ELSEVIER

Clinical Neurophysiology 112 (2001) 1917–1924

www.elsevier.com/locate/clinph

Warp-averaging event-related potentials

Kongming Wang*, Henri Begleiter, Bernice Porjesz

Department of Psychiatry, Box 1203, Neurodynamics Laboratory, SUNY Health Science Center at Brooklyn, 450 Clarkson Avenue, Brooklyn, NY 11203, USA

Accepted 28 June 2001

Abstract

Objective: To align the repeated single trials of the event-related potential (ERP) in order to get an improved estimate of the ERP.

Methods: A new implementation of the dynamic time warping is applied to compute a warp-average of the single trials. The trilinear modeling method is applied to filter the single trials prior to alignment. Alignment is based on normalized signals and their estimated derivatives. These features reduce the misalignment due to aligning the random alpha waves, explaining amplitude differences in latency differences, or the seemingly small amplitudes of some components.

Results: Simulations and applications to visually evoked potentials show significant improvement over some commonly used methods.

Conclusions: The new implementation of the dynamic time warping can be used to align the major components (P1, N1, P2, N2, P3) of the repeated single trials. The average of the aligned single trials is an improved estimate of the ERP. This could lead to more accurate results in subsequent analysis. © 2001 Published by Elsevier Science Ireland Ltd.

Keywords: Dynamic time warping; Event-related potentials; Single trials

1. Introduction

It is well known that the event-related potentials (ERPs) obtained in repeated trials experience variable latencies of the components (for example, P3 and N1). The distortion is non-linear in the sense that the timing of features (say P1, N1, P2, N2, P3) in one trial cannot be linearly mapped to the timing of another trial.

The conventional procedure for estimating the ERP is to average the repeated trials. Single trials are assumed to be in the form

$$s_i(t) = p(t) + e_i(t),$$

where $s_i(t)$ is the ith recorded signal, $p(t)$, the evoked potential which is the same for all trials, and $e_i(t)$, the background electroencephalogram (EEG). The amplitude and latency variations among the single trials may lead to underestimating the amplitudes of the components (e.g. P3 and N1 of the average). The latency of a component of the average could be dictated by few trials with highest amplitudes.

In Woody's method (Woody, 1967), cross-correlation or cross-covariance is calculated between a template and a low-pass filtered single trial. Single trial latency is defined as the time lag by which the template should be shifted to maximize the cross-correlation or cross-covariance. This would improve the average of the single trials if the latency variation were simple shift. Other latency correction averaging methods involve peak identification, grouping of the peaks, and aligning the segments between peaks (Aunon et al., 1981; McGillem et al., 1985). It is difficult to identify peaks unequivocally from the noisy single trials. It is more problematic to group the peaks to represent different components. Except for the well-defined component such as P3, one could end up with aligning the alpha waves that are actually random relative to each other in the individual trials. The alignment between peaks is most likely linear even though the distortion is non-linear.

Pham et al. (1987) proposed an maximum-likelihood method, an alternative to Woody's method. The model for single trials is

$$s_i(t) = p(t + \tau_i) + e_i(t)$$

and τ_i stands for latency shift of the ith trial. The above equation is Fourier transformed and a log-likelihood function is formed. The latency shifts of the single trials can be obtained by maximizing the likelihood function. For estimating the N1–P2 complex, their simulation showed that the mean squared error was 5–10 times smaller for their method than for Woody's. Jaskowski and Verleger (1999) extended the model by assuming variations of signal amplitude from trial to trial.

$$s_i(t) = a_i p(t + \tau_i) + e_i(t)$$

* Corresponding author. Tel.: +1-718-270-2115; fax: +1-718-270-4081.
E-mail address: kmw@mendel.cns.hscbklyn.edu (K. Wang).

1388-2457/01/$ - see front matter © 2001 Published by Elsevier Science Ireland Ltd. CLINPH 2001049
PII: S1388-2457(01)00640-X

where a_i stands for amplitude of the ith trial. For estimating P3 amplitude and latency of single trials, the simulations of Jaskowski and Verleger (2000) showed that the methods of Pham et al. (1987) and Jaskowski and Verleger (1999) are slightly better than that of Woody's method and peak picking.

Jung et al. (1999) applied independent component analysis (ICA) to single trial ERPs. It is reported that ICA separated stimulus-locked (P1/N1), response-locked (P3), and no-phase locked background EEG into different independent components. The alignment is to align the response-locked P3 wave by shifting the reaction time to the median reaction time. The stimulus-locked P1/N1 component is unaligned.

