Unrestricted Principal Components Analysis of Brain Electrical Activity: Issues of Data Dimensionality, Artifact, and Utility

Frank H. Duffy*, Kenneth Jones, Peter Bartels ***, Gloria McAnulty*, and Marilyn Albert******

Summary: Principal components analysis (PCA) was performed on the 1536 spectral and 2944 evoked potential (EP) variables generated by neurophysiologic paradigms including flash VER, click AER, and eyes open and closed spectral EEG from 202 healthy subjects aged 30 to 80. In each case data dimensionality of 1500 to 3000 was substantially reduced using PCA by magnitudes of 20 to over 200. Just 20 PCA factors accounted for 70% to 85% of the variance. Visual inspection of the topographic distribution of factor loading scores revealed complex loadings across multiple data dimensions (time-space and frequency-space). Forty-two non-artifactual factors were successful in classifying age, gender, and a separate group of 60 demented patients by linear discriminant analysis. Discrimination of age and gender primarily involved EP derived factors, whereas dementia primarily involved EEG derived factors. Thirty-eight artifactual factors were identified which, alone, could not discriminate age but were relatively successful in discriminating gender and dementia. The need to parsimoniously develop real neurophysiologic measures and to objectively exclude artifact are discussed. Unrestricted PCA is suggested as a step in this direction.

Key words: Spectral analysis; Evoked potentials; Principal components analysis; Singular value decomposition; Discriminant function analysis; Artifact; Dimensionality; Aging; Dementia.

Introduction

Principal components analysis (PCA) is a widely used statistical procedure (Bartels 1981a; Bartels 1981b; Cooley and Lohnes 1971; Hotelling 1933; Seal 1964) that has been applied to brain electrical activity including data derived from EEG spectral analysis and from EEG signal averaging (evoked potentials or EP) (Donchin 1966; Duffy et al. 1990; Harner and Riggio 1989; Harner et al. 1991; John et al. 1973; Kavanagh et al. 1976; Lopes da Silva 1987; Maier et al. 1987; Molfese et al. 1985; Rawlings et al. 1968; Rogers and Douglas 1984; Rösler and Manzey 1981; Scherg and Von Cramon 1985; Skrandies and Lehmann 1982; Suter 1970; Valdes et al. 1990; Van Rotterdam 1970; Wood and McCarthy 1984). In general, according to Bartels (Bartels 1981a, 1981b), PCA offers many advantages to the analysis of multivariate data sets for circumstances where initial, observed variables are manifold but may be redundant or highly intercorrelated as is the case for EEG. By PCA, variables are transformed into a new data set of parsimonious or mutually independent and uncorrelated (orthogonal) measures usually referred to as principal components or factors. Resulting factors are commonly expressed as the linear, weighted combination of observed variables and inspection of such factor loading coefficients assists in establishing the identity or meaning of the newly created factors. To improve interpretability, factor contrast can be enhanced while maintaining orthogonality through rotations in data space, often by the Varimax procedure (Kaiser 1958, 1959). Finally, selection of those factors carrying the bulk of information about the data set, e.g., by magnitude of the Eigenvalue (Kaiser 1960) or by the Bartlett criteria (Bartlett 1950) yields an estimate of the intrinsic data set dimensionality. In addition, subsequent statistical procedures, such as discriminant function analysis, are facilitated by the meaningful reduction of input variables afforded by PCA.

*Department of Neurology, Childrens Hospital and Harvard Medical School, Boston, MA, USA.

**Florence Heller Graduate School for Advanced Studies in Social Welfare, Brandeis University, Waltham, MA, USA.

***Optical Sciences Center, University of Arizona, Tuscon, AZ, USA.

****Departments of Neurology and Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

Accepted for publication: March 2, 1992.

Acknowledgements: This work was supported in part by NIA program project PO1AG049853 to M. Albert and the Mental Retardation Program Project P30HD18655 to J.J. Volpe. We thank our qEEG technologists Adele Mirabella, Susan Katz, Ellen Belles, and Marianne McGaffigan as well as our research secretaries for their unflagging support.

Correspondence and reprint requests should be addressed to Frank H. Duffy, MD, Department of Neurology, Childrens Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA.

As we have previously discussed (Duffy et al. 1990) commercially available quantified EEG (qEEG) devices now facilitate collection of very large data sets involving many thousands of variables. Variable number can be obtained by multiplying the number of scalp electrodes (typically 19 to 21) by the number of data points per electrode (typically 64 to 256). Prior to analysis users may reduce variable number on the basis of pre-existing convention such as restricting analysis to a selected subset of electrodes and to predetermined EEG spectral band(s) or EP latency range(s). Recently Abt has described Descriptive Data Analysis, a more rigorous process by which users may restrict analyses to data subsets by recognition of patterns within their data set (Abt 1987, 1988; Duffy et al. 1990). The traditional PCA technique for data reduction is not easily used in neurophysiology for most commonly available statistical packages limit the number of input variables to levels well below 1000 often due to actual or presumed limitations of available computer memory. For this reason, among others, when PCA has been applied to scalp recorded data it has been universally necessary to make prior simplifying assumptions to reduce input variable number before analysis. A common approach to PCA on EP data is to treat each electrode as if it were a separate case. By this simplification variable number is reduced to the number of sampled points in time at a single electrode. Resulting factor loading plots reflect variation over time but not space as spatial variation (electrode location) is sacrificed to achieve variable reduction. This process, referred to as temporal factor analysis, collapses the spatial variation into among subject variation. Alternately, the voltage profile across the scalp at each point in time may be treated as if it were a single case. Resultant topographic maps of factor loadings reveal variation over space, but not time, as temporal variation is sacrificed to achieve prior variable reduction. This process, often referred to as spatial PCA, collapses temporal variation into among subject variation. Although both procedures demonstrate a reduction of data dimensionality, simplifying assumptions place constraints such that potential relationships between input variables involving both space and time may not be detected.

Taking advantage of recent advances in low cost computational power, we undertook PCA of scalp recorded neurophysiologic data without such simplifying assumptions. The number of variables was equal to the actual number of sampled data points which ranged from 1536 for EEG spectral to 2944 for long latency EP data. The number of cases represented the number of actual subjects. Resulting factor loadings were visually inspected across both space and time using techniques commonly employed for topographic mapping of EP data (Duffy 1982; Duffy et al. 1979). In 202 medically healthy subjects we evaluated the EP to flash and click as well as EEG spectra during eyes open and eyes closed states. Our goal was fourfold:

(1) To estimate intrinsic data dimensionality.

(2) To evaluate and identify both biologically meaningful and artifactual factors.

(3) To estimate usefulness of resulting factors by their ability to classify subjects according to age, gender, and the presence or absence of dementia.

(4) To separately estimate the spurious classification power of factors derived from non-cerebral sources (artifact).

Methods

Subjects

Two hundred and two healthy, normal subjects, ranging in age from 30 to 80, were studied. All subjects had been screened to exclude systemic, neurologic and psychiatric illness. Any person with a history of alcoholism, drug abuse, learning disabilities, severe head trauma, epilepsy, hypertension, chronic lung disease, kidney disease, diabetes, coronary artery disease, cancer or psychiatric illness were excluded from the study. Most subjects received a series of standard laboratory procedures (i.e., CBC, SMA-20, EKG) and cognitive tests to further assure that participants were free of clinical disease. The vast majority of subjects were also screened by head CT scan. A few subjects refused this test for fear of radiation exposure but were none-the-less included if all other indices of health were obtained and within normal limits. All subjects were medication free at the time of evaluation. Many of the male subjects were included in a previous electrophysiological study of normative aging (Duffy et al. 1984b). The participants included both blue and white collar workers and, in general, represented a socioeconomic cross-section of the population. The subjects were divided into five decade groups as follows: 30-39 (N=29), 40-49 (N=41), 50-59 (N=29), 60-69 (N=57), 70-79 (N=46). Overall there were 89 males and 113 females of which 191 were right handed and 10 left handed by report; one subject was ambidextrous. The small number of left handers precluded analysis of this variable.

In addition to the healthy population, an additional 60 patients with the clinical diagnosis of early Alzheimer's dementia were studied. Many subjects were included in a previous study of the impact of Alzheimer's disease upon brain electrical activity (Duffy et al. 1984a). The diagnosis of Alzheimer's disease was made by a neurologist and independently supported by a psychiatrist and neuropsychologist. Medical conditions

known to produce dementia were excluded. All subjects were otherwise medically healthy and underwent the evaluation indicated above for the normal controls. Furthermore, to rule out multi-infarct dementia all demented patients had to have ischemic scores of 4 or less on the Hachinski scale (Hachinski 1978). Finally to limit the study to mildly demented patients, only those with scores of 30 or below on the Activities of Daily Living scale were included (Weintraub et al. 1982). Selection criteria for the current study were not based in any way upon prior neurophysiologic data derived from either the control or demented subjects.

Neurophysiologic Data Acquisition

All neurophysiologic data were obtained and analyzed in the BEAM Laboratory of Children's Hospital, Boston. EEG data were gathered during four behavioral states which included the following: resting with eyes open (EOP), resting with eyes closed (ECL), long latency flash visual evoked potential (VER), and long latency tone-pip auditory evoked potential (AER). Considerable care was taken to minimize artifact due to eye movement, eyeblink, muscle tension, mouth and tongue movement, and gross body or head movement. In the eyes open states, subjects were given fixation targets and were instructed to suppress blinking but were allowed frequent time outs or "blink holidays". In the eyes closed states thin, fully transparent soft film was laid gently over the eyes to provide the subject with feedback of residual involuntary eye-blinking. The EEG was continuously monitored to detect and avoid state change (e.g., drowsiness).

