Published in final edited form as:

J Neurosci Methods. 2009 March 30; 178(1): 228–236. doi:10.1016/j.jneumeth.2008.11.021.

Evoked Potential Variability

Lingli Hu¹, Nash N. Boutros³, and Ben H. Jansen^{1,2}

¹Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204-4005

²Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX 77204-4005

³Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI 48201

Abstract

An unsupervised correlation-based clustering method was developed to assess the trial-to-trial variability of auditory evoked potentials (AEPs). The method first decomposes single trials into three frequency bands, each containing activity primarily associated with one of the three major AEP components, i.e., P50, N100 and P200. Next, single-trial evoked potentials with similar post-stimulus characteristics are clustered and selectively averaged to determine the presence or absence of an AEP component.

The method was evaluated on actual AEP and spontaneous EEG data collected from 25 healthy participants using a paradigm in which pairs of identical tones were presented, with the first stimulus (S1) presented 0.5 s before the second stimulus (S2). Homogeneous, well-separated clusters were obtained and substantial AEP variability was found. Also, there was a trend for S2 to produce fewer ‘complete’ (and significantly smaller) responses than S1. Tests conducted on spontaneous EEG produced similar clusters as obtained from EP data, but significantly fewer stimuli produced responses containing all three EP components than seen in AEP data. These findings suggest that the clustering method presented here performs adequately to assess trial-to-trial EP variability. Also, the results suggest that the sensory gating observed in normal controls may be caused by the fact that the second stimulus generates fewer ‘responsive’ trials than the first stimulus, thus resulting in smaller ensemble averages.

Keywords

Auditory evoked potentials; clustering; single-trial variability; sensory gating

1 Introduction

Evoked potentials (EPs) are the brain’s responses to external sensory stimulation. The evoked responses in the mammalian cortex and the resulting behavioral responses exhibit a large variability (Arieli et al., 1996). This is true at the inter-individual level (Campanella et al., 1999) and at the trial-to-trial level (Haig et al., 1995; Zouridakis et al., 1997). The cause of this intermittent responsiveness is unknown, but some relation between evoked potential (EP) morphology and pre-stimulus EEG characteristics seems to exist (Brandt et al., 1991; Jansen

Corresponding Author: Ben H. Jansen, Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204-4005, U.S.A., (713) 743-4431 (office), (713) 743-4444 (FAX), E-mail: bjansen@uh.edu .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and Brandt, 1991; Rahn and Başar, 1993; Haig and Gordon, 1998; Kruglikov and Schiff, 2003).

The measurable changes in the ongoing electroencephalogram (EEG) produced by sensory stimulation either reflect additive, afferent activity, stimulus-related phase reorganization, or a combination of these effects. The EEG changes are often small and ensemble averaging is needed for enhancement. Given the relatively low signal-to-noise ratio of the evoked activity over the spontaneous EEG, it is conceivable that single trial responses either contribute *constructively* to the ensemble average, i.e., the post-stimulus activity is in phase with the ‘true’ ensemble average, or *deconstructively* when the post-stimulus activity is out of phase with the ‘true’ ensemble average. In this context, post-stimulus activity is considered to be a mixture of stimulus-related and spontaneous EEG activity. It is also possible that not all stimuli produce a response (Ford et al., 1994; Wagner et al., 2000; Rodionov et al., 2002), which would give rise to underestimating the magnitude of the response.

It has been reported that the response produced by healthy subjects to a second auditory stimulus is substantially smaller than to the first, identical stimulus, preceding the second one by 0.5 s (Bramon et al., 2004). It is unclear what the mechanism for this behavior is, but a reduced ability to phase reorganize the EEG following stimulus repetition has been reported (Jansen et al., 2003). In order to explore if repeated stimuli produce fewer responses than non-repeated stimuli, we developed a procedure to determine what components, if any, were present in single trial responses.

The next section describes the basic method, which is followed by a presentation of the experiments conducted to fine-tune the algorithm, and to evaluate its performance on actual data. A discussion concludes this paper.

2 Methods

The auditory EPs studied here typically contain three components: P50, a positive component with a latency between 30 and 80 ms; N100, a negative component between 80 and 150 ms; and P200, a positive component between 150 and 250 ms. We use a separate cross-correlation procedure to detect the presence of each of the three components. To reduce (but not completely eliminate) the number of ‘false alarms’, the cross-correlation procedures do not operate on the wide-band signal, but each works on a different frequency band, selected to capture the activity associated with the EP component being searched for. The cross-correlation values are used in a clustering procedure to group similar responses together. The ensemble averages for each cluster are examined to determine if they contain a P50, N100 or P200 component, and if they do, all members of that cluster are considered to have that EP component. In this way, each single trial response is coded to reflect which, if any, of the three EP components are present. The specifics of the method are explained next.

