excel spreadsheet. For example, put the correlation factor loadings in Column1. Put the standard deviations of the variables in Column2. In Column3, put the formula "= Column1 * Column2". Given the ease with which covariance scaling can be obtained for the presentation step, I think there is little reason not to use it for waveform plots of the factor loadings.

Kayser and Tenke (2005) argue that "While rescaling would indeed increase the similarity of corresponding waveforms, and in fact result in identical waveforms for the unrestricted solutions, the main point is that the sequence of the extracted factors differs due to differences in explained variance." This statement is incorrect in two ways. The first is that, as noted earlier, the only difference between what they call "standardized and unstandardized" covariance analyses is the presentation scaling so they are exactly the same, whether or not the solution is restricted or unrestricted. The difference is only a choice of how to present the data. Secondly, the statement is incorrect in that the sequence of the factors does not have to differ. The ordering of factors is an arbitrary decision and thus up to the author. An author can convert correlation-scaled factor loadings to covariance-scaling at the presentation step (or vice-versa), as described in the preceding paragraph, but retain the original ordering for the sake of comparability. Ultimately, if the ordering of the factors is important to a paper's analysis logic, it would be helpful for the authors to describe how the ordering is being determined and its role in the analysis process.

A further misstatement is made in the next sentence "As shown in our Fig. 4 for real ERP data, and in our Fig. 3 for an illustrative example constructed to explain the underlying principle, correlation loadings can result in erroneous high-variance factors." Again, choice of the presentation scaling only affects the appearance of the factor waveforms and does not affect their nature. Thus, while I agree that using correlation scaling at the extraction step results in degraded solutions in comparison to covariance scaling, the "erroneous" factors that they refer to for the "covariance matrix using standardized loadings" in Figures 3 & 4 differ from the "covariance matrix using unstandardized loadings" only in the appearance of the waveforms due to the choice of the presentation scaling.

How might scaling of the presented factor loadings produce such apparent "erroneous" factors? Not having access to their data, I cannot say with certainty. On the basis of my own experience I expect it is due to the scaling choice increasing the salience of the noise in relatively inactive time points. When using correlation scaling to present factor loadings, the inactive time points have been given equal weighting as the active time points. It is as if one had measured a group of mice (baseline noise), a group of sky

scrapers (P300), taken their mean heights, divided the heights by their respective standard deviations, and then graphed the resulting values. One might very well obtain a chart in which the mice appear to be the size of sky scrapers (or in this case, clearly erroneous peaks that do not correspond to ERP features in the grand average). Obviously, researchers are unlikely to desire such an analysis. Remember, PCA is often applied to questionnaire data where the scaling is meaningless. When the variables do share a meaningful scaling (feet for mice and skyscrapers, microvolts for ERPs), it does not make sense to use correlation scaling when presenting the factor loadings if the goal is to directly relate them to the original features (as in comparing the shape of the waveforms); doing so can cause a perfectly good PCA solution look like it has "erroneous" features.

This conclusion is not affected by the following statement made by Kayser and Tenke in their editorial: "However, if a retention criterion is applied before rotation, a non-identical set of factors may be submitted to the rotation procedure, resulting in the computation of different communalities, which leads to a different Kaiser's normalization, and finally different rotated loadings…" Once again, while we agree that the extraction step scaling will make a difference (and have acknowledged the value of this point), this is the only choice that is made prior to the rotation procedure. The Kaiser normalization (if used) makes the rotation choice moot and the presentation choice does not affect the analysis itself, only its presentation.

Analytical Philosphy

Moving on to the analytical issue, we encounter an issue where there is no mathematically correct answer and one is instead concerned with contrasting priorities and goals regarding analysis. The essential issue is that Kayser and Tenke (2003) advocated utilizing unrestricted solutions wherein all the factors are retained whereas Dien, Beal, and Berg (2005) advocated using a truncated solution on the grounds that having to examine so many factors would present multiple comparison problems.

In their editorial, Kayser and Tenke (2005) clarify that they are advocating a threshold approach in which only factors accounting for more than some level of variance are evaluated for meaningfulness. Their original report did not specify such an approach nor any criteria for setting such a threshold. The expanded procedure as described in the editorial therefore represents an important clarification and/or modification of the original report.