Data were obtained from 20 scalp electrodes (standard 10-20 placement plus OZ) and, for EOP and ECL, four other bipolar artifact electrodes strategically placed to monitor vertical and horizontal eye movements and muscle tension resulting from face, jaw, or scalp musculature. Bipolar channels were placed above and below the right eye (VEM), between both outer canthii (HEM), between both temporal muscles just below the zygomatic arch (BTM), and between bodies of both posterior neck muscles 1-2 cm below inion level (BOM). For VER and AER data only 23 data channels were employed; the occipital muscle channel (BOM) was sacrificed for a trial marker since our polygraph was limited to 24 channels. Following amplification by a Grass model 24-D EEG polygraph set to bandpass from 1-300 Hz, data were stored for subsequent analyses on a Honneywell 5600E 28-channel FM analogue 1" tape recorder (0-625Hz bandpass) along with appropriate trial and event markers. A through system sine wave calibration signal of 100 uV peak to peak at 10 Hz was recorded for all channels. Data were analyzed off-line after low pass filtering below 90 Hz and digitization at 256 Hz per channel for spectral analysis and 250 Hz for signal averaging. In this manner, contamination of the EEG spectral frequencies due to undersampling of higher frequency noise (aliasing) was avoided. Although our 256 Hz per channel sampling rate allowed for spectral definition of up to 128 Hz, data above 32 Hz were discarded from analysis of EEG spectral content as artifactual. Thus each spectrum consisted of 64, 0.5 Hz data values from 0.5 to 32 Hz. The 0.0 Hz data value was eliminated as artifactual. EEG data destined for spectral analysis were gathered in two second segments which were inspected off-line, and those containing artifact were eliminated from subsequent analyses. A minimum of one minute (30-two second segments) but often over two minutes of artifact free EEG were used to form the final mean spectrum for each subject. Spectral values were expressed as the log transform (Duffy 1988; Gasser et al. 1982; Zar 1984) of the square root value. EP segments containing eye blink or motion artifact were eliminated on the basis of individually adjustable over-voltage criteria. A minimum of 200 segments, to a maximum of 500 segments, were used to form the final averaged EP. All EPs were formed over 512 msec latency epochs (128, 4 msec data points) both before and after time of stimulation. The prestimulus latency epoch was used to assure adequacy of signal averaging, to verify absence of time locked prestimulus artifact, and to provide a zero microvolt reference point.

The VER was formed by visual stimuli (consisting of high intensity stroboscopic flashes) delivered from a sound-dampened Grass photostimulator Model PS-2, set at intensity 8, placed 25 cm from the subject's closed eyes. At such supra-maximal intensities dilatation is not necessary and was not used (Skalka and Holman 1986). A white-noise generator masked residual clicks. Stimuli were delivered on a pseudorandom basis. The mean interstimulus interval was 2.3 sec (range, 1.79 to 2.82 sec). To form the AER, auditory stimuli were delivered through binaural earphones at a supramaximal 92 db sound pressure level. The clicks consisted of 50 msec tone pips at 960 Hz with 10 msec rise and fall times. The delivery schedule was the same for the visual stimulation. EEG was carefully monitored for artifact and drowsiness during EP presentation.

Off-line data processing was performed on a Masscomp 5500 digital computer. A Nicolet software package was employed for off line digitization of tape recorded signals, artifact removal, spectral analysis, signal averaging, and topographic mapping (Duffy et al. 1979). The end product for each spectral analysis was 24 spectral waveforms and for each EP, 23 EP waveforms - one from each active electrode. Thus each EEG state (EOP and ECL) generated 1536 spectral variables per

subject: 64 values per channel times 24 channels (20 scalp, four artifact). Similarly each EP state (VER and AER) generated 2944 variables per subject: 128 values per channel times 23 channels (20 scalp and three artifact - the occipital mucle artifact channel was sacrificed for an event marker). Details of this procedure have been previously described (Duffy 1989).

Principal Components Analysis

In the PCA technique there are three fundamental options regarding which data to factor: 1. the raw data, 2. the centered data (i.e., means removed from each observation), and 3. the standardized data (i.e., centered and shifted to have unit variance). These options correspond to factoring 1. the raw cross products matrix, 2. the dispersion or variance-covariance matrix, and 3. the correlation matrix. We chose to factor the correlation matrix as we ultimately wish to look at individual differences in brain electrical activity as they relate to other clinically derived variables such as those resulting from neuropsychological testing. Deviations from the average response are of primary interest and that is what our factors represent. PCA did not include the 60 pathological subjects. To develop factor scores on this group it was necessary to standardize these 60 patients on the basis of the means and variances of the normal group.

Our approach was to factor the entire 1536 or 2944 variable by 202 subject matrix. One may envision this type of data reduction as determiniung the minimum dimensionality of a space spanned by 202 subject vectors each containing 1536 or 2944 variables. In the usual case for PCA the number of subjects is assumed to be greater than the number of variables. For our data, the opposite was true. Accordingly it is necessary to take advantage of several techniques adapted to this issue (Golub 1989; Horst 1965; Jones 1967). We chose the technique known as Singular Value Decomposition (Golub 1989) to facilitate solutions in a reasonable length of computer time with manageable memory requirements. As the number of cases was smaller than the number of variables, the maximum number of possible orthogonal roots was determined by the smaller case dimension of 202 (or 201 since the mean was removed).

Although the default minimal reduction of dimensionality from 1536 or 2944 to 201 might prove interesting since the roots would be orthogonal, the optimal and hoped for result would be to find a small number of roots describing a reasonably large amount of variance. A number of procedures have been developed to estimate the number of "significant factors" which traditionally include: 1. cumulative percent of trace, 2. Eigenvalues greater than one or the "rule of 1" (Kaiser 1960), and 3. Bartlett's test of significance of the residual matrix (Bartlett 1950). It is not uncommon for these estimates to disagree. In our experience Bartlett's test is more conservative, showing fewer "significant" factors than the rule of 1.

To maximimize factor interpretability the standard Varimax rotation procedure was employed (Kaiser 1958). Evaluation of PCA results (Table 1) revealed disagreement among the different procedures for estimating the precise number of "significant" factors. At least 20 factors were significant for all cases and accounted for 70 - 85% of the cumulative variance in each case. As a computational compromise we elected to limit Varimax rotation and subsequent discriminant analyses to the first 20 factors for each analysis. A consequence of this compromise is that we cannot exclude the possibility of useful information in factors above 20.

PCA and Varimax rotation were performed on a 25 MHz 80486 microprocessor with Weitek coprocessor and 32 megabytes of memory running PC Unix 5.4 from Esix. PCA took approximately one hour of CPU time including Varimax rotation.

Other Statistical Analyses

The prediction of age by the outcome factors was assessed by both multiple regression analysis with age as the independent variable and by linear discriminant function analysis with age as the "by decade" grouping variable. Assessment of the value of the factors in predicting gender and dementia involved linear discriminant function analysis. The 60 demented subjects were used only for the dementia analysis and were not included in the development of factors or for the gender and age analyses. The SPSS statistical package was employed.

Results

Dimensionality

Table 1 shows the first ten Eigenvalues, the number of Eigenvalues above 1.0, and the number of statistically significant facors by Bartlett's test and the variance explained by 20 factors. Based upon Bartlett's test, reduction in dimensionality by PCA was considerable: 71.8 for VER, 226.5 for AER, 59.1 for EOP and 66.8 for ECL. The percent variance explained by the first 20 factors was 71.66 for VER, 69.47 for AER, 86.06 for EOP, and 85.19 for ECL.

Factor Loading Characteristics

Tables 2 and 3 summarize the results of visual inspection of the spatio-temporal distrubution of the EP derived factors. Results for the spatio-spectral distrubution of the

Table 1. Results of unrestricted principal components analysis.

Data Type (Initial Variables)	First 10 Eigenvalues	Number of Eigenvalues >1 (Data Reduction Factor)	Number Significant by Bartlett (Data Reduction Factor)	% Variance For First 20 Factors
VER	358.786	130	41	71.66
(2944)	219.816	(22.6)	(71.8)	
	212.064			
	140.576			
	125.240			
	117.768			
	106.985			
	102.938			
	94.350			
	82.098			
AER	386.255	170	13	69.47
(2944)	288.889	(17.3)	(226.5)	
	218.908			
	181.410			
	124.603			
	113.072			
	95.233			
	85.560			
	71.285			
	64.005			
EOP	541.003	74	26	86.06
(1536)	157.725	(20.8)	(59.1)	
	126.876			
	69.402			
	64.493			
	50.512			
	48.076			
	37.516			
	33.148			
	29.652			
ECL	584.181	78	23	85.19
(1536)	163.456	(19.7)	(66.8)	
	85.800			
	58.772			
	54.044			
	47.516			
	41.728			
	35.893			
	33.886			
	30.064			

Table 2. VER factor definition.