2.1 Single trial decomposition

The Piece-wise Prony Method (PPM, Garoosi and Jansen, 2000) was used to decompose the EPs into the 0–2 Hz, 2–4 Hz, 4–8 Hz, 8–12 Hz and 12–16 Hz frequency bands. The PPM decomposes signals into exponentially decaying or increasing sinusoidal components with *arbitrary* duration and starting time. Also, these exponential components can have abrupt changes in amplitude, frequency, phase and exponential decay/growth rate. The choice of such “basis functions” is justified given that a variety of intra-cortical experiments have revealed that the response to a single electrical impulse in an awake and alert animal resembles a decaying sinusoidal oscillation (Başar and Ungan, 1973). Furthermore, ensemble averaged EPs have been modeled successfully as a summation of decaying oscillatory components (Turetsky et al., 1990).

The PPM approximation to a data segment is given by

$$\hat{x}[n] = \sum_{k=1}^p A_k \exp^{\frac{\alpha_k}{f_s}(n-b_k)} \cos(2\pi \frac{f_k}{f_s}(n-b_k) + \Theta_k) \cdot (u(b_k) - u(e_k)), \quad (1)$$

where f_s is the sampling frequency, p is the number of components, and component k has frequency f_k , amplitude A_k , rate of change α_k , and phase Θ_k . The arbitrary start and end time of the components (in terms of number of samples) are represented by b_k and e_k , respectively. The PPM uses an iterative procedure to determine the start and end times, amplitude, frequency, phase and exponential decay/growth rate of each sinusoid, starting at the lowest frequency band. Once an optimal decomposition has been obtained, it is subtracted from the original signal and an approximation to the next frequency band is determined. The procedure generally halts at 16 Hz.

PPM components from several frequency bands were combined to capture activity related to specific EP components. The decision which frequency bands to use for each EP component was made heuristically, taking into account the duration of each component. For example, P50 has latency between 30 and 80 ms, and is preceded by the N30, leading to a maximum duration of 50 ms. Interpreting this as a half period, the frequency components representing P50 should be 10 Hz or higher. Therefore, P50 is represented by the sum of PPM components in the 8–12 Hz and 12–16 Hz. This sum for trial i is referred to as $P50a_i$. Similarly, it was determined that N100 and P200 are primarily in the 2–8 Hz and 0–4 Hz frequency bands, respectively, and the sums for these bands for trial i are referred to as $N100a_i$ and $P200a_i$. A validation of these choices is presented in Section 3.1. It should be pointed out that neither of these bands shows activity *exclusively* associated with an EP component; in fact, each component shows up in more than one of these bands and the N100 and P200 band partially overlap. However, the bands were selected such that, for example, the 0–4 Hz band shows more P200-related activity than any of the other bands, with the same being true for the 2–8 Hz N100 band, and the 8–16 Hz P50 band.

2.2 Correlation-based clustering approach

Clustering is the process by which similar objects (single trials in this case) are grouped together. Objects are typically represented by (quantitative) features and the (e.g., Euclidean) distance between two objects is used as a measure of their similarity. In the agglomerative hierarchical clustering method used here, each single trial i represents a cluster C_i . In other words, there are as many clusters as single trials. The algorithm searches for the two nearest clusters, say C_i and C_j , and they are merged. This process is repeated until the nearest distance exceeds some threshold. Measuring the distance between two clusters can be done in different ways (Duda et al., 2001), but we have elected to use the ‘complete linkage’ method. In the complete linkage method, the distances between each element in cluster C_i and each element in C_j are computed, and the largest such distance is used to measure the distance between the two clusters.

In the present case, each single trial i is represented by the three activities $P50a_i$, $N100a_i$, and $P200a_i$. The similarity between trials is measured in terms of three correlation coefficients, one for each of the three activities. For example, the matrix *P50-corr* contains the cross-correlation coefficients (at lag zero) computed between the $P50a_i$ and $P50a_j$ over a fixed window of 30–80 ms post-stimulus for all $i, j = 1, 2, \dots, N$, where N is the number of trials. Similarly, the cross-correlation matrices *N100-corr* and *P200-corr* are computed within the 80–150 ms and 150–250 ms intervals, respectively. These intervals were selected because they span the typical latency ranges for the EP components.

In order to use the correlation values as a distance measure in the clustering procedure, the correlations are transformed so that highly similar trials with correlation close to one, will have a distance close to zero, and dissimilar trials (small or negative correlations) will have a large distance. Using P50-related activity as an example, this is accomplished by setting all elements of the cross-correlation matrix *P50-corr* with a value below the threshold *P50-corr-th* (see below) to a very large number (2^{10}). Next, the logarithmic transform $y = |\log_2(x)|$ is applied to convert the correlation matrix to the distance-like matrix *P50-dist*, and complete-linkage clustering is performed. The ensemble average is computed for each cluster to determine if a P50 component is present in the 30–80 ms interval. This is done by an algorithm that determines if a peak is present with an amplitude larger than the amplitudes at the edges of the 30–80 ms interval. If the component is present, the cluster is labeled as “positive”, otherwise it is classified as “negative”. This process is repeated for N100 and P200, respectively, using the corresponding correlation matrices (*N100-corr* for N100, and *P200-corr* for P200) and time intervals (80–150 ms for N100, and 150–250 ms for P200). Naturally, in case of N100, the algorithm will be looking for a valley with an amplitude less than seen at the edges of the 80–150 ms interval. By combining the results, each single trial can be assigned a three-bit binary code to indicate the presence or absence of P50, N100, and P200, respectively. For example, ‘101’ indicates a trial for which P50 and P200 are present, but N100 is absent, and ‘100’ indicates that P50 is present only.

