Nonlinear Alignment and Averaging for Estimating the Evoked Potential

Lalit Gupta*, *Member, IEEE,* Dennis L. Molfese, Ravi Tammana, *Member, IEEE,* and Panagiotis G. Simos

Abstract—**This paper addresses the problems associated with averaging brain responses evoked through a repetitive application of an external stimulus. In order to improve the estimate of the evoked potential (EP) through signal averaging, a method which incorporates nonlinear alignment of the EP's into the averaging operation is developed. The method makes no prior assumptions about the properties of the EP or which response in the set best characterizes the EP to be estimated. The nonlinear alignment procedure is designed to pairwise generate optimally aligned EP's by backtracking along the optimal alignment path. The nonlinear alignment and averaging operations are systematically combined to develop methods to estimate the EP. Results from a series of experiments conducted on simulated and real sets of responses show that, through nonlinear alignment and averaging, the events in the EP's are preserved and the estimates of the EP are quite robust.**

I. Introduction

THIS paper focuses on estimating the evoked potential (EP) signal recorded from the human scalp that occurs in response to an externally applied stimulus. The measured response is considered to be an additive superposition of the EP (signal) and the statistically independent background EEG activity (noise) [1]–[6]. Because the stimulus-induced changes in the EEG are small, the most widely used method for estimating the EP is by averaging a set of responses obtained through a repetitive application of the external stimulus. Averaging reduces the effects of noise thus leading to an improvement in the signal-to-noise ratio (SNR) of the measured response (noisy EP). The averaging operation is performed prior to analysis and classification; therefore, the results obtained and the conclusions drawn about the EP are highly dictated by the accuracy of the averaging operation. For example, in EP analysis, properties such as peak amplitudes and peak-to-peak latencies are measured from averaged responses. Additionally, principal component analysis is conducted on averaged responses to identify regions of variability in EP's across different conditions, electrode sites, and subjects. The outcomes of the analysis stage are also used to aid feature selection in EP classification. Averaged responses are also often used to represent typical EP class prototypes or templates in classification problems. Any improvement in the averaging of EP's will, therefore, clearly have a significant impact on improving the analysis and classification of EP signals. EP's are known to be nonstationary and therefore have characteristics that vary from trial to trial. It is well known that the EP's evoked by the same stimulus experience variable latencies in the start points as well as in the segments of the EP's. That is, the EP's experience nonlinear distortions which are typically inconsistent positions of segments and inconsistent spacing between segments from EP to EP. For the averaging operation to be meaningful, the EP's must be aligned or registered in time. If the segments are not aligned, the averaging operation will lead to a blurring (or the loss) of peaks and valleys in the averaged EP waveform. Approaches such as cross correlation-averaging [1], [2] and iterative Fisher scoring [3] have been proposed to estimate the latency shift by detecting a shift in the entire waveform or a shift in a prominent peak in each EP. The estimated latencies are used to register the responses prior to averaging. Essentially, these approaches assume that time alignment may be achieved by synchronizing the start points or the prominent peaks of the EP's. Nonlinear time variations such as expansion and compression in the segments are not taken into account and, therefore, present serious limitations in the validity of applying the averaging operation. Several researchers have identified and proposed more complex solutions to remedy the problems associated with using the latency of the entire waveform or a single prominent peak as the basis for alignment. Examples include latency correction averaging [4], continuous latency correction averaging and enhanced averaging [5]. The latency correction averaging technique involves peak identification, grouping of peaks, and aligning short segments in the vicinity of the peaks. The components of the latency correction averaging are generally disjoint and the continuous latency correction averaging is designed to convert the disjoint segments into a smooth curve using a weighted least-squares fit. The enhanced averaging method uses a deconvolution restoration approach to estimate components of the EP. A statistical pattern classification method has also been used to select responses with discriminable event-related signals in a defined time interval and the averages computed from the selected responses show enhancement of EP peaks in the chosen interval [6]. An approach based on optimally aligning and averaging responses without having to specify the events of the EP or making any assumptions about the features of the EP is developed in this paper. Additionally, it is assumed

Manuscript received December 17, 1993; revised October 25, 1995. *Asterisk indicates corresponding author.*

*L. Gupta is with the Department of Electrical Engineering, Southern Illinois University at Carbondale, Carbondale, IL 62901 USA (e-mail gupta@zeus.c-engr2.siu.edu).

D. L. Molfese and P. G. Simos are with the Department of Psychology, Southern Illinois University at Carbondale, Carbondale, IL 62901 USA.

R. Tammana is with the Department of Electrical Engineering, Southern Illinois University at Carbondale, Carbondale, IL 62901 USA.

Publisher Item Identifier S 0018-9294(96)02429-9.

that there is no knowledge of which response in the set best characterizes the EP to be estimated.

II. Nonlinear Alignment

Nonlinear alignment [also called dynamic time warping (DTW)] techniques were initially proposed to solve the time-registration problem between a test pattern and a reference pattern in speech recognition problems as it was found that a simple linear expansion or compression of the time axis was inadequate to compensate for the timing differences between the test and reference patterns [7]–[9]. Subsequently, DTW has been applied to other problems such as object recognition [10], unsupervised classification of EEG waveform segments [11], and classification of somatosensory EP's [12]. DTW is typically used to obtain a measure of dissimilarity remaining between two patterns after the patterns are optimally aligned. In the object recognition problems, nonlinear alignment is used to compensate for the variations in the boundary of an object that result from the segmentation of objects in different environmental conditions [10], [13]. Additionally, nonlinear alignment has also been used to align the outputs of neural net classifiers in partial object classification problems [14], [15]. In [11], it is demonstrated that minor changes in EEG's such as the occurrences of sharp waves and spikes can be detected by incorporating the DTW dissimilarity measure into a hierarchical clustering algorithm. No assumption of a reference pattern is made as the problem addressed focused on clustering EEG segments. In [12], a standard EP is generated by warping each EP with every other EP in the set. By defining an appropriate cost function, the EP with the least mean cost and variance across the set was selected as the EP that best represented the EP set. All EP's in the set were warped against the selected EP and averaged pointwise to produce the standard EP. The standard EP was then used as a reference in DTW for detecting abnormalities in patients with multiple sclerosis.