Data Factor Type Number	% Variance After Rotation	Status	Max Loading: Location, Latency & Description
VER 1	7.52	real	+.922 at FP1, 480 msec, broad bilat frontal
2	5.65	real	-.874 at FZ, 248 msec, medial frontal
3	4.60	real	-.843 at CZ, 92 msec, fronto-cental
4	4.29	real	-.868 at F4, 176 msec, fronto-cental
5	4.81	real	-.867 at O2, 388 & 488 msec, bilateral occipital oscillatory
6	3.19	artifact	-.892 at HEM, 396 msec, horizontal eye movement
7	2.31	real	-.762 at O2, 268 msec, occipital oscillatory
8	3.77	artifact	-.902 at CZ, 16 msec, muscle micro reflex/ERG
9	4.00	real	-.803 at P3, 212 msec, posterior quadrants
10	4.44	real	-.822 at FZ, 316, medial frontal
11	3.62	real	-.853 at P3, 296 msec, bi-parietal
12	2.71	real	-.860 at FZ, 124 msec, medial frontal
13	2.48	real	+.810 at OZ, 128 msec & also -.592 at OZ, 72 msec, occipital
14	2.79	real	-.859 at PZ, 360 msec, bi-parietal
15	2.58	artifact	-.819 at VEM, 144 and 472 msec, vertical eye movement/blink
16	2.41	artifact	-.848 at FP2, 60 msec & +.554 at FP2, 24 msec, ERG & blink
17	3.88	real	-.809 at CZ, at 412 msec, central
18	2.31	real	+.634 at O2, 172 msec, occipital oscillatory
19	2.22	artifact	+.713 at HEM, 192 msec, horizontal eye movement
20	2.06	artifact	+.818 at BTM, 376 msec, bilateral temporal muscle
	Total 71.66 %		

EEG derived factors are similarly shown in Tables 4 and 5. Note for all four sets of factors (Tables 2-5) that the percentage of variance accounted for by the factors does not necessarily show a continuous decrement for increasing factor number. This reflects reallocation of variance after the Varimax rotation. In general, visualization of factor loading patterns revealed smoothly varying functions without discontinuities; all images appeared "biological" in nature and many had the appearance of a simplified EP (e.g., figure 1) or spectral (e.g., figure 2) waveform. It was evident that certain factors were probably artifactual in nature on the basis of primary loading on one of the three (EP) or four (EEG) artifact channels (e.g., figure 5). In addition, factors were considered artifactual if loadings primarily involved single

Table 3. AER factor definition.

Data Factor Type Number	% Variance After Rotation	Status	Max Loading: Location, Latency & Description
AER 1	8.63	real	+.849 at FZ, 416 msec, bifrontal
2	5.35	real	-.885 at FZ, 148 msec, frontal
3	4.73	artifact	+.812 at F7 & -.574 at F8, 340-472 msec, horizontal eye movement
4	4.77	real	-.812 at P3, 304 msec, bi-parietal
5	3.77	real	+.894 at FZ, 104 msec, central-parietal
6	4.45	real	+.826 at FZ, 276 msec, frontal
7	2.72	artifact	+.889 at BTM, 200-500 msec, bilateral temporal muscle
8	3.18	real	+.866 at CZ, 64 msec, central
9	1.56	real	-.412 at CZ, 332 msec, central-parietal
10	2.71	artifact	-.780 at PZ, 480 msec, technical artifact
11	2.01	artifact	+.734 at CZ, 12 msec, muscle micro reflex
12	1.86	artifact	+.828 at PZ, 36 msec, early AER or muscle
13	3.86	artifact	+.828 at HEM, 52-440 msec, horizontal eye movement
14	7.14	real	-.906 at PZ, 216 msec, bi-parietal
15	3.04	artifact	-.860 at VEM, 244-500 msec, vertical eye movement/blink
16	3.84	real	-.748 at O1, 384 msec, broad bi-parietal-occipital
17	1.49	real	-.645 at T5, 144 msec, broad, biposterior
18	1.35	artifact	-.577 at VEM, FP1, FP2, 0-50 msec, ERG or eye artifact
19	1.24	artifact	+.545 at T6, 52 msec, focal right posterior temporal
20	1.75	real	-.672 at O1, 108 msec, broad biposterior
	Total 69.47 %		

Table 4. EOP factor definition.

Data Factor Type Number	% Variance After Rotation	Status	Max Loading: Location, Latency & Description
EOP 1	10.60	real	-.910 at OZ, 18-32 Hz, posterior beta
2	16.80	real	+.883 at CZ, 12-18 Hz, central slow beta
3	8.68	artifact	+.942 at FP1 & +.921 at VEM & FP2, 1.5 Hz, vertical eye movement or blink
4	4.15	artifact	+.945 at BTM, 1.5-32 Hz, bilateral temporal muscle
5	8.54	real	+.891 at F3, 26 msec, bifrontal beta
6	3.19	artifact	+.953 at BOM, pan spectral, posterior muscle
7	4.56	artifact	+.900 at T3 & +.725 T4, 14-32 Hz, bitemporal muscle
8	7.16	real	-.867 at CZ, 7 Hz, central
9	5.44	real	-.761 at PZ, 0.5-6 Hz, peaks 2.5 Hz, parietal delta
10	2.09	artifact	-.773 at VEM, 13-32 Hz, vertical eye movement/blink
11	1.79	real	+.625 at C3, 22-32 Hz, bilateral central beta
12	1.74	artifact	-.707 at FP2, 16-32 Hz, focal right prefrontal muscle
13	1.83	artifact	+.775 at HEM, 22-32 Hz, horizontal eye movement
14	1.45	artifact	+.840 at T6, 0 Hz, broad bi-posterior delta noise
15	1.23	real	-.575 at OZ, 10 Hz peak, classic alpha
16	2.30	artifact	+.683 at T5, 28 Hz, bi-posterior temporal muscle artifact
17	1.02	artifact	+.473 at HEM, 1 Hz, horizontal eye movement
18	1.12	artifact	-.481 at T4, 11.5 Hz, focal right mid temporal alpha
19	1.06	artifact	-.567 at F7, 15-32 Hz, focal left temporal beta
20	1.32	real	+.469 at CZ, broad peak at 23 Hz, central beta
	Total 86.06 %		

electrodes, especially bilateral mid temporal or posterior temporal electrodes in the higher spectral beta range (e.g., figure 6). Of the initial 80 factors, 38 were deemed artifactual on the basis of high loading on at least one artifact channel, or unusually focal loading on one or more artifact prone scalp channels despite absent artifact channel loading (usually focal beta) or EP factors loading at very early latencies suggestive of muscle microreflex (Bickford et al. 1964) or electroretinogram. Of the 38, 19 were recognized by high loading on an artifact channel and 19 were presumed artifactual by the other empirical criteria. Forty non-artifactual factors remained (14 VER, 11 AER, 8, EOP, 9 ECL). In Tables 2 to 5 the single electrode at which loading was greatest is given; in no instance were discrete, spatially isolated loadings noted for these "real" factors.

Figure 1 illustrates the factor loading pattern of AER factor 5 (Table 3). There is a clear peak at 104 msec. post stimulus latency (figure 1A) accompanied by a symmetrical fronto-central spatial distribution (figure 1B) with maximal loading of 0.894 at the mid frontal electrode, FZ. Note that there is minimal loading on the vertical eye movement channel (VEM) suggesting non-artifactual origins. This factor probably represents the classic N1 component of the AER. Note in general that the sign of

Table 5. ECL factor definition.

Data Factor Type Number	% Variance After Rotation	Status	Max Loading: Location, Latency & Description
ECL 1	18.46	real	-.876 at CZ, 14 Hz, central slow beta
2	12.48	real	+.834 at FZ, 27-30 Hz, fronto-central fast beta
3	10.30	real	+.876 at PZ, 0.5-6.5 Hz, peaks 2 Hz, posterior delta
4	5.76	real	+.824 at C4, 6-8.5 Hz, peaks 7.5 Hz, central theta
5	3.16	artifact	-.949 at BOM, pan spectral, peaks 15 Hz, posterior muscle
6	5.41	real	-.888 at OZ, 28 Hz, occipital beta
7	2.82	artifact	-.902 at VEM, 14-32 Hz, bifrontal eye muscle
8	3.42	artifact	+.904 at T3, 17-32 Hz, peaks 23.5 Hz, focal bi-midtemporal muscle
9	3.05	artifact	+.936 at BTM, 12-25 Hz, bitemporal muscle
10	3.83	real	-.813 at O1, 10.5 Hz, classic alpha
11	2.22	artifact	-.662 at FP1, 21-32 Hz, biprofrontal muscle
12	3.39	artifact	+.909 at VEM, FP1, FP2, 0.5-5 Hz, vertical eye movement and blink
13	1.22	real	+.432 at OZ, 9.5 Hz, biposterior alpha and 2nd harmionic
14	1.53	artifact	-.485 at T4, 14-32 Hz, focal right temporal muscle
15	1.99	artifact	-.780 at HEM, 17-32 Hz, horizontal eye movement
16	1.31	artifact	-.660 at T5, 21-32 Hz, focal left posterior temporal muscle
17	1.43	real	+.610 at CZ, 8.5 Hz, bianterior slow alpha
18	1.30	artifact	+.389 at T5, 0.5 Hz, biposterior temporal movement artifact
19	1.00	artifact	+.337 at OZ, 13 Hz, occipital muscle
20	1.11	real	-.452 at CZ, 26-28 Hz, central beta
	Total 85.19 %		

the correlation between the factor and its underlying variables (positive for AER factor 5) does not necessarily bear any meaningful relationship with the original polarity of the AER (negative for N1). Figure 2 shows the pattern for ECL factor 10 (Table 5). There is a clear peak at 10.5 Hz (figure 2A) accompanied by an occipital distribution (figure 2B) with maximal loading of 0.813 at left occipital electrode O1. Note the relatively low loadings on the occipital muscle channel (BOM) suggesting non-artifactual origins. This factor appears to correspond to classic occipital reactive alpha. Figure 3 illustrates VER factor 3 (Table 2). This factor has two post stimulus latency peaks, the larger negative loading of 0.843 at 92 msec. and a smaller positive loading of 0.452 at 160 msec. Note the slightly different spatial distributions for the two times of maximum loading. This factor demonstrates associations between variables generated at differing points in time and space.