The thresholds *P50-corr-th*, *N100-corr-th* and *P200-corr-th* were determined experimentally using actual EP data from multiple subjects (we used 25 healthy volunteers, see Section 3.1). We reasoned that some inter-trial latency variability will occur, and that trials that are shifted by more than one standard deviation should be considered ‘different.’ Therefore, ensemble averages were computed for each subject, and the P50, N100 and P200 latencies were measured. The standard deviation of the individual-specific latencies were determined, and the auto-correlation at a lag corresponding to this deviation was computed for each single trial *i*. For example, in case of P50, the 30 to 80 ms interval of *P50a_i* was used to compute this auto-correlation. Next, the average of all these auto-correlation values was computed across all trials and used as the correlation threshold *P50-corr-th*. This process was repeated for N100 and P200.

The clustering procedure was applied to the pooled data from all 25 subjects.

2.3 Performance assessment

The performance of the proposed clustering approach was assessed using spontaneous EEG and auditory EP data obtained from 25 healthy control subjects (mean age 42 with a range from 23 to 60 years) using pairs of identical, pure tones of 1000 Hz, with 4 ms duration and 1 ms rise and fall times. The tones within a pair were 0.5 s apart, and there was an 8 s interpair interval. Tones were delivered at 90 dB sound pressure level as measured at the ear using a measure-and-hold digital sound meter (Tandy Corp).

EEGs were recorded from seven channels (Fz, Cz, Pz, Fp1, Fp2, T3 and T4) with linked ears as the reference using silver/silver chloride disk electrodes, affixed with electrode paste, but only data from the Cz channel are reported here. Electrode impedances were kept at 5 kOhm or less. All data were inspected for artifacts. Recordings were rejected when EEG activity showed excessive drowsiness during the recording, and individual trials were rejected when activity in any channel exceeded 75 μV or eye-movement artifact was found. A total of 1062 artifact-free trials were obtained and 1 sec segments starting from 500 ms pre-stimulus to 500 ms post-stimulus of the first stimulus (S1), and of the second stimulus (S2) in each pair were extracted as the EP data. Segments of 500 ms to 1500 ms following each S2 were also extracted and considered to reflect spontaneous EEG.

Performance assessment focused on two aspects. First, the quality of the clusters was determined in terms of homogeneity, compactness, and distance to other clusters obtained from the S1 responses. Secondly, the reliability of the method was assessed by subjecting the spontaneous data to the same clustering procedure as the S1 responses. Spontaneous EEG should not contain EP components, so any trial identified as having one or more components reflects a ‘false alarm’. Given the low signal-to-noise ratio and the fact that spontaneous EEG shows essentially the same undulations as evoked activity, it is likely that the correlation-based clustering method will produce a substantial number of false alarms. However, the chance that all three components will be found in the same spontaneous EEG segment should be substantially smaller than for a post-stimulus segment. Similarly, the probability that all three components are absent in a post-stimulus segment should be much smaller than for spontaneous EEG. The performance assessment will provide insight into the actual size of the false alarm probability for the appearance of single components and/or combinations of components.

3 Results

3.1 Fine-tuning the clustering algorithm

First, we validated to what degree the three EP components could be represented by the three frequency bands mentioned in Section 2.1. The S1 responses (Cz channel) of the 25 healthy subjects were decomposed into the 0–4, 2–8, and 8–16 Hz bands, and grand averages for each band were computed together with the wide-band average. These averages are presented in Figure 1. As one can see, the 8–16 Hz band shows maximal activity coinciding with P50, while no such activity is seen in the 0–4 Hz band, but this band shows maximal activity corresponding to P200 (and a fair amount of N100 activity). The 2–8 Hz band shows all three components, but N100 is best represented by this frequency range. These observations support the choice of the frequency bands.

Next we established the correlation thresholds *P50-corr-th*, *N100-corrth* and *P200-corr-th* using the S1 responses from the 25 healthy volunteers. The component-specific activity (i.e., 8–16 Hz for P50, 2–8 Hz for N100 and 0–4 Hz for P200) for each response was obtained and the ensemble averages for each frequency band were computed for each of the 25 subjects. The standard deviations of the P50, N100, and P200 latencies across subjects were determined, and found to be 0.018 s, 0.020 s and 0.029 s, respectively. The autocorrelations at these lags, namely 0.44, 0.50 and 0.55, were used as the correlation thresholds for P50, N100 and P200, respectively.