In this paper, DTW is used in an unconventional way as the goal is not to find the DTW distance between two EP's but to align an entire set of EP's in a collection of measured responses so that there is no loss of information in the averaging operation. What makes this a complex problem is that no assumptions are made about the EP to be estimated. Neither is an assumption made about which response in the collection best characterizes the EP to be estimated as no response is truly noise-free, and it is not known if the selected response is free of latency variations. Therefore, a reference EP which could be used to align the EP's in the set does not exist. These assumptions are made realistically as the goal of this paper is to estimate an unknown signal; therefore, if possible, no assumptions about the signal should be made. Clearly, the complexity of the problem can be reduced by making prior assumptions about the signal.

Several versions of nonlinear alignment algorithms have been developed and can be grouped into two classes: symmetric and asymmetric. In the symmetric form, the time axes n and m of two patterns are transformed into an intermediate axis k, whereas, in the asymmetric form, the time axis of one pattern is warped onto the time axis of the other pattern. In the asymmetric algorithms, the samples of the reference pattern are optimally repeated or deleted to obtain the best alignment between a test and the reference pattern. The symmetric algorithms can be designed to ensure that no skipping of samples in any of the two patterns is allowed. The symmetric approach is preferred in the EP estimation problem as no knowledge of the types of nonlinear variations in the EP's is assumed; consequently, it is safer not to skip samples. For example, if one of the EP's has no latency variations and the other experiences compression in a segment, it would be much better to align the EP's by expanding (repeating samples) the compressed segment rather than aligning the EP's by compressing a segment (skipping samples) in the EP that does not contain any latency variations. Another reason for selecting a symmetric alignment algorithm is that the EP estimation problem, with the assumptions made, precludes the selection of a reference. If any response is selected to be the reference and the selected reference is a poor representation of the EP signal, then, averaging the aligned EP's after they are aligned with the reference using an asymmetric alignment algorithm will give a poor estimate of the EP signal.

In order to develop the method for EP estimation through nonlinear alignment and averaging, the general formulation of a symmetric nonlinear alignment algorithm and how it is used to generate a pair of aligned responses is described first. Consider two responses represented by sequences $r_x(n)$, $n = 1, 2, \cdots, N$, and $r_y(m)$, $m = 1, 2, \cdots, M$. To make the formulation very general, the durations of the two sequences are assumed to be unequal. The goal is to align the samples of the two sequences so that the latency discrepancies in the two sequences are minimized. Clearly, linear alignment by uniformly stretching or compressing the responses would not achieve the goal if the EP's contain nonlinear variations. Aligning the EP's by finding the position of best match by cross correlating the two sequences would also not necessarily align the segments in both EP's. In order to optimally align the samples, an alignment function of the form

$$W = w(1),\ w(2),\ \cdots,\ w(K),\ \text{with } w(k) = [i(k),\ j(k)]$$

must be determined so that the overall discrepancy D_{xy} between the two sequences is minimized (or equivalently, the overall similarity is maximized). W provides a mapping between the indexes of $r_x(\cdot)$ and $r_y(\cdot)$ via an intermediate axis k of length K. If, for example

$$W = [1,\ 1],\ [2,\ 2],\ [2,\ 3],\ [3,\ 4],\ \cdots,\ [N,\ M]$$

then sample $r_x(1)$ is aligned with $r_y(1), r_x(2)$ is aligned with $r_y(2), r_x(2)$ is aligned with $r_y(3), r_x(3)$ is aligned with $r_y(4)$, $\cdots$, $r_x(N)$ is aligned with $r_y(M)$. For each $w(k)$, a cost $d[w(k)]$ is assigned to reflect the discrepancy between the aligned samples. Examples of cost functions that could be used are the absolute difference, the square of the difference, and the Euclidean distance between the samples. The absolute difference is used in the experiments described in this paper.

The alignment function is, therefore, determined such that the overall cost (measure of discrepancy)

$$D_{xy} = \sum_{k=1}^{K} d[w(k)]$$

is minimized subject to the following constraints

1) Monotonicity: The alignment function must be monotonic to preserve the natural time ordering in the sequences. That is

$$i(k) \geq i(k-1)$$
$$j(k) \geq j(k-1).$$

For example, if $r_x(3)$ is aligned with $r_y(4)$, then $r_x(4)$ cannot be aligned with $r_y(3)$.

2) End-Point Alignment: The end-points (first and last samples) of the sequences must be aligned. That is

$$i(1) = j(1) = 1$$
$$i(K) = N$$
$$j(K) = M.$$

3) Continuity: The alignment function must not skip any samples in the two sequences, therefore

$$i(k) - i(k-1) \leq 1$$
$$j(k) - j(k-1) \leq 1.$$

Using a dynamic programming approach [7]–[9], the solution to the above optimization problem is given by solving the recursive equation

$$D[w(k)] = d[w(k)] + \min_{w(k-1)}\{D[w(k-1)]\}$$

with initial condition

$$D[w(1)] = d[w(1)].$$

The overall discrepancy between the two responses after alignment is given by

$$D_{xy} = (1/K)D[w(K)].$$

From the monotonicity and continuity constraints imposed on the alignment function, if $w(k) = [i,\ j]$, then $w(k-1)$ consists of $[i-1,\ j]$, $[i,\ j-1]$ and $[i-1,\ j-1]$ and the recursive equation becomes

$$D[i,j] = d[i,j] + \min\{D[i-1,j], D[i,j-1]\ D[i-1,j-1]\}.$$

Generally, $D[i,\ j]$ is computed only for $[i,\ j]$ values in a band along the diagonal of the $N \times M$ array because the optimal alignment path for sequences with similar features tends to fluctuate in the neighborhood of the diagonal. This not only restricts the region of search for the optimal alignment path in a meaningful manner but also reduces the number of computations required.