Dynamic time warping was developed to eliminate the non-linear timing differences between a speech signal and a template for speech recognition (Sakoe and Chiba, 1978). It has been applied for classification of EEG waveforms (Huang and Jansen, 1985; Picton et al., 1988) and for aligning ERPs (Picton et al. 1995; Gupta et al., 1996). While the method of Gupta et al. aligns signals based on amplitude, the method of Picton et al. aligns signals based on both amplitude and slope. Wang and Gasser (1997, 1999) proposed a modified version and applied it to estimate the non-linear latency variations among sample curves. Their approach is motivated by the loose assumption that the single trials can be modeled as

$$s_i(t) = a_i(t)p(g_i(t)) + e_i(t), \qquad i = 1, \cdots, I \tag{1}$$

where $a_i(t)$ stands for time-dependent amplitude and $g_i(t)$ stands for non-linear latency. The method was tested with simulations and applications to growth data and the results are promising (Wang and Gasser, 1999). This method is modified in this paper to align single trials.

The performance of these methods depends heavily on the signal to noise (background EEG and random noise) ratio (SNR). The SNR of single trials is very small (0.1–1), and it is necessary to filter out noise before applying these methods. Picton et al. (1995) suggested that warping is probably more applicable to combining the already averaged data from different subjects. We will apply the trilinear modeling of ERP (Wang et al., 2000) to filter single trials before aligning them by dynamic time warping. Some details on trilinear modeling are given in Appendix A.

We performed simulations to evaluate this method and compare its performance with that of other methods. Since the maximum-likelihood method is better than Woody's method and peak picking (Jaskowski and Verleger, 2000), we need only to evaluate 3 methods: the maximum-likelihood method, the previously employed dynamic time warping method (Gupta et al., 1996), and the method proposed in this paper. Simulations are pseudo-real in the sense that the ERP signal and amplitude/latency variations are simulated, while real EEG signals are used as background noise. The simulated results show that the proposed method performs much better than other methods. As an example, we will also apply the method to real ERP data.

2. Method

2.1. Background

Since the ERP of each trial is influenced by many factors (for example, background EEG, experimental environment, and learning), the amplitude and the development pace of the ERP fluctuate: the amplitude changes from trial to trial and the components (for example, N1 and P3) occur at different latencies from trial to trial. Therefore, model (1) is a reasonable model for single trial ERPs. But it is not uniquely defined (hence a 'loose model').

The amplitude fluctuations $a_i(t)$ are non-negative functions. This implies that a component of the ERP $p(t)$ is either present in the ith trial or absent (for example, one only observes N1 and P3 in the ith trial). Since the single trials are under same experimental conditions and evoked by same stimulus, it is natural to assume that the amplitude fluctuations $a_i(t)$ are random processes with mean $a(t) \equiv 1$.This leads to

$$\sum_{i=1}^{I} a_i(t)/I \approx 1, \qquad \text{for all } t.$$

The latency variations $g_i(t)$ are assumed to be monotone increasing functions of t. It means that after each stimulus is applied, the brain experiences certain events in the same order (for example, P1, N1, P2, N2, P3). In general, $g_i(t)$ are non-linear functions of t. Since the stimulus is the same for each trial, one expects that the brain responds to the stimulus in almost identical way. Yet some random effects (for example, background EEG effects, experimental environment changes, learning) could cause fluctuations in both amplitude and latency. Again it is natural to assume that $g_i(t)$ are random processes with mean $g(t) \equiv t$, the natural development pace. This leads to

$$\sum_{i=1}^{I} g_i(t)/I \approx t, \qquad \text{for all } t.$$

To estimate the ERP from these trials, the latency fluctuations have to be estimated first. That is, some warping functions $h_i(t)$ are estimated based on the single trials such that the features (peaks, change points, etc.) of the trials are aligned. For example, the P3 components of the aligned signals, $s_i(h_i(t))$, occur at the same latency which is the average position of the P3 latencies of the individual trials.

As a simple example, assume that the amplitude $a_i(t)$ can be modeled by positive constants ($a_i(t) \equiv a_i$ for a constant a_i and for all t). These constants depend on the brain status when the stimulus is applied. Then we would like to estimate $p(t)$ from the single trials. For this example, the warping functions are the inversions of the latency fluctuations,

$h_i(t) = g_i^{-1}(t)$.

The aligned curves are $s_i(h_i(t)) = a_i p(t) + e_i(h_i(t))$ and the average of the aligned curves are

$$\hat{s}(t) = \sum_{i=1}^{I} s_i(h_i(t))/I = \left(\sum_{i=1}^{I} a_i/I\right)p(t) + \sum_{i=1}^{I} e_i(h_i(t))/I. \quad (2)$$

As discussed above, we can assume that $\sum_{i=1}^{I} a_i/I \approx 1$. For the background EEG, one usually assumes that the average background $\sum_{i=1}^{I} e_i/I$ of single trials is approximately zero. This leads to the approximation

$\hat{s}(t) \approx p(t)$.