3.2 Validating the clustering procedure

The cluster procedure with the thresholds as determined above was applied to the 1062 artifact-free single S1 trials, and twenty-four clusters were obtained for the P50 component (averaging 43 trials per cluster, with 15 of the 24 clusters having more than 10 trials); 32 groups for the N100 component (averaging 33 trials per cluster, with 15 clusters containing more than 10 trials) and 52 groups for the P200 component (averaging 20 trials per cluster, and 18 clusters with more than 10 trials). Ensemble averages were computed for each cluster and searched for the presence or absence of each EP component, and the clusters (and their members) were assigned a three-bit code. The number of trials in each category is presented in Table 1. Three hundred twenty four (31%) trials were classified into the ‘111’ category and 55 (5%) trials were classified into the ‘000’ category. The probability of occurrence for P50, N100, and P200 can be computed from the data presented in Table 1. For example, the P50 probability would be the sum of percentage of trials assigned to the 100, 110, 101, and 111 class, divided by 100, and the N100 probability would be obtained by adding the 010, 110, 011, and 111 percentages. The resulting probabilities are presented in Table 2. It appears that N100 has the highest probability of occurrence, and P200 the lowest.

Assessing the validity of the clustering results was done in several ways. First, the quality of the clusters was measured in terms of cluster homogeneity, i.e., how close members of each cluster are to one another, and inter-cluster distance. Next, individual clusters and cluster categories were inspected for EP morphology. Finally, a comparison with spontaneous, i.e., non EP data was performed.

Cluster quality assessment—The homogeneity of each cluster was investigated by correlating its ensemble average with each trial assigned to that cluster. The average and standard deviation of these correlations were computed for each of the P50, N100 and P200 clusters. The average correlations were between 0.8 and 1 and the standard deviations ranged from 0.007 through 0.15, suggesting a fair degree of homogeneity.

The between-class scatter is usually measured by

$$S_B = \frac{1}{k} \sum_{i=1}^k (\vec{m}_i - \vec{m})^2, \quad (2)$$

where k is the number of clusters, $\vec{m}_i$ is the sample mean for cluster i and $\vec{m}$ is the overall mean. In the present case, the ensemble average of cluster i was taken as the sample mean for that cluster and the grand average of the ensemble averages across all P50 (or N100 or P200) clusters was referred to as the overall mean. However, the distance measure $(\vec{m}_i - \vec{m})^2$ in Equation 2 was replaced by the correlation coefficient between the ensemble average of cluster i and the grand average. In contrast to the classic distance-based measure, small values for S_B are desirable. The between-class scatter was found to be 0.056, 0.109 and 0.120 for P50, N100 and P200, respectively. In order to assign statistical significance to these values, the ratio of the within class variance to the total variance (Wilkes Lambda; Wilks, 1932) was computed. The distance transform $dist = 1 - correlation$ was applied to the correlations and the ratios obtained for P50, N100 and P200 were 0.228, 0.199 and 0.428, respectively. The associated F-ratio approximations ($F_{P50}[23, 1049] = 154.43$; $F_{N100}[28, 1030] = 148.07$ and $F_{P200}[43, 1010] = 31.39$) indicated significant differences ($p \leq 0.01$) among the clusters for all three EP components. This implies that trials in one cluster are well distinguished from the trials in the other clusters.

EP morphology—Figure 2 presents (wide-band) ensemble averages for some of the clusters obtained using P50-related activity (P50 clusters), N100-related activity (N100 clusters) and P200-related activity (P200 clusters), respectively. It should be pointed out that, for example, not all P50 clusters show a P50 component. In fact, a substantial variety in morphology is seen, with components missing or with opposite polarity. For example, a (some-what late) P50 can be observed for the first cluster, while the second cluster has a *negative* component at 50 ms and cluster 3 has a (negative-going) zero-crossing at 50 ms. Similarly, the second N100 cluster does not show a N100 component, and P200 is absent in the second and third cluster derived using P200-specific activity.

All trials with the same binary code were averaged and the resulting grand averages for each of the eight categories plus the wide-band average across all trials are presented in Figure 3. As one can see, the '111' category produces an AEP with all three components clearly present, and with larger amplitude than the overall average. Also, the binary code matches the AEP morphology closely. Interestingly, the '000' category is not characterized by a flat line, which would indicate true absence of a response. Rather, an 'anti-phase' EP is seen with a *negative* P50, *positive* N100, and *negative* P200.

Comparison between EP and spontaneous EEG—In a final test, spontaneous EEG segments were clustered. The grand average of the 1062 spontaneous EEG segments are presented in Figure 4, together with the grand average for the S1 responses. It is clear that the ensemble average of the spontaneous EEG segments does not contain an evoked response.

The spontaneous EEG data were clustering using the same procedure (and parameter settings) as used for the EP data. Twenty four groups were obtained for the P50 component (averaging 44 trials per group); 38 groups for the N100 component (averaging 28 trials per group); 49 groups for the P200 component (averaging 22 trials per group). These numbers are comparable to those obtained for the EP data.