For typical pattern recognition problems, the discrepancy is computed between a test pattern and all the reference patterns. The test pattern is then assigned to the class of the pattern that yields the minimum discrepancy (minimum mismatch rule). The primary goal in this paper is not to determine the discrepancy between the responses, but to generate an estimate of the EP through optimal alignment and averaging of the responses. The aligned versions of the two responses can be found if the optimal alignment path is known. If $D[i,\ j]$ is stored in an $N \times M$ array, the optimal alignment path W is found by starting at $D[w(K)] = D[N,\ M]$ and backtracking along the minimum $D[w(k-1)]$ until $D[w(1)] = D[1,\ 1]$ is reached. Once the optimal alignment path is determined through backtracking, two new aligned sequences $R_x[i(k)]$ and $R_y[j(k)]$, $k = 1,\ 2, \cdots,\ K$ can be formed. The samples of the sequences $R_x[i(k)]$ and $R_y[j(k)]$ have been optimally aligned and from the monotonicity and continuity constraints imposed, the optimal alignment is achieved by repeating samples from the sequences in an optimal manner. Due to the repetition of the samples, the duration of the intermediate axis k is $K \geq \max\ [N,\ M]$. It should also be noted that the two sequences being aligned are assumed to have the same scale factor because optimal alignment is determined by comparing the amplitudes of the samples of the sequences. If the sequences do not have the same scale factor, they must be scale normalized prior to alignment [10].

III. Pairwise Signal Averaging

As stated in the introduction, the most frequently used model to describe the relationship between the measured response $r(t)$ to an externally applied stimulus is the summation of the EP and the background EEG activity. That is

$$r(t) = s(t) + \varepsilon(t)$$

where $s(t)$ is the EP and $\varepsilon(t)$ is the ongoing EEG signal. The goal is to estimate $s(t)$ from a set of L responses $r_i(t)$, $i = 1,\ 2, \cdots,\ L$. The on-going EEG signal is activity not caused by the external stimulus and is, therefore, regarded as background noise. In this model, it is assumed that $s(t)$ and $\varepsilon(t)$ are independent, $s(t)$ is deterministic, and $\varepsilon(t)$ has zero mean. In practice, the signals are sampled and if the discrete signal formed by averaging L sampled responses $r_i(k)$, is

$$\bar{r}_L(k) = (1/L)\sum_{i=1}^{L} r_i(k), \quad k = 1, 2, \cdots,\ K$$

and if it is assumed that noise samples $\varepsilon_i(k)$ are uncorrelated, then it follows that

$$E[\bar{r}_L(k)] = s(k) \quad \text{and} \quad \sigma_r^2(k) = [1/L]\sigma_\varepsilon^2(k).$$

$E[\bar{r}_L(k)]$ and $\sigma_r^2(k)$ are the expected value and variance of sample k in $\bar{r}_L(\cdot)$, respectively, and $\sigma_\varepsilon^2(k)$ is the variance of the kth sample of $\varepsilon(\cdot)$. That is, the variance of each sample decreases by a factor L when L responses are averaged and the variability decreases as the number of responses used in the averaging process increases.

Given that the nonlinear alignment procedure generates two new sequences that are aligned, methods for obtaining an estimate of the EP from the response set can be developed by first assuming that L is an integer power of 2 and then noting that $\bar{r}_L(k)$ can be computed as

$$\bar{r}_L(k) = (1/2)[\bar{r}_{1,(L/2)}(k) + \bar{r}_{(L/2)+1,L}(k)]$$

where

$$\bar{r}_{1,(L/2)}(k) = [1/(L/2)] \sum_{i=1}^{L/2} r_i(k)$$

and

$$\bar{r}_{(L/2)+1,L}(k) = [1/(L/2)] \sum_{i=(L/2)+1}^{L} r_i(k).$$

That is, $\bar{r}_{1,\,(L/2)}(k)$ is the mean of the first half of the responses $r_1(k),\ r_2(k),\ \cdots,\ r_{(L/2)}$, and $\bar{r}_{(L/2)+1,\,L}(k)$ is the mean of the second half of the responses $r_{(L/2)+1}(k), r_{(L/2)+2}(k),\ \cdots,\ r_L$. By further decomposing the first half and second half of the responses into two equal sized sets, $\bar{r}_{1,\,(L/2)}\ (k)$ and $\bar{r}_{(L/2)+1,\,L}\ (k)$ can be computed as

$$\bar{r}_{1,(L/2)}(k) = (1/2)[\bar{r}_{1,(L/4)}(k) + \bar{r}_{(L/4)+1,(L/2)}(k)]$$
$$\bar{r}_{(L/2)+1,L}(k) = (1/2)[\bar{r}_{(L/2)+1,(3L/4)}(k) + \bar{r}_{(3L/4)+1,L}(k)].$$

The L responses can be decomposed into successively smaller sets until pairs of responses remain. The means of the pairs are computed and combined pairwise according to the steps outlined above. This method for computing the average $\bar{r}_L(k)$ is referred to as Method 1.

For the more general case when L is not an integer power of 2, the equation for computing $\bar{r}_L(k)$ can be written as

$$\bar{r}_L(k) = (1/L)[r_L(k)] + [(L-1)/L][\bar{r}_{L-1}(k)]$$

where

$$\bar{r}_{L-1}(k) = [1/(L-1)] \sum_{i=1}^{L-1} r_i(k).$$

That is, the average $\bar{r}_L\ (k)$ can be computed sequentially by first computing $\bar{r}_2\ (k)$ from the pair $r_1\ (k)$ and $r_2\ (k)$, $\bar{r}_3\ (k)$ from the pair $\bar{r}_2\ (k)$ and $r_3\ (k)$, $\ldots$, and, finally, $\bar{r}_L\ (k)$ from the pair $\bar{r}_{L-1}(k)$ and $r_L\ (k)$. This method for computing the average $\bar{r}_L\ (k)$ is referred to as Method 2.

IV. Nonlinear Alignment and Averaging Filters

Two filters to estimate the EP $s(n)$ from a set of L responses based on combining nonlinear alignment and the two pairwise sample mean computational methods are developed below. The resulting filters are referred to as nonlinear alignment and averaging filters (NLAAF's).