2.2. Estimating the latency fluctuations

The warping functions are formulated as a solution set of a weighted least-square problem to minimize the overall latency fluctuations among the single trials. The purpose is to align the components (P1, N1, P2, N2, P3) of the single trials. Several steps are taken to reduce the alignment error.

First, the performance of the methods (peak picking, maximum likelihood, dynamic time warping) depends heavily on SNR, which is very low in single trials (Fig. 4). Artifacts such as eye movements are removed by rejecting those trials with a value above a threshold (73.3 μV in our visual P3 experiment). We employ the trilinear modeling (Wang et al., 2000) to improve the SNR of the single trials. We use the average of the single trials as an estimate of the ERP signal and the difference between the single trials and this average as estimates of the background signals. We then get a crude estimate of the SNR of single trials. Then we use the estimated SNR to determine the number of components in the trilinear model. Specifically, the components of the trilinear model will represent 100(SNR/(1 + SNR)) percent of the total variance.

Secondly, large difference in amplitudes of the single trials can also lead to misalignment. Assume that two trials have no latency difference but have different amplitudes. Then dynamic time warping would try to align a segment of the trial with higher amplitude to a single point of the trial with lower amplitude if the alignment was based solely on amplitudes. This misalignment is due to the least-squares minimization. Two methods are employed to prevent this kind of misalignment.

One of the methods is to normalize the single trials. Each trial is normalized so that the maximum amplitude is one. When two trials have no latency difference, the two normalized trials are the same and misalignment could not occur. Another method is to align the derivatives of the single trials simultaneously. The estimated derivatives of the single trials are normalized in the same way. This will not only reduce the chance of alignment error, but also align other features such as the change points of the trials. Kernel smoothing method is employed to estimate the derivatives of the single trials (see Gasser and Müller, 1984).

We now formulate the least-square problem. Let $S_i(t)$ be the normalized data of ith trial and $D_i(t)$ its normalized derivative. Then the warping functions are given by the solution of the equation

$$\min_{\{h_i\},\alpha} \sum_{i=1}^{I} \int_0^T [\alpha^2 (S_i(h_i(t)) - \sum_{j=1}^{I} S_j(h_j(t)/I)^2 + (1-\alpha)^2 \times (D_i(h_i(t)) - \sum_{j=1}^{I} D_j(h_j(t))/I)^2] \mathrm{d}t. \quad (3)$$

The restrictions on the warping functions $\{h_i\}$ are: (a) continuous and monotone increasing; (b) no excessive alignment ($|h_i(t) - t|$ is less than a given upper bound); (c) same starting point ($h_i(0) = 0$). Eq. (3) is formed to find the warping functions such that each trial is aligned to the average of the aligned trials. With a penalty term $\sum P(h_i(t) - t)$ where P is a strictly convex function with minima 0, formula (3) will be a strictly convex problem and has a unique minima.

The alignment of the amplitudes (first term of Eq. (3)) and the alignment of the derivatives (second term of Eq. (3)) are tuned by a parameter α. The minimum of the total weight $\alpha^2 + (1-\alpha)^2$ has a minimum value 0.5 when $\alpha = 0.5$. This means that the alignment of the amplitudes and the alignment of the derivatives are equally weighted unless the data clearly suggest otherwise. The perfect warping function for the models discussed in Section 1 can be obtained by this modified dynamic time warping (Wang and Gasser, 1999).

2.3. Estimating the ERP

Since each warping function aligns a trial to the average of the aligned trials, the average of the warping functions

$$h(t) = \sum_{j=1}^{I} h_j(t)/I$$

is the average development pace of the single trials. Therefore, $h(t)$ is the estimated latency of the ERP.

The amplitude of the ERP at time $h(t)$ is estimated by the average of the aligned trials

$$\hat{p}(t) = \sum_{j=1}^{I} \tilde{s}_j(h_j(t))/I$$

where $\tilde{s}_j(t)$ is the smooth version of $s_j(t)$ obtained by the trilinear modeling. The estimated ERP is in parametrized form $(h(t), \hat{p}(t))$, $0 \le t \le T$. The restrictions on the warping functions imply that $h(0) = 0$.

2.4. Algorithm

The following descriptive algorithm is for understanding the method better and can be used as guidelines for programming.

Step 1: Apply trilinear modeling to improve the SNR of

single trials. This step uses data from all electrodes and all trials.

The rest of the algorithm is to align single trials from each electrode, one at a time. Let X be the single trial data from an electrode. Then X is an m by n matrix if there are m trials and n time points per trial.

Step 2: Estimate the derivatives of X. We apply kernel method to each row of X to compute the derivatives at sampling time points. Let X_d denote the derivatives. Then X_d and X have same size. Oversmoothing is employed to reduce the effects of alpha waves and noise.