My own position in this regard is that I feel uncomfortable with equating size of an effect with meaningfulness. However, this is a disagreement that cannot be resolved with mathematics and I do not believe that either of us have put this issue to any kind of conclusive test. While it is indeed striking that the F-tests in their 2003 study yielded more significance with the unrestricted solution, without knowing the true state of affairs it is unclear whether this represents an advance away from Type II error or a retreat into Type I error. It would be helpful to replicate this analysis with a reliable and well-characterized effect such as an oddball task. Kayser and Tenke describe having done so but did not present the actual results so it is not possible to evaluate them. I do think that Kayser and Tenke's proposal is not unreasonable and that further tests are quite merited. Indeed, only by examining many datasets, both real and simulated, will it be possible to make any claims of generalizability since this is not an issue that is amenable to mathematical proofs.

Promax Rotation

Some final comments regarding the Promax rotation may be helpful. Kayser and Tenke (2005)'s editorial made some positive comments about the simulation results with the Promax rotation, which we appreciate. They do evince some remaining skepticism, which is reasonable. I would therefore like to end with some final comments to further reinforce the recommendation to utilize Promax rotations. While they suggest that the orthogonal rotations yielded by Varimax might be considered to be more "parsimonious" and hence more useful, it is well to keep in mind that the principle of parsimony implicitly includes the condition of accuracy. For example, the most parsimonious description possible of an ERP dataset would be to claim that it contains only a single ERP component (with many peaks). While this is parsimonious in the sense of being very simple, it is clearly incorrect and not useful for typical ERP datasets. In the same fashion, I argue that orthogonal factors, while simple, are not accurate and hence not useful; in this sense, they are not in fact parsimonious. As has been said by others, the brain is not orthogonal.

Secondly, they note that the Promax results were obtained with simulation data and hence tests with real data would be desirable to have more confidence in this finding. Such a test is already available. As described elsewhere (Dien, Spencer, & Donchin, 2003), although the source of the P300 has not been conclusively isolated, the most likely source according to brain imaging studies, intracranial recordings, and lesion studies is the temporoparietal junction. A PCA solution of the P300 did in fact converge on this region as the source for the P300, but only when Promax was used as the rotation; a Varimax solution yielded a much less plausible source solution. This study therefore provides additional weight for the recommendation of using Promax solutions, based on real data.

Conclusion

In conclusion, therefore, I would like to echo the comments by Kayser and Tenke regarding the renaissance of the PCA method in the ERP literature. In order to help implement these recommendations, I have made PCA software tools written in Matlab available (http://www.people.ku.edu/~jdien/downloads.html). I hope that readers will find the studies by both Kayser and Tenke and by my co-authors, such as Manny Donchin, Kevin Spencer, and Don Tucker, to have been useful contributions to the field and expect that both our agreements and our disagreements with Kayser and Tenke will have been illuminating to readers. I appreciate their joining me on this communal effort to improve the utility of ERP measures and look forward to sharing future findings.

References

Chapman, R. M., & Mccrary, J. W. (1995). EP component identification and measurement by principal components analysis. *Brain and Cognition*, 27, 288-310.

Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, 40, 979-88.

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.

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

Dien, J. (1999). Differential lateralization of trait anxiety and trait fearfulness: evoked potential correlates. *Personality and Individual Differences*, 26(1), 333-56.

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), 1808-25.

Dien, J., & Frishkoff, G. A. (2004). 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., Cerbonne, A., & Tucker, D. M. (2003). Parametric analysis of event-related potentials in semantic comprehension: Evidence for parallel brain mechanisms. *Cognitive Brain Research*, 15, 137-53.

Dien, J., Frishkoff, G. A., & Tucker, D. M. (2000). Differentiating the N3 and N4 electrophysiological semantic incongruity effects. *Brain & Cognition*, 43, 148-52.

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, 637-50.

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), 665-78.

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., & 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. 555-72). Washington, DC: U.S. Government Printing Office.

Kayser, J., & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, 114(12), 2307-25.

Kayser, J., & Tenke, C. E. (2005). Trusting in or breaking with convention: Towards a renaissance of principal components analysis in electrophysiology. *Clinical Neurophysiology*, 116(8), 1747-53.

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.

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, 409-14.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal Analysis of the Late ERP Responses to Deviant Stimuli. *Psychophysiology*, 38(2), 343-58.
Van Boxtel, G. J. M. (1998). Computational and statistical methods for analyzing event-related potential data. *Behavior Research Methods, Instruments, & Computers*, 30(1), 87-102.