A. NLAAF1

This filter uses Method 1 to compute the average and it is assumed that L is an integer power of two. The L responses are first decomposed into $(L/2)$ pairs of responses $[r_1(n),\ r_2\ (n)],\ [r_3\ (n),\ r_4\ (n)],\ \cdots,\ [r_{L-1}\ (n),$ and $r_L(n)]$. In order to estimate $s(n)$, the pair $r_1(n)$ and $r_2(n)$ are aligned first using the above nonlinear alignment algorithm to generate two aligned responses $R_1(k)$ and $R_2(k)$. Due to the symmetry in the alignment algorithm, aligning $r_1(n)$ with $r_2(n)$ gives exactly the same result as aligning $r_2(n)$ with $r_1(n)$. The average $\bar{r}_{1,\,2}(k)$ of $R_1(k)$ and $R_2(k)$ is the average of the aligned versions of the pair of responses $r_1(n)$ and $r_2(n)$. Similarly, the averages $\bar{r}_{3,\,4}(k), \bar{r}_{5,\,6}(k),\ \cdots,\ \bar{r}_{L-1,\,L}(k)$ can be computed for the remaining pairs of aligned responses $[R_3(k), R_4(k)],\ [R_5(k), R_6(k)],\ \cdots,\ [R_{L-1}(k), R_L(k)]$, respectively. The $(L/4)$ pairs $[\bar{r}_{1,\,2}(k),\ \bar{r}_{3,\,4}(k)],\ [\bar{r}_{5,\,6}(k),\ \bar{r}_{7,\,8}(k)],\ \cdots,\ [\bar{r}_{L-3,L-2}(k),\ \bar{r}_{L-1,\,L}(k)]$ are aligned next to generate the aligned pairs $[\bar{R}_{1,\,2}(k),\ \bar{R}_{3,\,4}(k)],\ [\bar{R}_{5,\,6}(k),\ \bar{R}_{7,\,8}(k)],\ \cdots,\ [\bar{R}_{L-3,L-2}(k),\ \bar{R}_{L-1,\,L}(k)]$. The averages $\bar{r}_{1,4}(k),\ \bar{r}_{5,\,8}(k),\ \cdots,\ \bar{r}_{L-3,\,L}(k)$ can be computed from the $(L/4)$ aligned pairs. This systematic step-by-step pairwise aligning and averaging is continued until $\bar{r}_L(k)$ which is the estimate of $s(n)$ is finally computed as the average of the pair $[\bar{R}_{1,\,L/2}(k),\ \bar{R}_{(L/2+1),\,L}(k)]$.

B. NLAAF2

This filter uses Method 2 to compute the average and is more general than NLAAF1 because it places no restrictions on L, i.e., L need not be an integer power of two and L may be even or odd. In order to estimate $s(n)$, the responses $r_1(n)$ and $r_2(n)$ are aligned first using the nonlinear alignment algorithm to generate two aligned responses $R_1(k)$ and $R_2(k)$. The average $\bar{r}_2(k)$ of $R_1(k)$ and $R_2(k)$ is an estimate of $s(n)$ using the pair $r_1(n)$ and $r_2(n)$. The pair $\bar{r}_2(k)$ and $r_3(n)$ are aligned next to generate the aligned sequences $\bar{R}_2(k)$ and $R_3(k)$. The average $\bar{r}_3(k)$ of the pair $\bar{R}_2(k)$ and $R_3(k)$ using the sequential pairwise averaging rule of Method 2 is the estimate of $s(n)$ using the aligned versions of $r_1(n),\ r_2(n)$, and $r_3(n)$. This systematic step by step aligning-pairwise averaging procedure is continued until all L responses are used to estimate $s(n)$, that is, $\bar{r}_L(k)$ is the final estimate of the signal $s(n)$.

Although Method 2 yields an unbiased estimate of the average, experiments using NLAAF2 have shown that the alignment of a single signal with the average of the previous signals influences the alignment path significantly. For example, in the last step, the Lth response is aligned with the average of the previous $(L-1)$ responses, therefore, the Lth response has the same influence on the alignment path as the average of the other $(L-1)$ responses. In order to reduce the effects of this problem especially when L is large, the filter could be modified to use a combination of NLAAF1 and NLAAF2 when L is not an integer power of two. NLAAF2 is used in the first stage, and NLAAF1 is used in the subsequent stages. As an example, if $L = 96$, groups of three sequences can be averaged using NLAAF2 and the 32 averages can be combined using NLAAF1 to compute the final average. If $L = 100$, four responses could be discarded from the

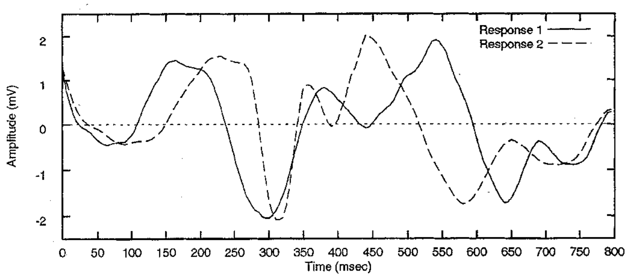


Fig. 1. Experiment 1: Responses prior to alignment.

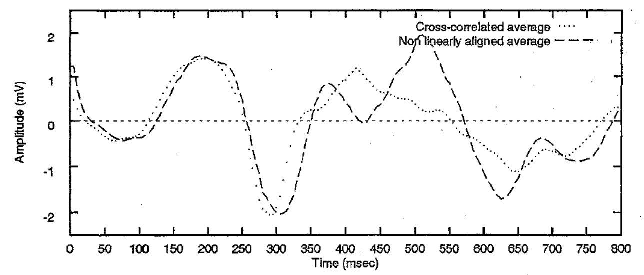


Fig. 3. Experiment 1: Nonlinearly aligned and cross-correlated averages.

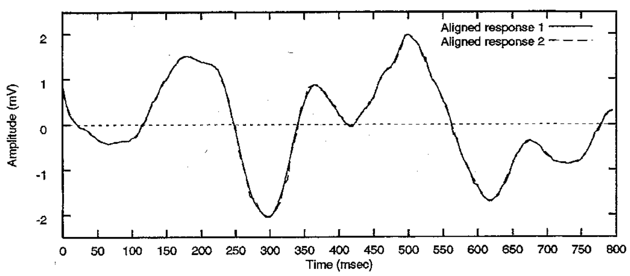


Fig. 2. Experiment 1: Responses in Fig. 1 after nonlinear alignment.