Step 3: Normalize data as follows:

$$\bar{X}(i,:) = X(i,:)/\max(|X(i,:)|)$$

$$\bar{X}_\mathrm{d}(i,:) = X_\mathrm{d}(i,:)/\max(|X_\mathrm{d}(i,:)|).$$

Here $X(i,:)$ stands for the ith row of X and $|\cdot|$ for absolute value.

Step 4: for each α in a grid on [0,1] (we used 3 values 0.3, 0.5, 0.7), run dynamic time warping on data $\bar{X}$ and $\bar{X}_\mathrm{d}$ using cost function (3). Let C_α be the cost. Then the best warping is indexed by $\alpha^* = \arg\min_\alpha C_\alpha$.

Step 5: Apply the best warping path obtained in step 4 to the data X. Average the aligned X as an estimate of ERP from the electrode. Note that the normalized data $\bar{X}$ and $\bar{X}_\mathrm{d}$ are only used to compute the best warping path, not the average.

Repeat steps 2–5 for each electrode. The algorithm is finished afterwards.

Computational cost depends on many factors: CPU speed, memory size, programming language, restrictions on warping functions ((b) after formula (3), number of α values to choose from, etc. We have not evaluated the computational cost.

3. Simulation

The simulated data are generated from the model

$$f_i(t) = a_i p(g_i(t)) + e_i(t).$$

Here a_i is an individual constant which represents the amplitude variation of trial i. Our visual P3 experiment runs 20–30 target trials and is recorded at 61 channels. Therefore, each simulation replication will have 61 channels and 25 repeated trials. The noise term $e_i(t)$ is obtained from recorded EEG data of 20 subjects. For each subject, the EEG data from each channel is cut into segments of the same length of the simulated ERP, $p(t)$. The collection has 2741 EEG segments from the 20 subjects. For each simulation replication, a random selection of 25 EEG segments from the collection is taken as the EEG data for this replication. The EEG data are then scaled to give a specified SNR:

$$\mathrm{SNR} = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}[a_i p(g_i(t))]^2}{\sum_{i=1}^{I}\sum_{t=1}^{T}[e_i(t)]^2}$$

The simulated ERP, $p(t)$, is plotted in Fig. 1. The ERP has several positive and negative components, simulating the basic pattern of visually evoked potentials.

Non-linear latency variation is generated by the formula

$$g_i(t) = t + b_i t(T - t)/T$$

with a constant b_i. To insure that g_i is strictly increasing, the condition $|b_i| < 1$ has to be satisfied. It is obvious that $g_i(0) = 0$ and $g_i(T) = T$.

For comparisons between the estimated ERP and the simulated ERP, we will require that

$$\frac{1}{I}\sum_{i=1}^{I} a_i = 1, \qquad \frac{1}{I}\sum_{i=1}^{I} g_i(t) = t.$$

It follows that the simulated ERP and the average of the aligned curves (the estimated ERP) should be identical. The latter condition implies that $\sum_{i=1}^{I} b_i = 0$.

The parameters are generated as follows. For a_i, first generate I normal random variables N_i, $i = 1,\ldots,I$. Then normalize them as $A_i = 0.5N_i/N$ where $N = \max_i|N_i|$. Set

$$a_i = 1 + A_i - \frac{1}{I}\sum_{j=1}^{I} A_j.$$

For the parameters b_i, we generate I normal random variables M_i, $i = 1,\ldots,I$ and let

$$b_i = 0.2\left(M_i - \frac{1}{I}\sum_{j=1}^{I} M_j\right).$$

It is easy to see that $|g_i(t) - t| = |b_i t(T - t)/T| \le T/4$. The upper bound for the restriction (b) on the warping functions should be set to $T/4$. This information is usually unavailable for real problems, but one does not expect excessive latency variation. The upper bound $T/4$ should be good for most real problems and we take it for the simulations of this section. Note that a smaller upper bound implies a narrow search and would speed up the dynamic time warping. For aligning the single trials in a visual P3

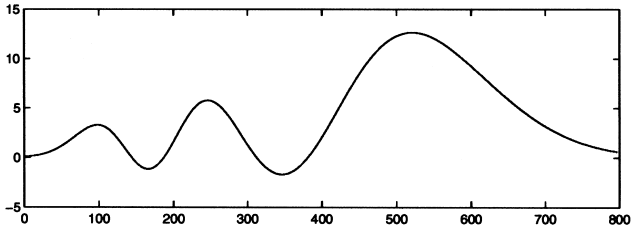

Fig. 1. The simulated ERP $p(t)$. Horizontal unit is in ms and vertical unit is in μV.

experiment, the upper bound is set at 60 ms in Section 4 based on experience.

The mean squared error of estimating the ERP, defined by

$$\text{MSEA} = \int_0^T [p(h(t)) - \hat{p}(t)]^2 \mathrm{d}t/T,$$

is used to evaluate the performance of the methods. For each method, we will compute the average MSEA (AMSEA) of all 61 channels for each replication. The simulation results with 200 replications are plotted in Fig. 2. The SNR for the simulated trials are randomly chosen, ranging from 0.2 to 1. Fig. 2 plots the ratio of the AMSEA for a method and the AMSEA of the conventional average.