Figure 5 presents the average spontaneous activity for each of the eight categories, along with the corresponding EP averages. One can see that the EP averages and the spontaneous EEG averages have similar shapes for each of the eight EP categories. However, it appears that the spontaneous EEG averages are smaller than the EP averages.

The distribution of spontaneous EEG segments over the eight categories is presented in Table 1. The EP data had many more trials containing the typical P50-N100-P200 response (coded as '111') than the spontaneous activity (30.51% vs 8.10%) with much fewer trials in the '000' category (5.18% vs 16.95%). A t-test showed that these two differences reached significance at $p \leq 0.01$ ($t[48] = 5.88$ for the '111' category; $t[48] = -6.89$ for the '000' category). It seems then that the false alarm rate for labeling a trial as complete hovers around 8%. The probability of occurrence for each of the three components was computed, and as is shown in Table 2, the N100 and P200 probabilities are substantially smaller for the spontaneous EEG than for the S1 responses.

In summary, the results presented here indicate that the clustering method achieved its goal of producing homogeneous, well-separated, and meaningful clusters.

3.3 S1 versus S2 comparison

Many studies have shown that the average S2 response is smaller than the S1 response in the paired tone paradigm (see for a review, Bramon et al., 2004). Here, we explore the hypothesis that the reduced average amplitude is due to the fact that S2 produces fewer complete responses than S1. The basic approach is to classify the S2 response using the clusters derived from the S1 responses as the template and then inspect the numbers of 'complete' and 'non-responsive' trials.

The S2 trials were classified by correlating their frequency-specific activity with the ensemble average of each S1-derived cluster. Correlation was computed over the component-specific interval only (i.e., 50–80 ms for P50, 80–150 ms for N100, and 150–250 ms for P200). If this correlation was larger than the correlation threshold, the trial was assigned to the group with the maximum correlation coefficient. Otherwise, a new group was started. No update/recalculation of cluster means took place. Only five trials could not be assigned to an existing cluster (one for P50, two for N100 and two for P200). The grand averages of S2 for each of the eight categories are presented in Figure 6, along with the grand averages of S1 for that category. As one can see, the S1 and S2 averages are comparable, but overall S2 seems to be smaller than S1. The number of S2 trials classified into each EP category is presented in Table 1. As one can see, S2 produced 243 'complete' (111) and 81 'non-responsive' (000) trials (out of 1062), while these numbers were 324 and 55 for the S1 trials, respectively. However, these S1 and S2 percentages did not meet the more rigorous criteria of statistical significance, but the differences were significant for $p < 0.1$ ($t[48] = 1.88$ for 'complete' trials; $t[48] = -1.99$ for 'non-responsive' trials). Again, the probabilities of occurrence for P50, N100 and P200 were computed, and as shown in Table 2, all components had a lower probability of occurrence

for S2 responses than for S1 responses, but were substantially above the spontaneous EEG probabilities (except for P50).

To investigate the amplitude differences, the area under the curve was computed for the trials that had a particular EP component ('positive' trials) and these values were compared between the S1 and S2 responses. For example, the '1xx' category (with 'x' indicating the "don't care") was studied for the P50 trials and 'x1x' for N100. Using P50 as an example, the area under the curve was computed as the sum of the amplitude values from baseline in the 30 ms to 80 ms interval. This was repeated for N100 and P200 using component-specific time ranges. The t-test results, presented in Table 3, indicate that the S2 responses had significantly smaller amplitudes ($p \leq 0.01$) than S1 response for all three EP measures. We also studied the 'complete' category (111). In this case, the peak-to-peak amplitudes were measured for P50, N100 and P200. The t-test results, presented in Table 3, indicate that S1 trials produced significantly ($p \leq 0.01$) larger response than the S2 trials for the N100 and P200 components, but no difference was found for P50.

4 Discussion and Conclusions

A correlation-based unsupervised clustering technique for selective averaging of single-trial EPs has been introduced, using a combination of single trial decomposition and subsequent clustering to group trials with identical characteristics. Similar methods do exist, relying on bandpass filtering (Ruchkin et al., 1988; vonSpreckelsen and Bromm, 1988), principal component analysis (PCA) (Bernat et al., 2005; Dien et al., 2005), wavelet transforms (Bartnik et al., 1992; Thakor et al., 1993; Klein et al., 2006), or independent component analysis (ICA) (Kobayashi et al., 1999; Vigario et al., 2000; Jung et al., 2001; Iyer and Zouridakis, 2007). Features extracted from these components are then used to cluster or classify the trials on the basis of pre- or post-stimulus characteristics (Geva and Pratt 1994; Haig et al., 1995; Zouridakis et al., 1997). The problem with many of these approaches is that the decomposition methods used are often selected primarily on the basis of mathematical convenience (e.g., orthogonality or independence of the components), and that the performance highly depends on the choice of features and similarity index. In contrast, the decomposition method employs a piece-wise functional approximation, which eliminates many of the problems associated with methods requiring orthogonal basis functions. Also, our correlation-based clustering method avoids the need to select features. An additional strength of the method is its ability to deal with inter-subject variations in shape and latency of the EPs. Because correlation is used as measure of similarity, the method is insensitive to amplitude effects, and 'excessive' latency shifts result in the creation of additional clusters.