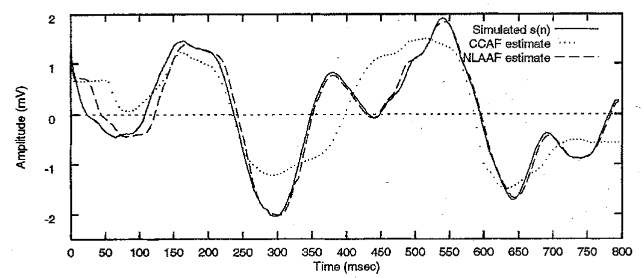


Fig. 4. Experiment 2: Only nonlinear distortion (no noise).

computations. The grouping of the responses in the first stage must be done such that the number of averages going into the second stage is an integer power of two.

The symmetric nonlinear alignment algorithm optimally aligns the responses without skipping samples in order to preserve the events in the responses being aligned. Additionally, the step-by-step procedures ensure that no events are lost in the alignment-averaging operations over all the responses. The duration of the aligned responses is generally larger than the maximum of the durations of the two responses prior to alignment; therefore, the duration of the estimate has a tendency to grow with each additional alignment step. The final estimate $\bar{r}_L(k)$ could be duration normalized, however, in order to avoid the additional computations associated with an increase in the durations, the estimate at the end of each alignment-averaging stage is normalized to have a duration equal to the original responses.

V. Experiments and Results

In order to demonstrate the effectiveness of the NLAAF's in estimating the EP signal, a series of experiments were conducted using real measured responses. To aid visual analysis, the sampled responses are displayed as continuous signals in the figures that follow. In all experiments, the number of signals L was selected to be integer power of two so that the NLAAF1 could be used.

A. Experiment 1

The first experiment was designed to demonstrate the generation of optimally aligned sequences using the nonlinear alignment procedure. Fig. 1 shows two responses which have similar features, however, due to a 64% compression (simulated) between 190 ms and 540 ms in response 2, the features in the two responses are not registered in time. Fig. 2 shows the two new responses formed by applying the nonlinear alignment algorithm. The responses were also aligned through cross correlation and Fig. 3 shows the averages of the cross-correlated and the nonlinearly aligned responses. Fig. 2 clearly shows how well the features of the two responses are aligned in time. Additionally, Fig. 3 shows that averaging after nonlinear alignment preserves the features whereas averaging after cross-correlation (or directly) tends to blur the out-of-phase features in the responses.

B. Experiment 2

The next set of experiments were designed to evaluate the performance of the NLAAF in estimating an EP signal. Noise and nonlinear variations such as expansions in segments, compressions in segments, and inconsistent timing differences in the segments of the EP's were modeled to simulate a set of test responses (distorted EP's) from a conventionally time-locked averaged response. The responses used to form the average were collected from the F3 scalp location of an adult subject. The signals were recorded in response to a series of three successive picture stimuli, while the subject was engaged in a matching task. The averaged response represented the simulated EP $s(n)$ to be estimated and the distorted EP test set represented the measured responses $r_i(n)$. By using this approach, the estimate $\bar{r}_L(k)$ generated from $r_i(n)$ could be compared with the simulated $s(n)$ visually as well as by using an appropriate measure such as the discrepancy D_{sr} (degree of mismatch) to determine the robustness of the estimate. Each signal in the test set was simulated to experience either expansion or compression in a segment of the signal with equal probability. The location of the segment

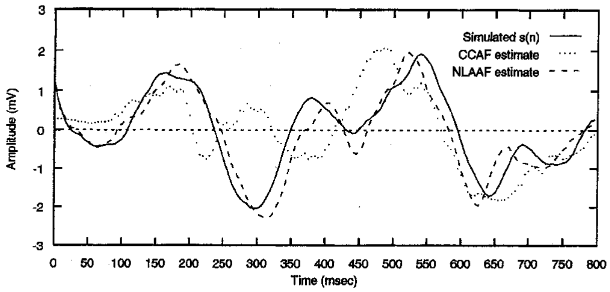


Fig. 5. Experiment 2: Nonlinear distortion and noise with SNR = 1.

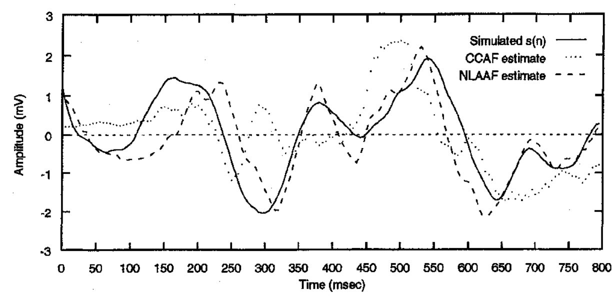


Fig. 6. Experiment 2: Nonlinear distortion and noise with SNR = 0.5.

was to the left or right of the midpoint (400 ms) of the waveform with equal probability. The size of the segment was 300 ms and the percentage of expansion/compression varied randomly between 25% and 45%. To simulate noise, noise sequences generated by a 10th-order autoregressive (AR) model were added to the nonlinearly distorted $s(n)$. The AR model parameters were estimated using prestimulus EEG segments sampled at 200 Hz for a period of 1 s. The EEG data were collected from the F3 scalp location of the same subject. The AR model was driven by a white noise sequence with zero mean and a variance selected to simulate responses having a specified SNR. The SNR as used in these experiments is the ratio of the signal variance to the noise variance. A 10th order model was chosen as it has been shown to provide a realistic model for EEG data [16], [17]. The NLAAF was applied to test sets of various sizes and examples of typical estimation results for test sets containing 64 simulated responses are shown in Figs. 4–7. For comparison, the signals estimated by the Woody filter [1], i.e., the cross-correlation averaging filter (CCAF) are also shown in the figures. The discrepancies between the simulated EP and the estimates are summarized in Table I. The superiority of the NLAAF estimate is clearly evident visually as well as from the discrepancy values. Additional results are presented in Table II. For each SNR, the experiment was repeated 100 times and the average discrepancies between the simulated EP and the estimates are listed in the table. The ordering of the responses in the test set were shuffled randomly and the experiments were repeated to determine if the performance of the NLAAF was dependent on the ordering of the responses used in the step-by-step alignment-averaging operations. Only minor differences in the discrepancy values were noted and these differences were regarded as insignificant.