Figures 4 and 5 illustrate two artifactual factors detected by high artifact channel loading and figure 6 shows an artifactual factor so designated on the basis of loadings on artifact prone scalp electrodes. Figure 4 shows AER factor 13 which appeared to represent vertical eye movement artifact (Table 3). Note the relatively noisy and low level appearance of the scalp electrode loadings with maxima seen in the prefrontal electrodes of 0.3 to 0.4 (FP1 and FP2, figure 4B). On the other hand VEM artifact channel demonstrates a loading of 0.860 (figure 4A top middle and figure 4B, top buttons). Figure 5 depicts ECL factor 15 representing horizontal eye movement muscle artifact (Table 5). Note once again (Fig 5A) the low amplitude, noisy, scalp electrode factor loading plots reaching a maximum of only 0.1 to 0.2 somewhat greater in the anterior temporal electrodes (F7, F8 - Fig 5B). Nonetheless this factor loads 0.780 on the horiziontal eye movement artifact channel (HEM, figure 5B left button and figure 5A top left trace) and 0.380 on the bilateral temporal muscle channel artifact (BTM). Although maximum loading was achieved at 25.5 Hz, note the continuous high loadings from 13 to 32 Hz over the entire beta band. ECL factor 8 (only topographic map shown - Fig 6) was designated artifactual on the basis of heavy loadings in the bilateral mid temporal electrodes (T7, T8) despite absent loadings on BTM or HEM. Loadings reached 0.904 for T7 and 0.661 for T8. The topographic map of the factor loading scores at the spectral point of maximum loading demonstrated characteristic high values in both temporal regions with a pattern recognized in clinical mapping studies as bi-temporal muscle artifact. Although maximal loading was achieved at 23.5 Hz, there was a broad distribution of high loadings from 12 to 32 Hz involving the entire beta band, a pattern also suggesting muscular rather than cerebral origin.

Statistical Evaluation

Age

All 202 normal subjects were grouped into the five decades indicated above. Linear discriminant analysis was performed on the basis of the 42 artifact free factors (Table 6A). The first cannonical discriminant function was highly statistically significant, Wilk's lambda 0.166, $p < 0.0001$, with an overall classification success of 66%. Note that the random outcome for this five group dis-

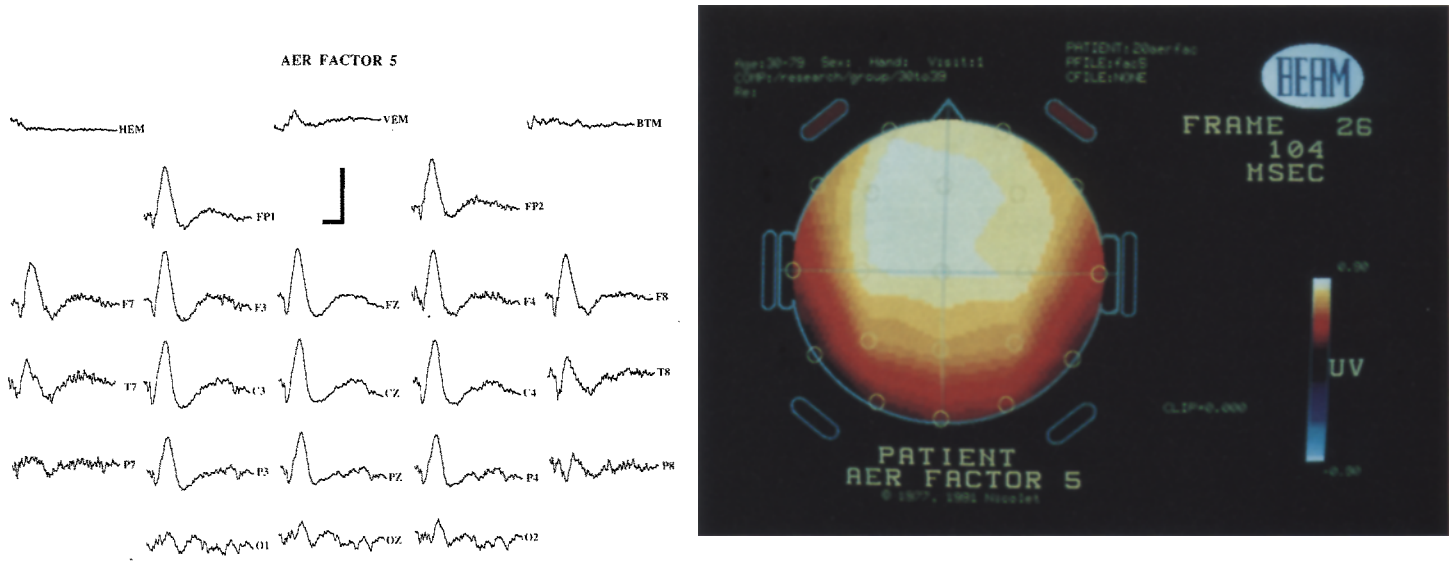


Figure 1. This figure displays the spatial and temporal extent of AER factor 5. In figure 1A the factor loadings are displayed as if they were evoked potential data. The 2944 factor loadings are divided into 23 curves of 128 loadings each corresponding to the 23 original evoked potentials. The name of the appropriate electrode is given to the right of each curve. Each curve corresponds to 512 msec time. Scale shown above indicates a loading value of 1.0 (vertical line) and 100 msec latency (horizontal line). Figure 1B is a schematic map of the head from vertex view with nose above, right ear to the right, left ear to the left, and occiput below. Data within the head outline represent the topographic distribution of factor loadings at 104 msec, peak of the curves shown in figure 1A. The head outline is surrounded by six "buttons" reflecting artifact channel data. The two buttons above the eye correspond to vertical eye movement and blink arifact (VEM). The button by the left ear corresponds to horizontal eye movement and/or temporal muscle artifact (HEM). The button to the right of the right ear reflects bilateral temporal muscle activity (BTM). The occipital two buttons correspond to occipital muscle activity (BOM). See text for electrode placement description. Red-orange corresponds to positive loadings and blue to negative loadings. Maximum loading is at electrode FZ, +0.899. This curve is believed to represent non-artifactual data corresponding to the N1 long latency AER component latency.

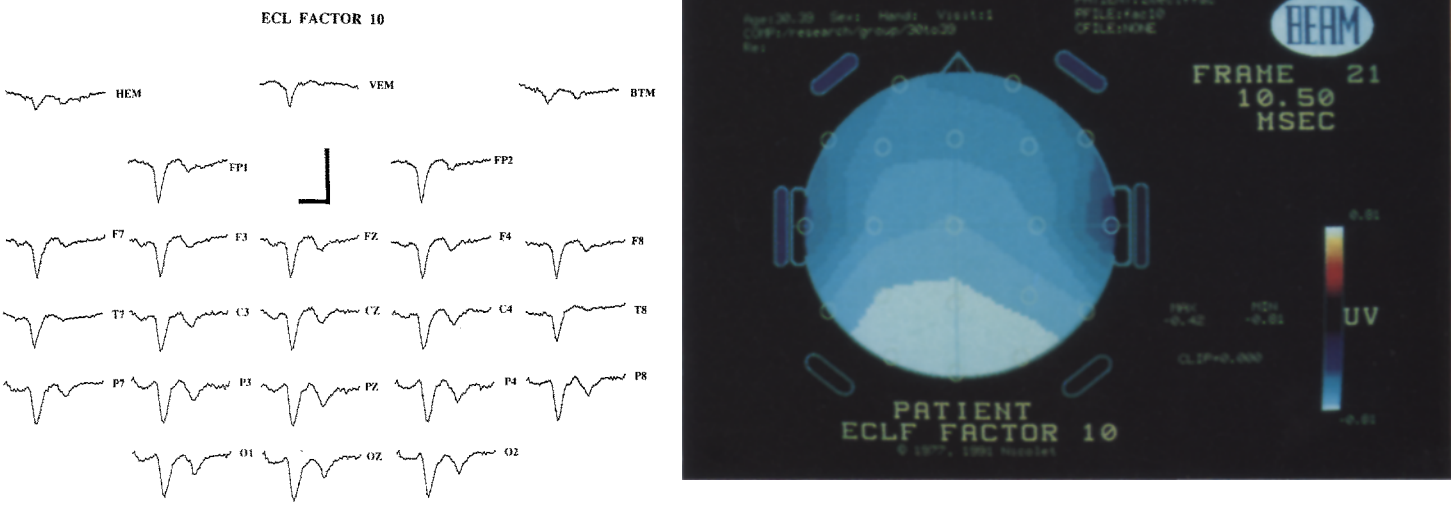


Figure 2. Both figures 2A and 2B are displayed similarly to the format of figure 1. Data correspond to the spatio-spectral distribution of ECL factor 10. The scale corresponds to a loading score of 1.0 (vertical line) and 10 Hz (horizontal line). Curves correspond to the original 24 EEG spectral curves of 64 data values from 0.5 to 32 Hz. Figure 2B depicts the topographic distribution of the loading factors at 10.5 Hz, peak of the curves in figure 2A. Maximum loading is achieved at electrode O1, -0.813. This factor represents non-artifactual data corresponding to the classic alpha spectral peak. Note minimal loadings in the artifact buttons.