The capability of this correlation-based clustering technique was demonstrated using EP and spontaneous EEG trials recorded from 25 normal subjects. Using the EP data, we were able to ascertain that homogeneous, well-separated clusters were obtained, which is a hallmark of a well-functioning clustering procedure. Also, our results confirmed previous reports (Jin et al., 1994; Haig et al., 1995; Zouridakis et al., 1997) that there are a number of different kinds of responses among single trials and that not all stimuli produce responses. Specifically, we found that only 30% of the single trials produced 'complete' responses, which is close to Haig et al.'s (1995) finding that 40% of trials had a morphology resembling the averaged EP wave form. The Haig et al. experiment is similar to our study in the sense that both aimed to seek the natural groupings of single-trial responses. However, Haig et al. measured the similarity between single trials by computing the squared error distortion, which ignores the characteristics of the EP components. In contrast, our technique used the correlation coefficient as the similarity measurement, which takes into account EP morphology and component latency. Also, PCA was used in Haig's study to decrease the vector size and the components selected were not

physiologically meaningful. In our study, the components were carefully selected to be reflective of the EP components.

Not all of the post-stimulus activity is directly related to sensory information processing, but contains a (potentially substantial) spontaneous EEG component. It is therefore possible that the resulting clusters are mainly determined by this spontaneous component. This was investigated by clustering *spontaneous* data. We did find that both spontaneous and evoked activity produced similar clusters, but only about 8% of the spontaneous segment were classified as being ‘complete’ responses, suggesting that the false alarm rate is relatively low. Also, the fact that about 30% of the evoked potentials produced complete responses suggests that stimulus processing results more frequently in the occurrence of particular waveforms at specific latencies than in spontaneous EEG. This provides additional support for the phase-resetting hypothesis regarding EP generation (Kolev and Yordanova, 1997; Laskaris and Ioannides, 2000; Winterer et al., 2000; Makeig et al., 2002; Jansen et al., 2003; Klimesch et al., 2004). This theory postulates that the specific afferent associated with a stimulus causes a phase shift in one or more oscillatory EEG components. As a result, the phase distribution obtained from an ensemble of EP responses will transition from a uniform distribution pre-stimulus, to a non-uniform distribution displaying a preferred phase post-stimulus. This is shown for example in Jansen et al., 2003. Parenthetically, even if the phase reorganization model is incorrect, as claimed by some (Kruglikov and Schiff, 2003; Yeung et al., 2004; Makinen et al., 2005), it will be likely that the stimulus-specific activity will result in responses that typically reach a peak or valley at a latency commensurate with the EP components seen in ensemble-averages, thus leading to large inter-trial correlations and meaningful clustering results.

Our results suggest that the sensory gating observed in normal controls may be caused by the fact that the second stimulus generates fewer ‘responsive’ trials which leads to smaller ensemble averages. Also, the probability that a specific component would occur was smaller for S2 responses than S1 responses. However, we also found that the S2 amplitude is significantly smaller than for S1. The neural mechanism responsible for blocking and/or attenuating the response is not known, but several elementary neuronal mechanisms have been proposed (Freedman et al., 1991; Freedman et al., 1996; Grill-Spector et al., 2006). It remains to be seen, however, which mechanism is actually employed by the central nervous system to prevent it from flooding by too much irrelevant information.

Acknowledgments

The research described here was supported in part by grant R01 MH58784 from the National Institute of Mental Health.

References

- Arieli A, Sterkin A, Grinvald A, Aertsen A. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* 1996;273:1868–1871. [PubMed: 8791593]
- Bartnik EA, Blinowska KJ, Durka PJ. Single evoked potential reconstruction by means of wavelet transform. *Biological Cybernetics* 1992;67:175–181. [PubMed: 1627686]
- Başar E, Ugan P. A component analysis and principles derived for the understanding of evoked potentials of the brain: studies in the hippocampus. *Kybernetik* 1973;12:133–140. [PubMed: 4712969]
- Bernat EM, Williams WJ, Gehring WJ. Decomposing ERP time-frequency energy using PCA. *Clin Neurophysiol* 2005;116(6):1314–1334. [PubMed: 15978494]
- Bramon E, Rabe-Hesketh S, Sham P, Murray RM, Frangou S. Meta-analysis of the P300 and P50 waveforms in schizophrenia. *Schizophr Res* 2004;70:315–329. [PubMed: 15329307]
- Brandt ME, Jansen BH, Carbonari JP. Pre-stimulus spectral EEG patterns and the visual evoked response. *Electroencephalogr Clin Neurophysiol* 1991;80:16–20. [PubMed: 1703944]