C. Experiment 3

The last experiment was designed to estimate the match and mismatch EP's from a set of actual (not simulated) match and mismatch responses to an external visual stimulus. Discriminating between match and mismatch conditions is a problem of considerable interest in EP research and numerous experiments have employed measures of scalp-recorded brain activity to study this issue [18]–[27]. In such binary classification tasks, subjects must partition sequentially presented external test stimuli into two categories: match (same) or mismatch (different). A match condition occurs when the subject decides that the second test stimulus was the same as the previously presented stimulus and a mismatch condition occurs when the subject decides the second test stimulus was different from the previously presented stimulus.

Electrophysiological activity time-locked to the onset of each visual stimulus served as the raw measured response in this experiment. The responses were collected from a 29-year-old adult subject. The experimental task involved presenting a series of visual stimuli on a micro-computer screen. On each trial, a printed word was presented first followed by the picture of an object. Upon presentation of a picture, the subject had to decide whether it matched in meaning the printed word that was seen. A two-choice ("match" or "mismatch") response was required by the subject.

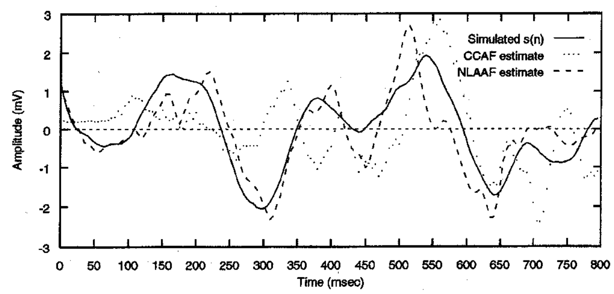


Fig. 7. Experiment 2: Nonlinear distortion and noise with SNR = 0.1.

TABLE I
EXPERIMENT 2: DISCREPANCY MEASURES FOR THE ESTIMATES SHOWN IN FIGS. 4–7

	Fig. 4	Fig. 5	Fig. 6	Fig. 7
NLAFF	0.034	0.091	0.132	0.144
CCAF	0.161	0.171	0.191	0.234

TABLE II
EXPERIMENT 2: DISCREPANCY MEASURES AVERAGED OVER 100 REPETITIONS

	SNR				
	1	0.7	0.5	0.3	0.1
NLAFF	0.098	0.105	0.113	0.134	0.148
CCFF	0.166	0.174	0.186	0.197	0.237

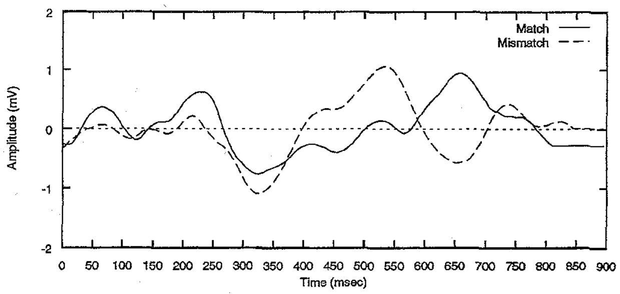


Fig. 8. Experiment 3: CCAF match and mismatch estimates at F3.

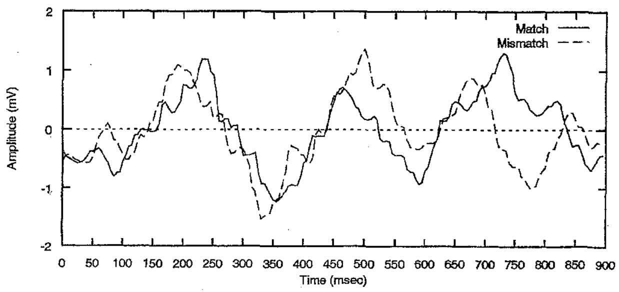


Fig. 9. Experiment 3: NLAAF match and mismatch estimates at F3.

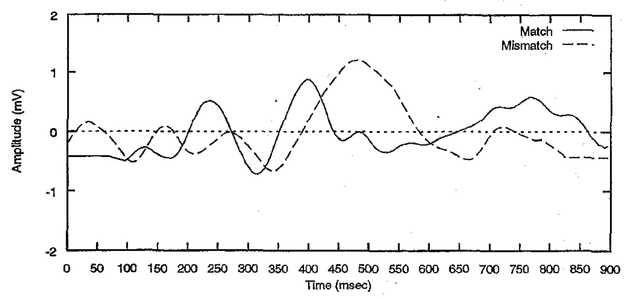


Fig. 10. Experiment 3: CCAF match and mismatch estimates at P3.

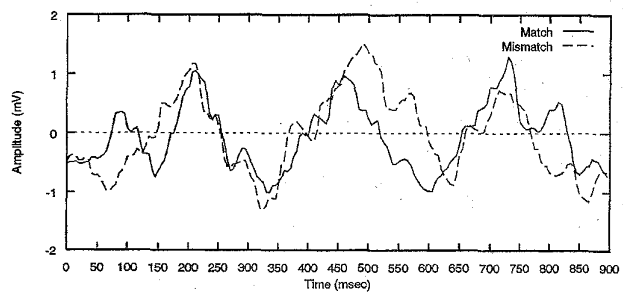


Fig. 11. Experiment 3: NLAAF match and mismatch estimates at P3.

The interstimulus interval was varied randomly from 5 to 6 s across stimuli. The trials were self initiated by the subject. The response was recorded from six scalp locations (F3, Fz, F4, P3, Pz, and P4) according to the 10/20 System [28] using a standard electrode cap. Electrooculogram (EOG) was recorded from a right canthal and a right suborbital electrode (bipolar montage). All scalp-electrode inputs were referred to linked earlobes. Scalp-electrode impedances were kept below 5 KOhm while ear-electrode impedances were within 1 KOhm with each other (both pre- and post-test). The response was input to a set of differential amplifiers (Grass Model 12) with gain set at 20 000. Filters were set at 0.1 and 30 Hz corresponding to 50% low- and high-frequency cutoff points, respectively. The 60-Hz notch filter was also engaged. The signal was digitized online by sampling every 5 ms starting 100 ms before stimulus onset and continuing for 900 ms. Thus, each response consisted of 180 samples. After testing, individual responses were manually rejected when either an incorrect key-press response had been given, or the signal was contaminated by a physiological artifact—defined as a peak-to-peak amplitude greater than 70 μV in the eye channel or 100 μV in any one scalp-electrode channel. Using these criteria, a total of 32 responses to "match" and 32 responses to "mismatch" events were recorded at each scalp location.