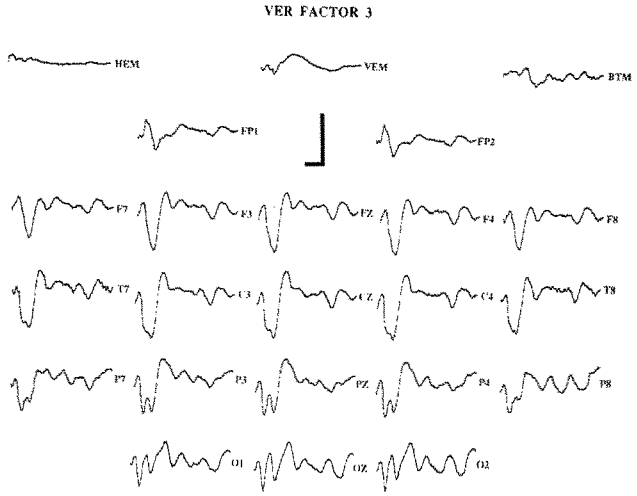
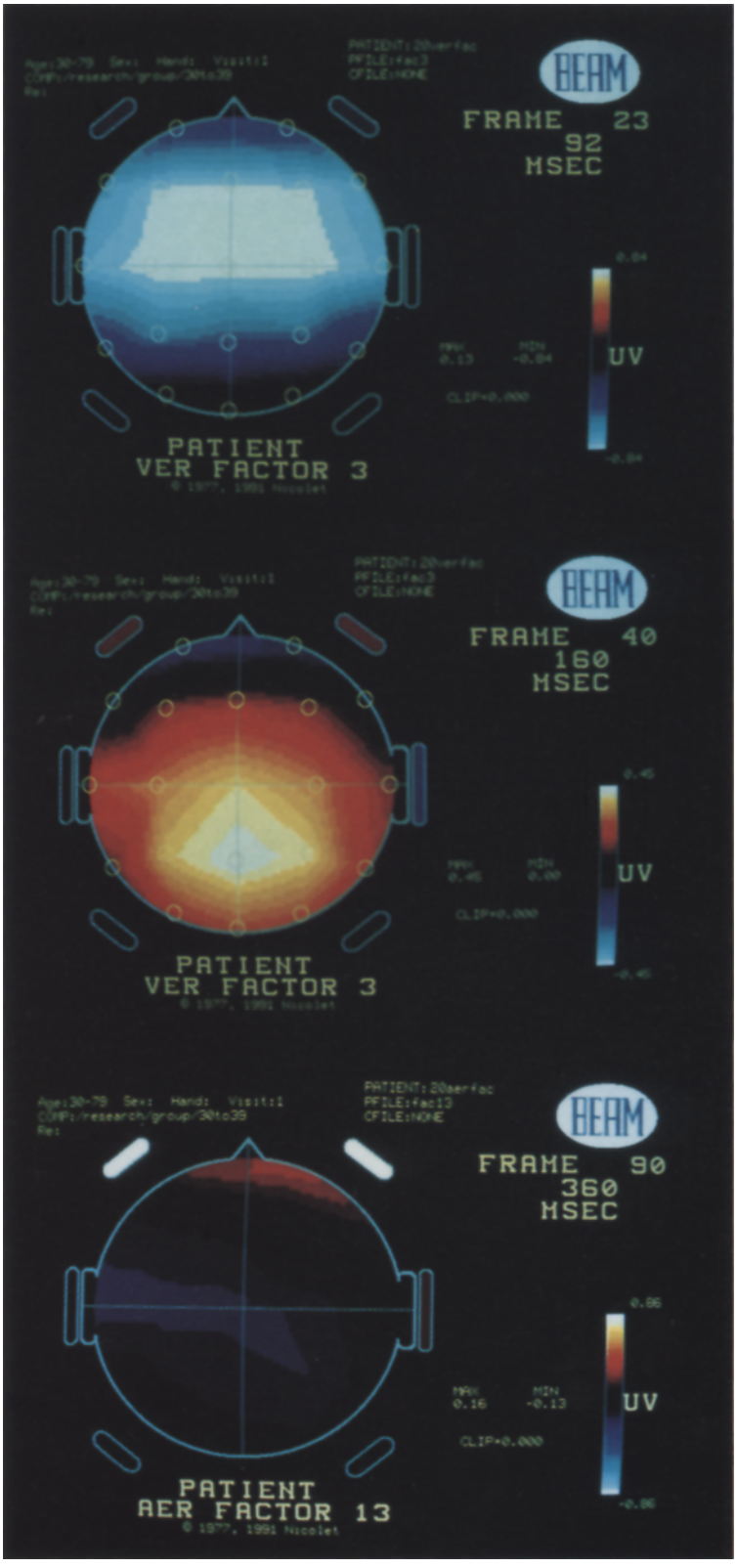


Figure 3. Figure 3A is displayed as for figure 1A and figures 3B and 3C are displayed as for figure 1B. Figure 3 shows VER factor 3. Figure 3B depicts the distribution of negative loadings at the 92 msec peak, -0.843 at electrode FZ. Figure 3C depicts the positive loadings at 160 msec, maximal at electrode PZ, -0.452. This factors illustrates the natural tendency of original data to associate across space and time.



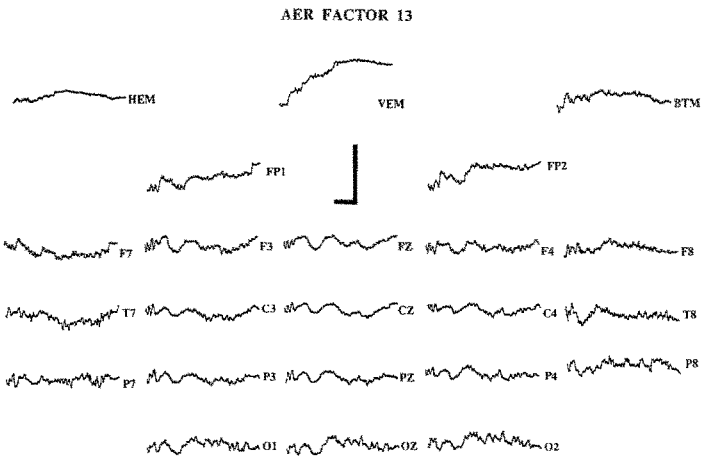
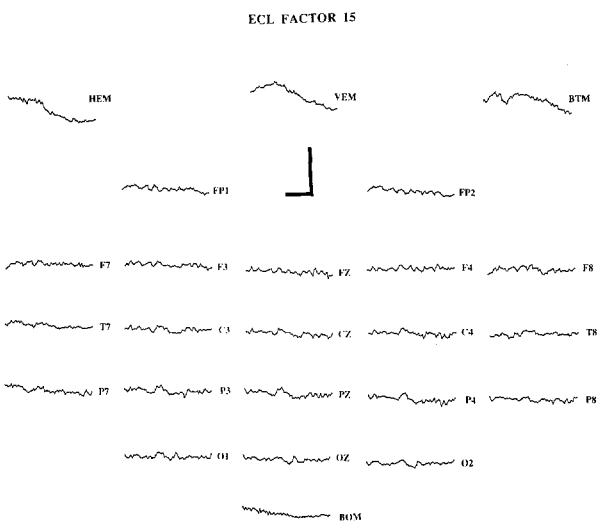


Figure 4. Figure 4 is shown as for figure 1. Illustrated is the spatio-temporal distribution of AER factor 13. It is an artifacual factor which loads maximally on the vertical eye movement channel (VEM) shown above in figure 4A - (VEM button is white, maximum). The topographic distribution at 360 msec is displayed in figure 4B, the time of maximal VEM loading, -0.860. This corresponds to a typical vertical eye movement artifact.



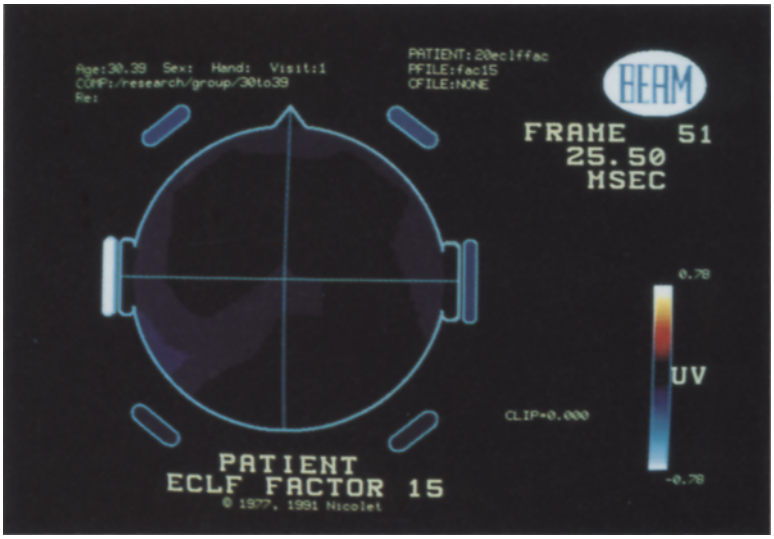


Figure 5. Figure 5 is shown as for figure 2. Illustrated is the spatio-spectral distribution of ECL factor 15 representing a typical horiziontal eye movement muscle artifact. The topographic map in figure 5B shows the frequency of maximum loading at 25.5 Hz, maximal in the horizontal eye movement channel (HEM button) at + 0.780 and bilateral temporal muscle channel (BTM button) at +0.380.