- Campanella S, Gomez C, Rossion B, Liard L, Debatisse E, Dubois S, Delinte A, Bruyer R, Crommelinck M, Guerit JM. A comparison between group average and individual evoked potential analysis. *Clin Neurophysiol* 1999;29(4):325–338.
- Dien J, Beal DJ, Berg P. Optimizing principal components analysis of event-related potentials: matrix type, factor loading weighting, extraction, and rotations. *Clin Neurophysiol* 2005;116(8):1808–1825. [PubMed: 15996897]
- Duda, RO.; Hart, PE.; Strok, DG. *Pattern Classification*. New York: Wiley-Interscience; 2001. p. 552–555.
- Ford JM, White P, Lim KO, Pfefferbaum A. Schizophrenics have fewer and smaller P300s: a single trial analysis. *Biol Psychiatry* 1994;35:96–103. [PubMed: 8167215]
- Freedman R, Waldo M, Bickford-Winner P, Nagamoto H. Elementary neuronal dysfunction in schizophrenia. *Schizophr Res* 1991;4:233–243. [PubMed: 1645590]
- Freedman R, Adler LE, Myles-Worsley M, Nagamoto H, Miller C, Kisley M, McRae K, Cawthra E, Waldo M. Inhibitory gating of an evoked response to repeated auditory stimuli in schizophrenic and normal subjects. *Arch Gen Psych* 1996;53:1114–1121.
- Garoosi V, Jansen BH. Development and evaluation of the piecewise Prony method for evoked potential analysis. *IEEE Trans Biomed Engng* 2000;47:1549–1554. [PubMed: 11125589]
- Geva AB, Pratt H. Unsupervised clustering of evoked potentials by waveform. *Med Biol Eng Comput* 1994;32(5):543–550. [PubMed: 7845071]
- Grill-Spector K, Henson R, Martin A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 2006;10(1):14–23. [PubMed: 16321563]
- Haig AR, Gordon E, Rogers G, Anderson J. Classification of single-trial ERP sub-types: application of globally optimal vector quantization using simulated annealing. *Electroencephalogr Clin Neurophysiol* 1995;94(4):288–297. [PubMed: 7537201]
- Haig A, Gordon E. Pre-stimulus EEG alpha phase synchronicity influences N100 amplitude and reaction time. *Psychophysiology* 1998;25:316–329.
- Iyer D, Zouridakis G. Single-trial evoked potential estimation: comparison between independent component analysis and wavelet denoising. *Clin Neurophysiol* 2007;118(3):495–504. [PubMed: 17188929]
- Jansen BH, Brandt ME. The effect of the phase of pre-stimulus alpha activity on the averaged visual evoked response. *Electroencephalogr Clin Neurophysiol* 1991;80:241–250. [PubMed: 1713834]
- Jansen BH, Agarwal G, Hegde H, Boutros NN. Phase synchronization of the ongoing EEG and auditory EP generation. *Clin Neurophysiol* 2003;114:79–85. [PubMed: 12495767]
- Jin Y, Potkin SG, Wu J, Dang C, Sandman CA. Single trial P50 in normals and schizophrenia subjects. *Biol Psychiatry* 1994;35:631.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and visualization of single-trial event-related potentials. *Hum Brain Mapp* 2001;14(3):166–185. [PubMed: 11559961]
- Klein A, Sauer T, Jedynek A, Skrandies W. Conventional and wavelet coherence applied to sensory-evoked electrical brain activity. *IEEE Trans Biomed Eng* 2006;53(2):266–272. [PubMed: 16485755]
- Klimesch W, Schack B, Schabus M, Doppelmayr M, Gruber W, Sauseng P. Phase-locked alpha and theta oscillations generate the P1-N1 complex and are related to memory performance. *Brain Res Cogn Brain Res* 2004;19:302–316. [PubMed: 15062867]
- Kobayashi K, James CJ, Nakahori T, Akiyama T, Gotman J. Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. *Clin Neurophysiol* 1999;110(10):1755–1763. [PubMed: 10574290]
- Kolev V, Yordanova J. Analysis of phase-locking is informative for studying event-related EEG activity. *Biol Cybern* 1997;76:229–235. [PubMed: 9151420]
- Kruglikov SY, Schiff SJ. Interplay of electroencephalogram phase and auditory-evoked neural activity. *J Neurosci* 2003;23:10122–10127. [PubMed: 14602828]
- Laskaris NA, Ioannides AA. Exploratory data analysis of evoked response single trials based on minimal spanning tree. *Clin Neurophysiol* 2000;112:698–712. [PubMed: 11275544]
- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ. Dynamic brain sources of visual evoked responses. *Science* 2002;295:690–694. [PubMed: 11809976]