Application of the NLAAF and CCAF to the 32 match and 32 mismatch responses gave the match and mismatch signal estimates shown in Figs. 8–11. For brevity, only the results for left-hemisphere locations F3 and P3 are presented as these results were typical of the results obtained from all six scalp electrode sites. As expected, the NLAAF minimized the blurring in the estimated match and mismatch signals. This is evident when the results of NLAAF are compared with the results of the CCAF. The peak-to-peak voltages in the nonlinearly aligned and averaged waveforms are larger and are also more well defined for both the match and mismatch conditions. Studies have shown that the most consistent effect of match-mismatch conditions is an increased negativity in the EP region between 240 and 400 ms and an increased positivity between 400 and 600 ms following mismatch events [18], [19], [21], [24], [25], [27]. This effect is consistent and enhanced in the match and mismatch estimates obtained using NLAAF. In order to further evaluate the performance of the filters, the average discrepancy between the match estimate and the 32 match responses was determined at each location. Similarly, the average discrepancy between the mismatch estimate and the 32 mismatch responses was determined for each location. These results are summarized in Table III. The average discrepancy is a measure of the overall dissimilarity between the estimate and the responses used in computing the estimate. The dissimilarity decreases with a decrease in the average discrepancy. A smaller average discrepancy indicates that the estimate is a better representation (hence, more robust) of the signals. The smaller average discrepancies obtained by the NLAAF for the match and mismatch estimates at all six locations indicate that the NLAAF estimates are more robust than the CCAF estimates. The above observations in conjunction with the simulation results of the previous sets of experiments lead to the conclusion that the signal estimates obtained from NLAAF are quite robust and the comparisons with the CCAF estimates emphasize this robustness.

VI. Conclusion

This paper focused on solving the alignment problems associated with signal averaging in estimating the EP from a set of responses elicited by an externally presented stimulus. As the problem addressed estimating an unknown signal, no

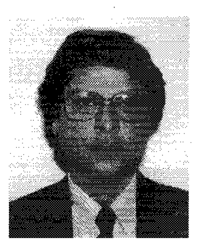
TABLE III
EXPERIMENT 3: AVERAGE DISCREPANCY MEASURES

	CCAF		NLAAF	
	Match	Mismatch	Match	Mismatch
F3	0.499	0.461	0.414	0.379
Fz	0.480	0.469	0.381	0.367
F4	0.484	0.427	0.407	0.346
P3	0.422	0.410	0.356	0.319
Pz	0.419	0.428	0.364	0.345
P4	0.398	0.409	0.356	0.313

assumptions were made about the signal to be estimated. A nonlinear alignment algorithm was used to optimally align the events in the responses without the need for a reference in the alignment procedure. The detection of features such as peaks and valleys was not necessary, neither was it necessary to select a best EP representative from the set of responses. Two pairwise averaging methods were formulated and combined with nonlinear alignment to develop the nonlinear alignment and averaging filters for EP estimation. The undesirable effect of aligning a set of aligned responses with a single response in NLAAF2 was identified and a method to minimize this effect was suggested. For the EP estimation problem, NLAAF1 can always be applied as the number of EP's collected can be controlled, therefore, the restriction on L is of little concern. The series of experiments conducted clearly demonstrate the robustness of the NLAAF in estimating a signal in the presence of nonlinear time variations and noise. Due to the accuracy of the NLAAF estimates, improved measurements are possible, therefore, significant improvements can be expected in the analysis and classification of EP's based on the NLAAF EP estimates.