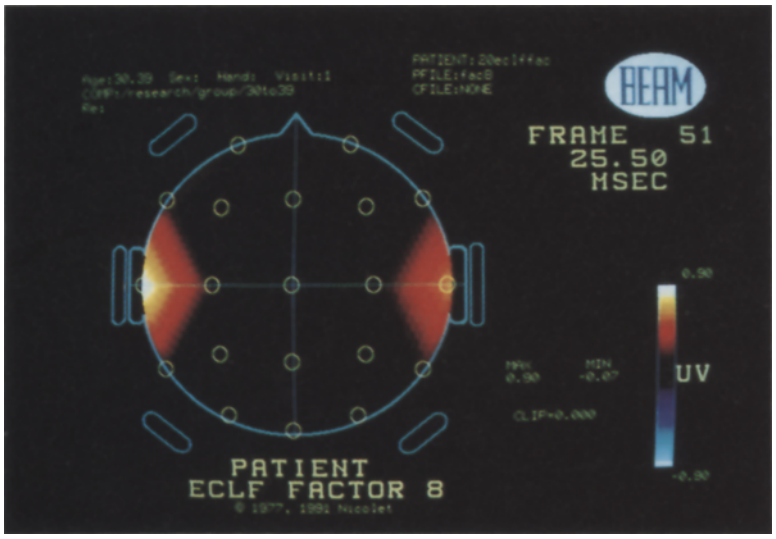


Figure 6. Figure 6 shows the topographic map for artifactual ECL factor 8 at 25.5 Hz, where loading was maximal at the left mid temporal electrode, T7 (+0.904) , and the right mid temporal electrode, T8 (+0.661). Display convention is as for figure 2B. Note the clearly artifactual nature corresponding to temporal muscle activity but without heavy loading on BTM or HEM.

criminant would be 20% in contrast to a 50% random outcome for a two group problem. Of the 10 factors most highly correlated with the discriminant function, eight were derived from EP data (five VER, three AER) and two from EEG data (EOP). Most factors involved symmetrical frontal and central areas. The two EEG factors involved parietal delta and occipital alpha.

When the same five group discriminant was restricted to the 38 artifactual variables, a different result was seen (Table 6B). Wilk's lambda increased to 0.494 and did not quite achieve significance at the 0.05 level. Factors correlating with the discriminant function included early muscle artifact and ERG signals, blink artifact, horizontal eye movement, and temporal muscle activity.

Multiple regression restricted to the 42 non-artifactual factors with age now treated as a continuous rather than a grouping variable yielded an overall Multiple R of 0.795, explaining 63 % of the variance (Table 6C). ANOVA on the residual was quite significant, $p < 0.0001$. Of the 10 most significant variables, nine were derived from EP data (five VER, four AER) and only one from EEG (EOP). Variables were similar to those in the 42 real factors discriminant analysis described above.

Gender

Using the 42 real factors, 88% of males and females were correctly discriminated overall with roughly equal success for both genders (male=92%, females=85%). Wilk's lambda was 0.45, significant at $p < 0.0001$ (Table 7A). Of the 10 most highly correlated factors with the discriminant function, five were EP derived (three VER, two AER) and five EEG derived (three EOP, two ECL). Factors included many sources including occipital oscillatory VER (factor 18), central theta (ECL factor 4), occipi-

Table 6. Statistical analyses: prediction of age.

A) AGE BY LINEAR DISCRIMINANT ANLAYSIS OF FIVE DECADE USING 42 "REAL" FACTORS

Wilk's lambda = 0.166 **Chisquare = 319.069 DF = 168** **p <= 0.0001**

10 most highly correlated factors with discriminant functions:

Factor	Correlation
AEP factor 6	0.282
VEP factor 10	-0.251
VEP factor 4	0.239
AEP factor 2	0.236
VEP factor 13	-0.190
VEP factor 3	0.188
VEP factor 17	0.181
EOP factor 9	0.176
EOP factor 15	0.161
AEP factor 16	-0.097

Group Membership Prediction: Overall Success = 65.84%

actual	% predicted group membership				
(N)	30	40	50	60	70
30 (29)	79.3	13.8	3.4	0.0	3.4
40 (41)	9.8	58.5	12.2	10.3	7.3
50 (29)	6.9	13.8	69.0	10.3	0.0
60 (57)	3.5	7.0	10.5	50.9	28.1
70 (46)	0.0	0.0	6.5	13.0	80.4

B) AGE BY LINEAR DISCRIMINANT ANALYSIS OF FIVE DECADE USING 38 "ARTIFACT" FACTORS

Wilk's lambda = 0.494 **Chisquare = 130.646 DF = 108** **p <= ns**

7 most highly correlated factors with first discriminant function:

Factor	Factor
AEP factor 12	EOP factor 19
VEP factor 16	EOP factor 10
AEP factor 13	AEP factor 3
VEP factor 6	

C) AGE BY MULTIPLE REGRESSION USING 42 "REAL" FACTORS

Multiple R = 0.795 **Variance explained = 63.24**

ANOVA on Residual: F = 6.32 (42, 158), p <= 0.0001

10 Most significant Variables by T Test:

Factor	T value	Factor	T value
AEP factor 2	4.13	EOP factor 5	3.06
AEP factor 6	3.72	VEP factor 4	3.03
VEP factor 10	-3.70	VEP factor 17	2.81
AEP factor 20	-3.49	AEP factor 8	2.56
VEP factor 3	3.47	VEP factor 13	-2.28

Table 7. Statistical analysis: prediction of gender.

A) GENDER BY LINEAR DISCRIMINANT ANALYSIS USING 42 "REAL" FACTORS

Wilk's lambda = 0.445 Chisquare = 144.877 DF = 42 p <= 0.0001

10 most highly correlated factors with discriminant function:

Factor	Correlation	Factor	Correlation
VEP factor 18	0.363	AEP factor 2	-0.181
ECL factor 4	-0.302	ECL factor 2	0.177
EOP factor 1	-0.265	AEP factor 6	-0.162
VEP factor 3	0.214	VEP factor 11	-0.152
EOP factor 9	-0.210	EOP factor 2	0.147

Grour Membership Prediction: Overall Success = 88.12%

actual % predicted group membership (N)	F	M
F(113)	85.0	15.0
M(89)	7.9	92.1

B) GENDER BY LINEAR DISCRIMINANT ANALYSIS USING 38 "ARTIFACT" FACTORS

Wilk's lambda = 0.697 Chisquare = 65.288 DF = 38 p <= 0.0038

10 most highly correlated factos with discriminant functions:

Factor	Factor
VEP factor 16	EOP factor 3
ECL factor 18	EOP factor 6
ECL factor 9	ECL factor 5
ECL factor 8	VEP factor 8
EOP factor 14	EOP factor 13

Group Membership Prediction: Overall Success = 73.76

actual % predicted group membership (N)	F	M
F(113)	72.6	27.4
M(89)	24.7	75.3

tal beta (EOP factor 1) among others.

Using the 38 artifactual factors (Table 7B), correct gender discrimination was reduced, but not to the chance level, at 74% overall (males=75%, females=73%). Wilk's lambda at 0.70 was increased but remained significant ($p < 0.004$). Factors most correlated with the discriminant function included ERG and blink (VER factor 16) and many temporal-occipital muscle factors.

Dementia

Table 8 summarizes linear discriminant function analyses between the 202 normal healthy adults and the 60 Alzheimer's patients. On the basis of the 42 real factors (Table 8A), Wilk's lambda was 0.52, significant at $p < 0.0001$. Ninety percent of subjects were correctly classified; normals=92%, Alzheimer patients=85%. Six of the 10 factors most correlated with the discriminant function were EEG derived (three EOP, three ECL) and four were EP derived (two AER, two VER). EEG derived factors represented the four most correlated factors. Factors, among others, represented central theta (EOP factor 8), parietal delta (EOP factor 9 and ECL factor 3), and slow beta (ECL factor 1). EP factors were largely frontal.

Restriction of analyses to the 38 artifactual factors resulted in little degradation of result Table 8B). Wilk's lambda decreased slightly (0.51) and was highly significant ($p < 0.0001$). Overall classification was slightly reduced to 87%; normals=87%, Alzheimers=88%. However, artifact derived factors were slightly better in discriminating just the Alzheimer's patients. Half the discriminating factors derived from eye movement or muscle activity in the eye channels (AER factors 3 and 15, EOP factors 17 and 3, and ECL factor 12). Remaining factors represented artifactual delta and beta or other artifact.