- Makinen VT, Tiitinen H, May PJ. Auditory event-related responses are generated independently of ongoing brain activity. *NeuroImage* 2005;24:961–968. [PubMed: 15670673]
- Rahn E, Başar E. Enhancement of visual evoked potentials by stimulation during low pre-stimulus EEG stages. *Int J Neurosci* 1993;72:123–136. [PubMed: 8225796]
- Rodionov V, Goodman C, Fisher L, Rosenstein GZ, Sohmer H. A new technique for the analysis of background and evoked EEG activity: time and amplitude distributions of the EEG deflections. *Clin Neurophysiol* 2002;113(9):1412–1422. [PubMed: 12169322]
- Ruchkin DS, Johnson R Jr, Mahaffey D, Sutton S. Toward a functional categorization of slow waves. *Psychophysiology* 1988;25(3):339–353. [PubMed: 3406333]
- von Spreckelsen M, Bromm B. Estimation of single-evoked cerebral potentials by means of parametric modeling and Kalman filtering. *IEEE Trans Biomed Eng* 1988;35(9):691–700. [PubMed: 3169821]
- Thakor NV, Xin-Rong G, Yi-Chun S, Hanley DF. Multiresolution wavelet analysis of evoked potentials. *IEEE Trans Biomed Eng* 1993;40:1085–1094. [PubMed: 8307591]
- Turetsky B, Raz J, Fein G. Representation of multi-channel evoked potential data using a dipole component model of intracranial generators: application to the auditory P300. *Electroencephalogr Clin Neurophysiol* 1990;76:540–556. [PubMed: 1701121]
- Vigário R, Särelä J, Jousmäki V, Hämäläinen M, Oja E. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans Biomed Eng* 2000;47(5):589–593. [PubMed: 10851802]
- Wagner P, Roschke J, Grozinger M, Mann K. A replication study on P300 single trial analysis in schizophrenia: confirmation of a reduced number of 'true positive' P300 waves. *J Psychiatric Res* 2000;34:255–259.
- Wilkes SS. Certain generalizations in the analysis of variance. *Biometrika* 1932;24:417–474.
- Winterer G, Ziller M, Dorn H, Frick K, Mulert C, Wuebben Y, Herrmann WM, Coppala R. Schizophrenia: reduced signal-to-noise ratio and impaired phase-locking during information processing. *Clin Neurophysiol* 2000;111:837–849. [PubMed: 10802455]
- Yeung N, Bogacz R, Holroyd CB, Cojen JD. Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. *Psychophysiology* 2004;41:822–832. [PubMed: 15563335]
- Zouridakis G, Jansen BH, Boutros NN. A fuzzy clustering approach to EP estimation. *IEEE Trans Biomed Eng* 1997;44(8):673–680. [PubMed: 9254981]

Figure 1.
Grand averages (thin lines) for 8–16 Hz band (top left), 2–8 Hz (top right) and 0–4 Hz band (bottom), together with the wide-band average (thick line).

Figure 2. Ensemble averages for some of the clusters obtained using P50-specific data (top row), N100 data (middle row), and P200 data (bottom row). Vertical lines indicate the typical latency for each component, and a tick mark indicates the moment of stimulus presentation (at time zero).

Figure 3.
The grand averages of the wide-band EPs for each of the eight EP categories (thick line) and all trials (thin line).

Figure 4.
Ensemble averages of EP data (thick line) and spontaneous data (thin line).

Figure 5.
Grand averages of the S1 (thick line) and spontaneous activity (thin line) for each of the eight categories.

Figure 6.
The grand averages for each of the eight EP categories for the S1 (thick line) and S2 (thin line).

Table 1
The number (and percentage) of trials assigned to each category, for the S1 and S2 responses, and the spontaneous EEG. The categories are binary coded to indicate the presence ('1') or absence ('0') of the P50, N100, or P200 component.

P50	N100	P200	S1	S2	EEG
0	0	0	55 (5%)	81 (8%)	180 (17%)
0	0	1	56 (5%)	78 (7%)	78 (7%)
0	1	0	119 (11%)	122 (11%)	77 (7%)
0	1	1	161 (15%)	153 (14%)	49 (5%)
1	0	0	86 (8%)	91 (9%)	295 (28%)
1	0	1	83 (8%)	90 (9%)	143 (14%)
1	1	0	178 (17%)	204 (19%)	154 (14%)
1	1	1	324 (31%)	243 (23%)	86 (8%)

Table 2

The estimated probability of occurrence for P50, N100, and P200 for the S1 and S2 responses, and the spontaneous EEG.

component	S1	S2	EEG
P50	0.67	0.61	0.67
N100	0.75	0.68	0.33
P200	0.61	0.54	0.33

Table 3
Average S1 and S2 amplitudes and t-test statistics for the amplitude differences between these two responses for various response types using the same coding scheme as in Table 1, with ‘x’ indicating the “don’t care”.

type	comp.	S1-A	S2-A	p	t/df
lxx	P50	10.21	7.73	0.003	2.95/1297
x1x	N100	-54.95	-24.10	5.6E-19	-9.02/1502
xx1	P200	116.67	66.06	1.4E-09	6.10/1186
l11	P50	9.31	9.33	0.97	-0.04/536
l11	N100	-16.18	-12.41	0.0001	-3.89/565
l11	P200	24.48	18.50	3.31E-06	4.70/562