REFERENCES

[1] C. D. Woody, "Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals," *Med. Biol. Eng.*, vol. 5, pp. 539–553, 1967.
[2] D. G. Wastell, "Statistical detection of individual evoked responses: An evaluation of Woody's adaptive filter," *Electroencephalog. Clin. Neurophysiol.*, vol. 42, pp. 835–839, 1977.
[3] J. Mochs, W. Kohler, T. Gasser, and D. T. Pham, "Novel approaches to the problem of latency jitter," *Psychophysiol.*, vol. 25, no. 2, pp. 217–226, 1988.
[4] J. I. Aunon, C. D. McGillem, and D. G. Childers, "Signal processing in evoked potential research: Averaging and modeling," *CRC Crit. Rev. Bioeng.*, pp. 323–367, 1981.
[5] C. D. McGillem, J. I. Aunon, and C. A. Pomalaza, "Improved waveform estimation procedures for event-related potentials," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 6, pp. 371–379, June 1985.
[6] A. S. Gevins, N. H. Morgan, S. L. Bressler, J. C. Doyle, and B. A. Cutillo, "Improved event-related potential estimation using statistical pattern classification," *Electroencephalog. Clin. Neurophysiol.*, vol. 64, pp. 177–186, 1986.
[7] H. Sakoe and S. Chiba, "Dynamic programming optimization for spoken word recognition," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-26, pp. 43–49, 1978.
[8] L. R. Rabiner and S. E. Levinson, "Isolated and connected word recognition, theory and selected applications, *IEEE Trans. Commun.*, vol. COM-29, pp. 621–659, 1981.
[9] H. F. Silverman and D. P. Morgan, "The application of dynamic programming to connected speech recognition," *IEEE Acoust. Speech Signal Processing Mag.*, pp. 7–41, July 1990.
[10] L. Gupta and M. D. Srinath, "Invariant planar shape recognition using dynamic alignment," *Pattern Recognition*, vol. 21-3, pp. 235–239, 1988.
[11] H. C. Huang and B. H. Jansen, "EEG waveform analysis by means of dynamic time-warping," *Int. J. Biomed. Comput.*, vol. 17, pp. 135–144, 1985.
[12] K. Roberts, P. Lawrence, A. Eisen, and M. Hoirch, "Enhancement and dynamic time warping of somatosensory evoked potential components applied to patients with multiple sclerosis," *IEEE Trans. Biomed. Eng.*, vol. BME-34, no. 6, pp. 397–405, June 1987.
[13] L. Gupta and K. Malakapalli, "Robust partial shape classification using invariant breakpoints and dynamic alignment," *Pattern Recognition*, vol. 23, no. 10, pp. 1103–1111, 1990.
[14] L. Gupta and A. M. Upadhye, "Nonlinear alignment of neural net outputs for partial shape classification," *Pattern Recognition*, vol. 24, no. 10, pp. 943–948, 1991.
[15] L. Gupta, J. Wang, A. Charles, and P. Kisatsky, "Three-layer perceptron based classifiers for the partial shape classification problem," *Pattern Recognition*, vol. 27, no. 1, pp. 91–97, 1994.
[16] B. H. Jansen, J. R. Bourne, and J. W. Ward, "Autoregressive estimation of short segment spectra for computerized EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 9, pp. 628–638, Sept. 1981.
[17] B. H. Jansen, "Analysis of biomedical signals by means of linear modeling," *CRC Crit. Rev. Bioeng.*, vol. 12, pp. 343–392, 1985.
[18] J. Polich, "Semantic categorization and event-related potentials," *Brain Language*, vol. 26, pp. 304–321, 1985.
[19] M. D. Rugg, "Dissociation of semantic priming, word and nonword repetition effects by event-related potentials," *Quarterly J. Exp. Psych.*, vol. 39A, pp. 123–148, 1987.
[20] D. Friedman, S. Sutton, L. Putnam, C. Brown Jr., and L. Erlenmeyer-Kimling, "ERP components in picture matching in children and adults," *Psychophysiol.*, vol. 25, pp. 570– 589, 1988.
[21] D. L. Molfese, P. A. Morse, and C. J. Peters, "Auditory evoked responses to names for different objects: Cross-modal processing as a basis for infant language acquisition," *Dev. Psych.*, vol. 26, pp. 780–795, 1990.
[22] P. Domalski, M. E. Smith, and E. Halgren, "Cross-modal repetition effects on the N4," *Psychological Sci.*, vol. 2, pp. 173–178, 1991.
[23] D. L. Molfese, P. A. Morse, and R. Cornblatt, "Event related potential correlates of language comprehension in traumatic brain injured adults," *Brain Language*, (in press).
[24] A. F. Kramer and E. Donchin, "Brain potentials as indices of orthographic and phonological interaction during word matching," *J. Exp. Psych.: Learning, Memory, Cognition*, vol. 13, pp. 76–86, 1987.
[25] A. Kok and A. J. Rooyakkers, "ERP's to laterally presented pictures and words in a semantic categorization task," *Psychophysiol.*, vol. 23, pp. 672–683, 1986.
[26] L. Gupta, D. L. Molfese, and R. Tammana, "An artificial neural network approach to ERP classification," *Brain Cognition*, vol. 27, pp. 311–330, 1995.
[27] P. G. Simos, "Event-related potentials in a two choice task involving within form comparisons of pictures and words," masters thesis, Southern Illinois Univ., 1993.
[28] H. Jasper, "The ten-twenty electrode system of the International Federation of Societies for Electroencephalography: Appendix to report of the committee on methods of clinical examination in electroencephalography," *Electroencephal. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.



Lalit Gupta (M'86) received the B.E. (Hons.) degree in electrical engineering from Birla Institute of Technology and Science, Pilani, India, in 1976, the M.S. degree in digital systems from Brunel University, Middlesex, England, 1981, and the Ph.D. degree in electrical engineering from Southern Methodist University, Dallas, TX, in 1986.

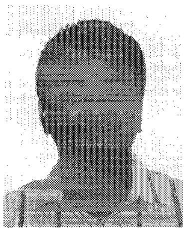
Since 1986, he has been with the Department of Electrical Engineering, Southern Illinois University at Carbondale, where he is currently an Associate Professor. His research interests include pattern recognition, digital signal processing, neural networks, and computer vision.

Dr. Gupta serves as an associate editor for the *Pattern Recognition Journal* and is a member of the Pattern Recognition Society, and the International Neural Network Society.

Dennis L. Molfese received the B.S. degree in psychology from Oklahoma City University, in 1969, and both the M.S. degree and the Ph.D. in psychology from the Pennsylvania State University, University Park, in 1970 and 1972, respectively.

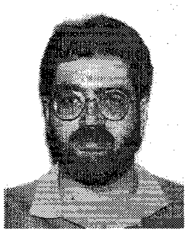
Since 1972 he has been at Southern Illinois University at Carbondale where he is currently Professor of Psychology and Physiology, and as well as Chairman of the Department of Behavioral and Social Sciences, School of Medicine. His research interests include neuroelectrical correlates of language, cognition, and learning, brain–behavior relationships, language development, and learning disabilities.

Dr. Molfese is Editor-in-Chief of the journal, *Development Neuropsychology*, and a member of 15 professional organizations including the American Psychological Society (Fellow), American Psychology Association, International Neuropsychological Society, and the Society for Research in Child Development.



Ravi Tammana (S'90–M'94) received the B.E. degree in electronics and communication engineering from Osmania University, Hyderabad, India, in 1985, the M.S. degree in electrical engineering and the Ph.D and degree in engineering sciences, in 1989 and 1994, respectively, from Southern Illinois University at Carbondale.

Since 1994, he has been working as a Systems Analyst at Audre Inc., San Diego, CA. His research interests include pattern recognition, neural networks, and digital signal processing.



Panagiotis G. Simos received the B.A. degree in psychology from the University of Crete, Greece, in 1990. He received the Ph.D. degree in experimental psychology/biopsychology from Southern Illinois University at Carbondale, in 1995.

He is currently a Research Fellow at the Department of Neurosurgery at the University of Texas, Houston. During his graduate studies, he was involved in research in the area of human development neurolinguistics using event-related potentials. His research focuses on the functional mapping of the frontal lobes using magnetoencephalography and magnetic resonance imaging. His clinical expertise is in intraoperative monitoring of the electrophysiological activity of the brain and the peripheral nervous system.