Table 8. Statistical analysis: prediction of dementia

A) DEMENTIA BY LINEAR DISCRIMINANT ANALYSIS USING 42 "REAL" FACTORS

Wilk's lambda = 0.518 Chisquare = 157.194 DF = 42 p <= 0.0001

10 most highly correlated factors with discriminant function:

Factor	Correlation	Factor	Correlation
EOP factor 8	0.440	VEP factor 2	0.262
EOP factor 9	0.375	VEP factor 2	0.256
ECL factor 3	-0.358	AEP factor 1	0.237
ECL factor 1	-0.345	EOP factor 2	0.229
VEP factor 12	0.286	AER factor 20	0.228

Grour Membership Prediction: Overall Success = 90.46%

actual %	predicted group membership	
(N)	N	A
N(202)	92.1	7.9
A(60)	15.0	85.0

B) DEMENTIA BY LINEAR DISCRIMINANT ANALYSIS USING 38 "ARTIFACT" FACTORS

Wilk's lambda = 0.512 Chisquare = 161.185 DF = 38 p <= 0.0001

10 most highly correlated factos with discriminant functions:

Factor	Factor
AEP factor 3	ECL factor 12
EOP factor 17	EOP factor 3
ECL factor 19	EOP factor 18
AEP factor 10	ECL factor 18
EOP factor 14	AER factor 15

Group Membership Prediction: Overall Success = 87.02%

actual %	predicted group membership	
(N)	N	A
N(202)	86.6	13.4
A(60)	11.7	88.3

Discussion

Principal components analysis (PCA) was performed on the brain electrical activity of 202 healthy adults employing singular value decomposition, a computational algorithm which facilitates analysis of data sets with large numbers of initial variables and smaller numbers of subjects. The goal was to derive factors from the thousands of variables characteristically produced in qEEG mapping studies. Prior studies have often made simplifying assumptions such as restricting analysis to the time dimension and neglecting space so as to reduce variable number to a more manageable size. Such computational restrictions are no longer necessary given the ready availability of fast microprocessors with cheap memory and PCA algorithms optimized to manage computational singularities (Golub 1989). As anticipated our results demonstrated a clear clustering of evoked portential (EP) factor loading scores at regions defined by both time and space (figures 1 and 3). VER factor 3 was remarkable in that two separate space - time cluster points were clearly visible (figure 3), one at 92 msec in the fronto-central region and another at 160 msec in the parietal vertex region. Similarly spectral factor loading patterns clustered at regions defined by both frequency and space. ECL factor 10, for example, appears to represent eyes closed alpha on the basis of the occipital location and 10.5 Hz peak loading (figure 2). Almost all factor loading score profiles, when visualized in two dimensions (time-space or frequency-space), demonstrated considerable variation across both dimensions. Those showing little variation in one dimension (e.g., EOP factor 6 and AER factor 12) were clearly artifactual in origin. The importance of allowing all data points to freely associate when performing neurophysiologic PCA is thereby illustrated.

Silberstein et al. (1992) suggest that much of the spatial variation in scalp recorded data, as determined by spatial PCA, results from physical considerations such as the spherical nature of the skull. These results call into question the need to evaluate the spatial dimension at all. However, Valdes et al. (1990) feel that such physical considerations dominate spatial PCA only for signals of low spatial frequency. It may be that restriction of PCA exclusively to the spatial data dimension serves to over emphasize physical characteristics and facilitates contamination by spherical harmonics. Use of unrestricted PCA, as reported herein, would be freer of such potential contamination for factors are free to demonstrate associations across space and time (or frequency).

Visualization of factor loading scores in time-space (EP data) or frequency-space (EEG data) yielded in almost every case patterns that were recognizable from experience evaluating unprocessed EP or EEG spectral data (figure 1 to 6). On the basis of heavy loadings upon channels placed so as to be maximally sensitive to artifact, or upon other unique characteristics such as loading on a single scalp channel, so called "artifact" factors were readily identified. Of the 80 factors studied almost half (38) were clearly artifactual, attesting to the high amount of potential scalp electrode contamination that must be expected even when great care is taken to both reduce artifact at time of data collection and at time of subsequent off-line analysis Remaining "real" factors ranged from those obviously corresponding to reactive occipital alpha (ECL factor 10 and EOP factor 15) to those coresponding to known EP components such as the AER P2 (AER factor 14).

PCA also provides an indication of the underlying dimensionality of data structures. It is important and reassuring to note the relatively low intrinsic dimensionality of neurophysiologic data (Table 1) since such numbers are often used to estimate the potential for false positives or capitalization on chance (Duffy et al. 1986; Oken and Chiappa 1986). It would be clearly inappropriate to make such estimates solely on the basis of input variable number. Reduction in dimensionality by PCA, calculated by dividing the number of variables input to PCA by the number of "significant" factors produced, was considerable. Unfortunately there is no absolute and/or universally accepted technique to estimate the number of "significant" factors resulting from PCA and thereby unequivocally estimate the intrinsic dimensionality of the underlying data structure. Results for two common techniques, the "rule of 1" and Bartlett's test are both shown in Table 1. By the "rule of 1", reduction in dimensionality ranged from 17 to 22 across all modalities. By Bartlett's test it was somewhat greater, varying from 59 to 227. Between 69% and 86% of the total variance was accounted for by just the first 20 factors. This reduction in dimensionality should come as no surprise. It stems from the high intercorrelation among primary neurophysiologic variables clearly recognizable by the smooth, continuous appearance data when evaluated across time or frequency and space. Indeed any data point standing apart from or not easily predicted by its neighbors is commonly deemed artifactual. Thus although there have been legitimate questions as to whether factors should be taken as one to one analogues of underlying biological generators (Achim et al. 1988; Rösler et al. 1981; Scherg et al. 1985; Van Rotterdam 1970; Wastell 1979; Wood et al. 1984), it is widely agreed that PCA is useful in providing practical estimates of electrophysiological data dimensionality and that when performed such dimensionality is far less than the number of variables collected (Duffy et al. 1990). Although a study of 202 healthy adults may appear large, the number of subjects is relatively small when compared to the number of variables entered into analysis. Accordingly the ultimate answer as to underlying demensionality must await for an even larger study. In this report case number was limited by our wish to study only subjects whose health was well documented; a process demanding of both resources and time.

It is, of course, insufficient to assume that factors represent the useful essence of the data structure without some practical test of their utility. Accordingly we grouped our healthy subjects into five decade groups from 30 to 80. The resulting discriminant function was highly significant and decade was predicted with an overall classification success of 66% in a situation where chance outcome would be 20% (Table 6A). Using multiple regression, neurophysiologic factors accounted for 63% of the variance in predicting age as a continuous variable (Table 6C). Again using the 42 "real" factors 88% of all subjects correctly categorized by gender (Table 7A). Demented and healthy, normal subjects were 90% correctly identified (Table 8A). Thus a priori factors derived solely from the neurophysiologic data showed potential for use in clinical diagnosis and research.

It is also interesting to note the relatively high correlation of EP factors with the discriminant functions. As can be seen in Tables 6-8, EP data provided the highest correlating factor for two (age and gender) of the three discriminants. Results suggest that long latency data may be underutilized in clinical practice and research perhaps owing to inefficiencies in extracting useful measures from raw data by visual inspection alone. EP factors derived from PCA may ultimately extend the value of such data by providing useful a priori measures or components.

The biological or medical "meaning" of all factors has yet to be fully determined. Almost half are artifactual and about one quarter appear, on the basis of their load-

ing patterns, to represent familiar entities (alpha, N1,P2, etc). But final definition will rest upon completion of a series of evaluations where factors are utilized in correlational or discriminant analyses. Final factor "meaning" will be derived by a table for each factor indicating the types of studies in which it proved useful. For example based upon our own pilot studies one can infer that EOP factor 8 (central theta - Table 4) is useful in discrimination of dementia (Table 8A). Much more work is needed before all factors can be more clearly understood.

Most surprising was the relative success of discriminations based solely upon factors recognized as artifactual. Although unsuccessful in predicting age (Table 6B), discriminant functions based upon the 38 artifactual factors fared well predicting gender (74%, Table 7B) and dementia (87%, Table 8B). Thus neurophysiologic artifact can be so group specific as to have discriminating power. Dementia, for example, was discriminated largely on the basis of artifactual factors deriving from eye movement. Classification of just the demented patients was slightly better for the artifactual discriminant (88%) than with the real discriminant (85%) - see Table 8. This might have been predicted since we have observed that demented patients appear to have more difficulty controlling eye movements than normal subjects and many have accentuated blink rates. Differences in muscle tone and subsequent muscle artifact and in early eye potentials appear to be in part responsible for the successful gender discrimination. We would not, on the basis of our experience, have been able to anticipate this group specific artifact.

In summary, PCA constitutes a powerful tool for meaningful reduction of all types of data. Its application to large neurophysiologic data sets may prove a fruitful direction to follow especially when input variables are allowed to freely associate in an unrestricted manner. Factors resulting from unrestricted PCA on qEEG data sets demonstrate complex associations across time or frequency and space demonstrating the need to avoid artificial restriction of PCA to one dimension or another. Uses of PCA include recognition of artifact, estimation of data dimensionality, and the parsimonious derivation of measures for subsequent descriptive and/or discriminant analyses. It is our hope that unrestricted PCA will eventually prove useful as a supplement to other methods for neurophysiologic data reduction including the neurometric approach of John et al. (1977, 1988) and Descriptive Data Analysis of Abt (1987, 1988).

Artifact is surprisingly pervasive, substantially contributing to almost 50% of resulting factors, and is not necessarily evenly distributed across categories when subjects are grouped. Under some circumstances this might lead to successful group classification on the basis of artifact alone. Although artifact sensitive electrodes were extensively used for our analyses not all artifactual factors demonstrated loading on these channels. Approximately half of the artifactual factors had to be recognized by empirical criteria and prior experience with such data. Much progress needs to be made in the generation of apriori measures from neurophysiologic data that are artifact free on the basis of objective rather than subjective criteria. The work of Semlitsch (Semlitsch et al. 1986), who demonstrated that eye blink could be removed by partial covariance of eye blink channels with scalp channels, is an important step in the right direction. The possibility of extending this approach to muscle artifact is a next step.

References

Abt, K. Descriptive data analyses: A concept between confirmatory and exploratory data analysis. Methods of Information in Medicine, 1987, 26: 77-88.

Abt, K. . Descriptive data analysis (DDA) in Quantitative EEG studies. In D. Samson-Dollfus,J. D. Guieu,J. Gotman, and P. Etvènon (Eds.), Statistics and Topography in Quantitative EEG (pp. 150-160). Amsterdam: Elsevier, 1988.