The average ratio of the 200 replications is 0.89 for the method of Gupta et al., an 11% improvement over the conventional averaging. The average ratio is 0.73 for the maximum-likelihood method of Pham et al., a 27% improvement. Since the maximum-likelihood method estimates only a shift, we only used the data from 300 to 800 ms, or the P3 component, to estimate the shift. Furthermore, data are lower-pass filtered at 6 Hz, following the idea of Jaskowski and Verleger (2000), before applying the maximum-likelihood method. The average ratio is 0.56 for our new method, a 44% improvement over the conventional averaging method. So in terms of MSEA, the modified dynamic time warping performs much better. After all the maximum-likelihood method is not designed for handling non-linear latency variations. This simulation shows that both data normalization and aligning derivatives are necessary to reduce the misalignment. Finally, to see how important it is to improve SNR before alignment, we applied the modified dynamic time warping and conventional averaging to a simulated clean data set (without adding noise to the simulated signals). The alignment is very good, a 76% improvement over the conventional averaging method. Recall that the improvement is 44% for noisy data.

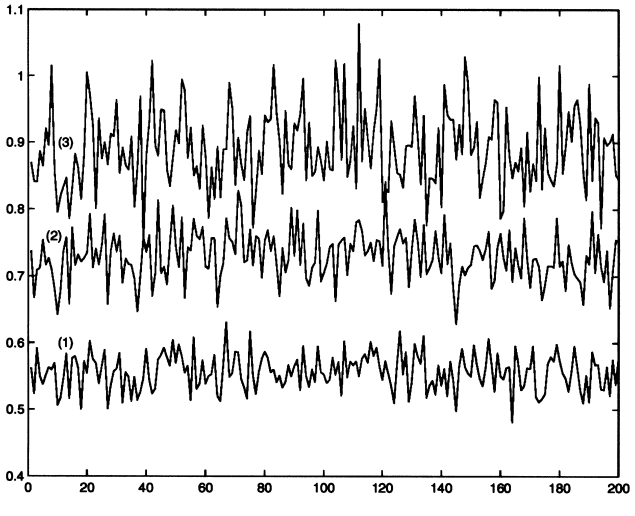

Fig. 2. Each line represents the ratio of the average MSEA of a method and the average MSEA of the conventional averaging method in 200 replications. (1) The modified dynamic time warping; (2) the maximum-likelihood method of Pham et al.; and (3) the dynamic time warping of Gupta et al. Horizontal unit is the index of the replications.

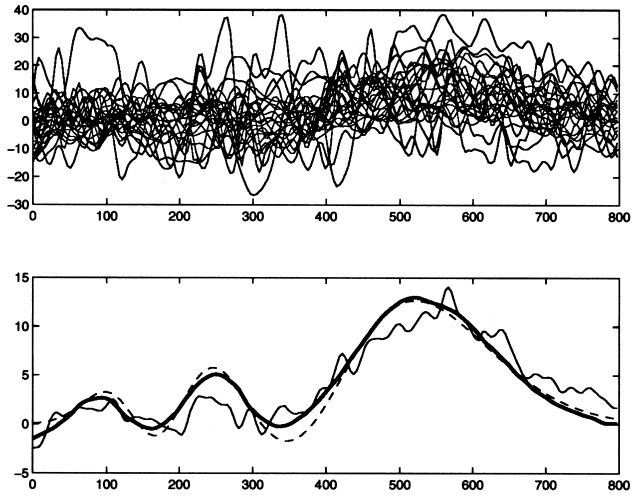

Fig. 3. A typical run of the simulation. Top: the simulated data from a channel. Bottom: the true ERP signal (dashed line), the estimated ERP by the modified dynamic time warping (wide solid line), and the conventional average (thin solid line). Horizontal unit is in ms and vertical unit is in μV.

A typical replication of the simulation is plotted in Fig. 3. Fig. 4 shows the effects of noise on alignment.

