Theoretical and empirical rationale for using unrestricted PCA solutions to identify and measure ERP components

Jürgen Kayser and Craig E. Tenke
Department of Biopsychology, New York State Psychiatric Institute, New York

http://psychophysiology.cpmc.columbia.edu

Introduction

• Although principal components analysis (PCA) is widely used to determine “data-driven” ERP components, it is unclear if and how specific methodological choices may affect factor extraction. The effects of three variations, i.e., 1) Type of association matrix (correlation / covariance), 2) Form of Varimax rotation (scaled / unscaled), and 3) Number of components extracted and rotated, were considered and systematically investigated when applying temporal PCA (tPCA) to ERP data.

Theoretical Rationale

• The usefulness of the extracted factors can be evaluated by specific knowledge about the variance distribution of ERPs, which are characterized by large near the end of the recording epoch and at ERP component peaks.
• Whereas a covariance matrix preserves this information, it is lost by a correlation matrix that assigns equal weights to each sample point, yielding the possibility that small but systematic variations may form a factor.
• These considerations were evaluated and confirmed with simulated ERP data (see Figures 1–3).

Methods

• Real ERP data, collected from healthy, right-handed adults during a visual half-field study (see Figure 4), were repeatedly submitted to tPCA (BMDP-4M, Dixon 1992). Columns of the data matrix represented time (110 sample points from -100 to 1,000 ms), and rows consisted of subjects (16), conditions (4), and electrode sites (30).
• IPCAs were performed for three extraction / rotation criteria:
  1) Covariance matrix / Varimax rotation on raw data
  2) Correlation matrix / Varimax rotation
  3) Covariance matrix / Varimax rotation on standardized variables

• 110 IPCAs were computed for each extraction / rotation condition, by systematically increasing the number of components to be extracted from 1 to 110 (= number of variables).

Extraction Method:

A) FORM = COVA.
B) FORM = CORR.

Figure 5. Sequences of factor loadings of the correlation-based solutions (A) and overlaid loadings of Factor 3 for restricted (. . .) and liberal (•) extraction criteria (B).

Figure 6. Grand average ERPs for 16 healthy adults for positive and negative visual stimuli at 30 recording sites, averaged across hemifield of presentation (250 ms exposure in a visual half-field paradigm). Data from Kayser et al. 2000 Int J Psychophysiol 36(3-4):

Results

• Limiting the number of components changed the morphology of some components considerably (see Figures 5B and 6B).
• However, more liberal or unlimited extraction criteria did not degrade or change high-variance components. Instead, their interpretability was improved by more distinctive time courses with narrow and unambiguous peaks (i.e., low secondary loadings; see Figures 5A and 6A).
• Some physiologically meaningful ERP components that are small in amplitude and topographically localized (e.g., P1) were found to have a PCA counterpart (e.g., Factor 130; see Figure 8A), that were lost with restricted solutions due to their low overall variance contributions.
• Covariance-based factors had more distinct time courses (i.e., lower secondary loadings) than the corresponding correlation-based factors (Figures 5B and 6B), thereby allowing a better interpretation of their electrophysiological relevance.
• Correlation-based solutions were likely to produce artificial factors that merely reflected small but systematic variations when the ERP waveform intersected the baseline (i.e., zero; cf. Factors –•– in Figures 5A and 6A).
• Scaling covariance-based PCA factors before rotation approximation correlated-based solutions, and ultimately yielded the same coefficients (factor loadings) when all components were rotated (see Figure 6A).
• The same systematic approach using auditory ‘oddball’ ERP data yielded comparable results for a different set of task-specific PCA factors.

Figure 7. Overlap factor loadings and factor score topographies of the first 10 covariance- (A, C) or correlation-based (B, D) PCA components extracted from the unrestricted (•) solution, identified by peak latencies of factor loadings.

Figure 3. Time course of factor loadings for the first PCA factors extracted from the correlation matrix for pseudo ERP data with (B) and without noise (A). The covariance-based PCA extracted a component (factor 1), that accurately reflected the introduced variance shape for both data sets. The correlation-based PCA only produced a component (factor 1) that indicated the direction, but not the size of variations from zero (i.e., from baseline). Similarly, the constant low-level offset was disproportionally reflected in another component (factor 3) for the noise-free data.

Figure 2. A) ERP ‘group’ average of noise data set. B) ERP ‘group’ average of noise data set.

Figure 4. Grand average ERPs for 16 healthy adults for positive and negative visual stimuli at 30 recording sites, averaged across hemifield of presentation (250 ms exposure in a visual half-field paradigm).