Achim, A., Richer, F., Alain, C., and Saint-Hilaire, J. M. A test of model adequacy applied to the dimensionality of multichannel average auditory evoked potentials. In D. Samson-Dollfus,J. D. Guieu,J. Gotman, and P. Etvenon (Eds.), Statistics and Topography in Quantitative EEG (pp. 161-171). Amsterdam: Elsevier, 1988.

Bartels, P. H. Numerical evaluation of cytologic data IX. Search for data structure by principal components transformation. Anal. Quant. Cytol., 1981a, 3: 167-177.

Bartels, P. H. Numerical evaluation of cytologic data VIII. Computation of the principal components. Anal. Quant. Cytol., 1981b, 3: 83-90.

Bartlett, M. S. Tests of significance in factor analysis. Brit. J. of Stat. Psychol., 1950, 3: 77-85.

Bickford, R. G., Cody, D. T., Jacobsen, J. L., and Lambert, E. H. Fast motor systems in man: Physiopathology of the sonomotor response. Trans. Am. Neurol. Assoc., 1964, 89: 56-58.

Cooley, W. W., and Lohnes, P. R. Multivariate Data Analysis. New York: J. Wiley and Sons, 1971.

Donchin, E. A multivariate approach to the analysis of average evoked potentials. IEEE Trans. Biomed. Eng., 1966, BME-13: 131-139.

Duffy, F. H. Topographic display of evoked potentials: Clinical applications of brain electrical activity mapping (BEAM). Ann. N.Y. Acad. Sci., 1982, 388: 183-196.

Duffy, F. H. . Clinical decision making in quantified electroencephalographic analysis. In D. Samson-Dollfus,J. Gotman,J. D. Guieu, and P. Etevenon (Eds.), Statistics and Topography in Quantitative EEG (pp. 9-26). Paris: Elsevier, 1988.

Duffy, F. H. . Topographic mapping of brain electrical activity: Clinical applications and issues. In K. Maurer (Eds.), Topographic Brain Mapping of EEG and Evoked Potentials (pp. 19-52). Heidelberg, Germany: Springer-Verlag, 1989.

Duffy, F. H., Albert, M. S., and McAnulty, G. Brain electrical

activity in patients with presenile and senile dementia of the Alzheimer's type. Ann. Neurol., 1984a, 16: 439-448.

Duffy, F. H., Albert, M. S., McAnulty, G., and Garvey, A. J. Age - related differences in brain electrical activity mapping of healthy subjects. Ann. Neurol., 1984b, 16: 430-438.

Duffy, F. H., Bartels, P. H., and Neff, R. N. A response to Oken and Chiappa. Ann. Neurol., 1986, 19: 494-495.

Duffy, F. H., Burchfiel, J. L., and Lombroso, C. T. Brain electrical activity mapping (BEAM): A method for extending the clinical utility of EEG and evoked potential data. Ann. Neurol., 1979, 5: 309-321.

Duffy, F. H., Jones, K., Bartels, P., Albert, M., McAnulty, G. B., and Als, H. Quantified neurophysiology with mapping: Statistical inference, exploratory and confirmatory data analysis. Brain Topog., 1990, 3: 3-12.

Gasser, T., Bacher, P., and Mocks, J. Transformation towards the normal distribution of broad band spectral parameters of the EEG. Electroencephalogr. Clin. Neurophysiol., 1982,. 53: 119-124.

Golub, G. H. Matrix Computations (2 ed.). Baltimore, MD: Johns Hopkins University Press, 1989.

Hachinski, J. Cerebral blood differentiation of Alzheimer's disease from multi-infarct dementia. In R. Katzman,R. D. Terry, and K. T. Bick (Eds.), Alzheimer's Disease: Senile Dementia and Related Disorders (pp. 97-104). New York: Raven Press, 1978.

Harner, R. N., and Riggio, S. Application of singular value decomposition to topographic analysis of flash-evoked potentials. Brain Topog., 1989, 2: 91-98.

Harner, R. N., Warner, J. A., and Riggio, S. Localization of disordered brain function by singular value decomposition and spatial rotation of orthogonal features of the multichannel flash-evoked potential(in press). Brain Topog., 1991.

Horst, P. Factor Analysis of Data Matrices. New York: Holt, Rinehart, Winston, 1965. Hotelling, H. Analysis of complex statistical variables into principal components. J. Ed. Psychol., 1933, 24: 417-441.

John, E. R., Karmel, B. Z., Corning, W. C., Easton, P., Brown, D., Ahn, H., John, M., Harmony, T., Prichep, L., Toro, A., Gerson, I., Bartlett, F., Thatcher, R., Kaye, H., Valdes, P., and Schwartz, E. Neurometrics. Science, 1977, 196: 1393-1410.

John, E. R., Prichep, L. S., Fridman, J., and Easton, P. Neurometrics: Computer-assisted differential diagnosis of brain dysfunctions. Science, 1988, 239: 162-169.

John, E. R., Walker, P., Cawood, D., Rush, M., and Gehrmann, J. Factor analysis of evoked potentials. Electroencephalogr. Clin. Neurophysiol., 1973, 24: 33-43.

Jones, K. J. Relation of R-analysis factor scores and Q-analysis factor loadings. Psychol. Reports, 1967, 20: 247-249.

Kaiser, H. F. The application of electronic computers to factor analysis. Ed. Psych. Measmt., 1960, 20: 141-151.

Kaiser, H. J. A Varimax criterion for analytic rotation in factor Analysis. Psychometrika, 1958,. 23: 187-200.

Kaiser, H. J. A computer program for Varimax rotation factor analysis. Ed. Psych. Measmt., 1959, 19: 413-420.

Kavanagh, R. N., Darcey, T. M., and Fender, D. H. The dimensionality of the human visual evoked scalp potential. Electroencephalogr. Clin. Neurophysiol., 1976, 40: 633-644.

Lopes da Silva, F. H. . Event-related potentials: Methodology and quantification. In E. Niedermeyer and F. Lopes da Silva (Eds.), Electroencephalography Basic Principles, Clinical Applications and Related Fields (pp. 763-772). Baltimore-Munich: Urban and Schwarzenberg, 1987.

Maier, J., Dagnelie, G., Spekreijse, H., and van Dijk, B. W. Principal components analysis for source localization of VEPs in man. Vision Res., 1987, 27: 165-177.

Molfese, D. L., Buhrke, R. A., and Wang, S. L. The right hemisphere and temporal processing of consonant transition durations: electrophysiological correlates. Brain Lang., 1985,. 26, 49-62.

Oken, B. S., and Chiappa, K. H. Statistical issues concerning computerized analysis of brainwave topography. Ann. Neurol., 1986, 19: 493-494.

Rawlings, R. R., Rohrbaugh, J. W., Begleiter, H., and Eckardt, M. J. Spectral methods for principal components analysis of event-related brain potentials. Comput. and Biomed. Res., 1968, 19: 497-507.

Rogers, L. J., and Douglas, R. R. Comparative factor analysis models for an empirical study of EEG data, III: resolution of stationary generating signals with confirmation by variable deletion. Internat. J. Neurosci., 1984, 24: 167-183.

Rösler, F., and Manzey, D. Principal components and varimax-rotated components in event-related potential research: some remarks on their interpretation. Biol. Psychiatry, 1981, 13: 3-26.

Scherg, M., and Von Cramon, D. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol., 1985, 62: 32-44.

Seal, H. L. . Principal components. In H. L. Seal (Eds.), Multivariate Statistical Analysis for Biologists (pp. 101-122). New York: John Wiley and Sons, 1964.

Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. Pschophysiol., 1986, 23: 695-703.

Silberstein, R. B., Cadusch, P. J., and Schier, M. A. Effects of conduction on spatial components analysis applied to brain electrical activity. Brain Topography, 1992, 4(4): 267-276.

Skalka, H., and Holman, J. Effect of pupillary dilatation in flash VER testing. Doc. Ophthalmol., 1986, 63: 321-324.

Skrandies, W., and Lehmann, D. Spatial principal components of multichannel maps evoked by lateral visual half-field stimuli. Electroencephalogr. Clin. Neurophysiol., 1982, 54: 662-667.

Suter, C. M. Principal component analysis of average evoked potentials. Exp. Neurol., 1970, 29(2): 317-327.

Valdes, P., Bosch, J., Grave, R., Hernandez, J., Riera, J., Pascual, R., and Boscay, R. Frequency domain models of the EEG. Brain Topography, 1992, 4(4): 309-319.

Van Rotterdam, A. Limitations and difficulties in signal processing by means of the principal-components analysis. I.E.E.E. Trans. Biomed. Eng., 1970,. BME-17: 268-269.

Wastell, D. G. On the independence of P300 and the CNV: A short critique of the principal components analysis of donchin et al. (1975). Biol. Psychiat., 1979, 9: 171-176.

Weintraub, S., Baratz, R., and Mesulam, M. M. Assessment of daily living in Alzheimer's disease. In S. Corkin,K. L. Davis, and J. H. Growdon (Eds.), Alzheimer's Disease: A Review of Progress (pp. 189-192). New York: Raven Press, 1982.

Wood, C. C., and McCarthy, G. Principal component analysis of event-related potentials: Simulation studies demonstrate

misallocation of variance across components. Electroenceph. Clin. Neurophysiol., 1984, 59: 249-260.

Zar, J. H. Biostatistical Analysis. Englewood Cliffs, N. J.: Prentice Hall, 1984.