4. Application to VP3 data

In the visual P3 experiment, subjects are presented with 280 visual stimuli with a uniform inter-stimulus interval of 1.6 s. There are 210 non-target stimuli in the shape of an outline of a square, 35 target stimuli in the shape of an X, and 35 novel stimuli, each a different colored polygon or

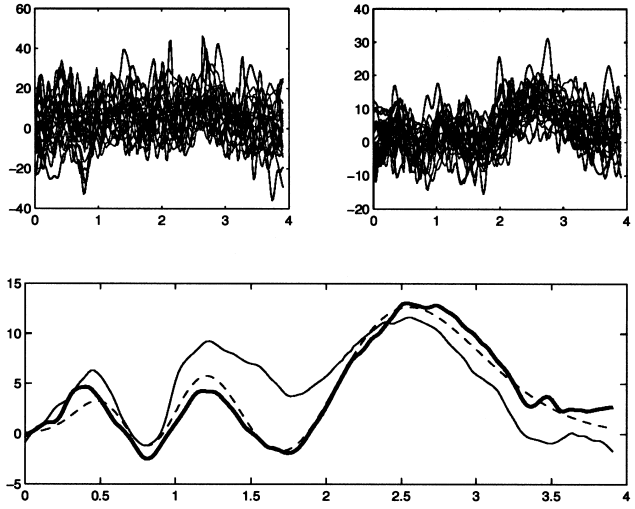

Fig. 4. Top-left: simulated data with SNR 0.2. Top-right: simulated data with SNR 1. Bottom: simulated ERP as in Fig. 1 (dash line), warp-average of the data with SNR 0.2 (solid line), and warp-average of the data with SNR 1 (wide solid line).

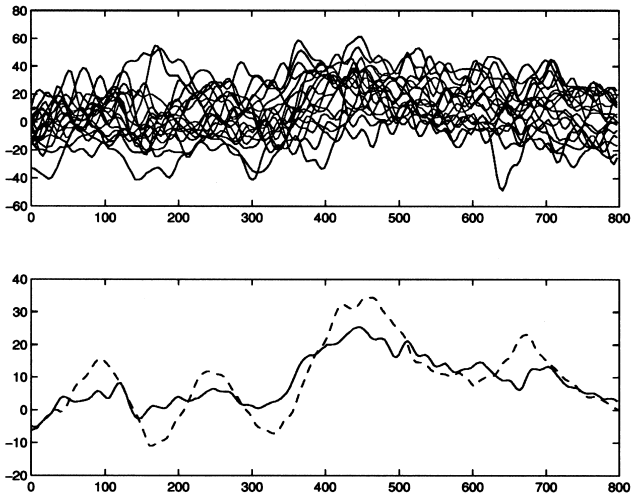

Fig. 5. Application to visual ERP data. Top: the 25 target trials recorded at PZ electrode from a subject. Bottom: the estimated ERP by the modified dynamic time warping (dashed line) and the conventional average (solid line). Horizontal unit is in ms and vertical unit is in μV.

other geometrical figure. The different types of stimuli are presented in random order. Subjects are instructed to respond to target stimuli by pressing a button. The visual P3 data are recorded from 61 channels (the international 10/20 system). After amplification by a factor 10 000, artifact threshold is set at 73.3 μV. Any trial with a value above the threshold is rejected. The number of the recorded single trials is different for different subjects and experimental conditions due to the artifact thresholding.

Fig. 5 plots the data from a subject, recorded at PZ electrode. There are 25 clean (no artifact) target trials recorded from the subject. We apply the modified dynamic time warping to the target trials and obtain an estimate (warp-average) of the ERP. The result is plotted in Fig. 5.

5. Discussion

Many methods have been proposed for computing an improved average of single trials. The objective is to eliminate the latency differences of the components across the repeated trials. The correlation techniques (Woody, 1967; Wastell, 1977) and the iterative Fisher scoring (Möcks et al., 1988) estimate a latency shift in the entire waveform or a shift in a prominent peak in each trial. The single trials are then registered (aligned) according to the estimated shift and averaged. Non-linear time variations such as expansion or compression between consecutive components are not taken into account. The latency correction averaging technique (Aunon et al., 1981; McGillem et al., 1985) identifies peaks (components) of the single trials and then aligns the segments between peaks. Identifying peaks across noisy single trials is a difficult task. These and other parametric methods (Gasser et al., 1983; Gevins, 1984) consider the measured response to be an additive superposition of the evoked potential (ERP signal) and the background EEG activity (noise). They then make assumptions on the statistical properties of signal and noise. It is often assumed that signal and noise are statistically independent. All those assumptions are of course approximations. Some methods do not make assumptions on the statistical properties of signal and noise (Gevins et al., 1986; Laskaris et al., 1997). These methods apply pattern classification or moving average techniques to select good trials for averaging. The idea is that task-related signals may not be obtained in some single trials and such trials should be excluded for averaging. The ICA approach of Jung et al. (1999) aligns only P3 component by shifting reaction times to the median reaction time. The numbers of components in the component classes (stimulus-locked, response-locked, and no-phase locked) vary across subjects and the ICA components cannot be ordered in variance that they represent. One would need a systematic method to identify the stimulus-locked, response-locked, and no-phase locked ICA components.

Dynamic time warping provides a non-linear, non-parametric, and smooth alignment of evoked potentials. It is an automated method that aligns potentials based on the entire signals. This reduces the human error of picking up corresponding components across trials. The method was initially developed for speech analysis and speech recognition (Sakoe and Chiba, 1978; Rabiner and Schmidt, 1980; Parsons, 1986). The idea is that people speak the same word or sentence in different speed and loudness. For recognizing a sample speech, the sample is warped onto a template by eliminating the timing differences of the features between the sample and the template. The template can be obtained as an aligned average of the spoken signals by many people. Wang and Gasser (1997, 1999) proposed a modified dynamic time warping for computing an aligned average of sample curves observed from general experiments. The features that are aligned by the modified method are peaks of the sample curves and their derivatives. The derivatives are estimated from noisy data by kernel smoothing. Gupta et al. (1996) proposed several implementations for aligning evoked potentials.

It is straightforward to apply dynamic time warping to single trials since an aligned average of single trials is simply a template for ERP. Picton et al. (1988) applied dynamic time warping to align the brain-stem auditory evoked potentials recorded from different subjects prior to averaging. They compared dynamic time warping with experienced human interpreter. They found that warping is very accurate in identifying the waves of normal brain-stem auditory evoked potentials with error rates between 0 and 4%, and reasonably accurate in identifying the peaks in abnormal wave forms with error rates between 3 and 18%. Eisen et al. (1986) used dynamic time warping to align somatosensory evoked potentials to a standard wave derived from normal subjects. They found that the mean costs for such alignment are 1.305 for normal subjects and 5.089 for

patients with definite or possible multiple sclerosis. This indicates that the alignment using dynamic time warping can help in clinical diagnosis of patients. Other applications of dynamic time warping are reported by Jansen and Huang (1985), Roberts et al. (1987), and Huang and Jansen (1985).

Our implementation is based on the modified method of Wang and Gasser (1997, 1999). The single trials are smoothed by trilinear modeling to improve the SNR. A statistical method (kernel smoothing) is employed to estimate the derivatives of the smoothed single trials. In aligning the single trials, both the recorded signals and the estimated derivatives are aligned. The amounts of alignment of the recorded signals and the estimated derivatives are weighted according to the data. The involvement of derivatives helps to make more accurate alignment, especially the components with smaller amplitudes. The signals and the derivatives are normalized for the purpose of computing the time differences among the single trials. This prevents the misalignment in which amplitude differences are explained in latency differences.

A main drawback of some early methods is the alignment of alpha waves. Several steps of the new implementation help to resolve this problem. The components of the trilinear method are obtained by singular value decomposition (SVD). Each component is a linear combination of the signals from all trials and all electrodes. This reduces the alpha wave because of its randomness with respect to the time-locked feature of ERP. Oversmoothing reduces the effects of alpha wave on estimated derivatives that are used to compute the alignment. Finally, normalization of single trials and their derivatives seems also to reduce the effects of alpha wave on alignment.

We have focused on computing an aligned average from single trials. As it has been done before (Picton et al., 1988; Eisen et al., 1986), the modified dynamic time warping should improve the accuracy of accessing variability of the single trials or the recorded ERPs from different subjects, the accuracy of clinical diagnosis, and the accuracy of classifying patients.

Acknowledgements

This research is supported by N.I.H. grants AA 05524 and AA 02686. We want to thank Vladimir Kotlyarevsky, Marty Krakowsky, Aquanette Sass, Arthur Stimus, Sandi Watson for providing the ERP data, and David Chorlian for many discussions.

Appendix A. Trilinear modeling

Trilinear modeling is to represent the data X_i as

$$X_i = BA_iC \tag{A1}$$

where each column of B is a spatial component and each row of C is a temporal component. The matrix A_i is the loading matrix of trial i.

For estimating temporal components, each time point is a variable and each electrode produces an observation. The concatenation of the trial data matrix X_i by variables gives the combined data matrix

$$D = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \end{pmatrix}.$$

SVD of the data D results in

$$D = USV$$

where S is a diagonal matrix and where the rows of V are orthogonal temporal components. The diagonal elements of S give the variances explained by the temporal components, respectively. If the first K temporal components explain at least a specified portion of temporal variance, then the first K rows of V will be chosen as the temporal components. The rest will be considered as negligible components or noise.

For estimating spatial components, each electrode is considered to be a variable and each time point produces an observation. In this case, the concatenation of the trial data matrix X_i by variables gives the combined data matrix

$$D_s = (X_1, X_2, \ldots).$$

Then SVD of D_s results in

$$D_s = U_sS_sV_s.$$

As in the estimation of the temporal components, we choose the first K_s columns of U_s as spatial components and consider the rest of the columns of U_s as negligible components or noise.

With the estimated spatial and temporal components (the first K_s columns of U_s and the first K rows of V), the data matrix X_i of subject i can be decomposed as in Eq. (A1). This representation is unique because of the orthogonality of the components. It is desirable that the loading matrix A_i has as few significant entries as possible. This is also neurophysiologically meaningful because it leads to the interpretation that the activity of a spatial component (source distribution) is mainly captured by one temporal component. To this end, we compute the average loading matrix A of all trials $A = \sum_{i=1}^{m} A_i/m$ if we have m trials) and its SVD decomposition $A = PRQ$. Since R is diagonal, the best rotation matrix for rotating temporal components is Q and the best rotation matrix for rotating spatial components is P. Multiplying the first K_s columns of U_s by P and the first K rows of V by Q, we obtain the common spatial components B and the common temporal components C of the formula (A1), respectively.

With B and C at hand, one can compute the loading matrix for trial i as $A_i = B'X_iC'$. The ERP in trial i is then estimated by BA_iC.

References

Aunon J, McGillem C, Childers D. Signal processing in evoked potential research: averaging and modeling. CRC Crit Rev Bioeng 1981:323–367.

Eisen A, Roberts K, Lawrence P. Morphological measurement of the SEP using a dynamic time warping algorithm. Electroenceph clin Neurophysiol 1986;65:136–141.

Huang H, Jansen B. EEG waveform analysis by means of dynamic time-warping. Int J Biomed Comput 1985;17:135–144.

Gasser T, Möcks J, Verleger R. SELAVSO: a method to deal with trial-to-trial variability of evoked potentials. Electroenceph clin Neurophysiol 1983;55:717–723.

Gasser T, Müller H. Estimating regression functions and their derivatives by the kernel method. Scand J Statist 1984;11:171–185.

Gevins A. Analysis of the electromagnetic signals of the human brain: milestones obstacles and goals. IEEE Trans Biomed Eng 1984;31:833–850.

Gevins A, Morgan N, Bressler S, Doyle J, Cutillo B. Improved event-related potential estimation using statistical pattern classification. Electroenceph clin Neurophysiol 1986;64:177–186.

Gupta L, Molfese D, Tammana R, Simos P. Nonlinear alignment and averaging for estimating the evoked potentials. IEEE Trans Biomed Eng 1996;43:348–356.

Jansen B, Huang H. Automated morphological analysis by means of dynamic time warping. Electroenceph clin Neurophysiol 1985;60:282–284.

Jaskowski P, Verleger R. Amplitudes and latencies of single trial ERP's estimated by a maximum-likelihood method. IEEE Trans Biomed Eng 1999;46:987–993.

Jaskowski P, Verleger R. An evaluation of methods for single-trial estimation of P3 latency. Psychophysiology 2000;37:153–162.

Jung T, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski T. Analyzing and visualizing single-trial event-related potentials. Adv Neural Inform Proc Sys 1999;11:118–124.

Laskaris N, Fotopoulus S, Papathanasopoulos P, Bezerianos A. Robust moving averages with Hopfield neural network implementation for monitoring evoked potential signals. Evoked Potentials 1997;104:151–156.

McGillem C, Aunon J, Pomalaza C. Improved waveform estimation procedures for event-related potentials. IEEE Trans Biomed Eng 1985;32:371–379.

Möcks J, Kohler W, Gasser T, Pham D. Novel approaches to the problem of latency jitter. Psychophysiology 1988;25(2):217–226.

Parsons T. Voice and speech processing, New York, NY: McGraw-Hill, 1986.

Pham D, Möcks J, Köhler W, Gasser T. Variable latencies of noisy signals: estimation and testing in brain potential data. Biometrika 1987;74:525–533.

Picton T, Hunt M, Mowrey R, Rodriguez R, Maru J. Evaluation of brainstem auditory evoked potentials using dynamic time warping. Electroenceph clin Neurophysiol 1988;71:212–225.

Picton T, Lins O, Scherg M. The recording and analysis of event-related potentials. In: Boller F, Grafman J, editors. Handbook of neuropsychology, vol. 10. Amsterdam: Elsevier, 1995. pp. 3–73.

Rabiner L, Schmidt C. Application of dynamic time warping to connected digit recognition. IEEE Trans Acoust Speech Signal Process 1980;28:377–388.

Roberts K, Lawrence P, Eisen A, Hoirch M. Enhancement and dynamic time warping of somatosensory evoked potential components applied patients with multiple sclerosis. IEEE Trans Biomed Eng 1987;34:397–405.

Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 1978;26:43–49.

Wang K, Gasser Th. Alignment of curves by dynamic time warping. Ann Statist 1997;25:1251–1276.

Wang K, Gasser Th. Synchronizing sample curves nonparametrically. Ann Statist 1999;27:439–460.

Wang K, Begleiter H, Porjesz B. Trilinear modeling of event-related potentials. Brain Topogr 2000;12:263–271.

Wastell D. Statistical detection of individual evoked responses: an evaluation of Woody's adaptive filter. Electroenceph clin Neurophysiol 1977;42:835–839.

Woody C. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med Biol Eng 1967;5:539–553.