Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Using Single-Trial EEG to Estimate the Timing of Target Onset During Rapid Serial Visual Presentation

An Luo and Paul Sajda

Abstract— **The timing of a behavioral response, such as a button press in reaction to a visual stimulus, is highly variable across trials. In this paper we describe a methodology for single-trial analysis of electroencephalography (EEG) which can be used to reduce the error in the estimation of the timing of the behavioral response and thus reduce the error in estimating the onset time of the stimulus. We consider a rapid serial visual presentation (RSVP) paradigm consisting of concatenated video clips and where subjects are instructed to respond when they see a predefined target. We show that a linear discriminator, with inputs distributed across sensors and time and chosen via an information theoretic feature selection criterion, can be used in conjunction with the response to yield a lower error estimate of the onset time of the target stimulus compared to the response time. We compare our results to response time and previous EEG approaches using fixed windows in time, showing that our method has the lowest estimation error. We discuss potential applications, specifically with respect to cortically-coupled computer vision based triage of large image databases.**

I. INTRODUCTION

During perceptual decision making, behavioral response times can vary significantly. The cortical origins of such response time variability, specifically during rapid serial visual presentation, have been previously studied [1], with results showing that the variability appears to arise during the characteristic P300 response. Such variability can be seen from the perspective of estimating the onset time of the stimulus, where the estimation requires an EEG signal which is detectable single-trial and is locked to the visual stimulus.

In this paper we describe a method for estimating the visual stimulus onset time using feature selection in the space of the EEG data. We use a mutual information based feature selection (MIFS) method to identify signatures in the EEG which are distributed across sensor and time and are informative about the image class. We then use a linear discriminator to project the identified feature space to a one-dimensional space that best discriminates target from non-target responses. These steps are computed using training data. We apply the discriminator across time and use the timing of maximal discrimination in the EEG sensor space and the mean response time computed from the training data to estimate the onset time of the visual stimulus on test data. We compare this method with previous methods which use fixed time EEG features and show that the estimation of the onset time of the visual stimulus is better when using our feature selection method.

This work was supported by DARPA and NGA (HM1582-05-C-0043)

An Luo and Paul Sajda are with the Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA al2082@columbia.edu and ps629@columbia.edu

II. METHODS

A. Paradigm

The experimental paradigm consisted of two types of video clips (each is 10 seconds long): 1) a “nature” clip without people in it (“distractor clip”) and 2) a clip with a person(s) in it, from the first frame to the last (“target clip”). A set of ten clips was concatenated to form a trial, where one (and only one) of the ten clips was a ”target” clip. The target clip could be from the second to the last clip in a trial. Each trial started with a fixation image (also 10 seconds long, not part of the ten video clips) which was a white cross on the black background. The trial sequences were presented ten times faster than the normal speed. There were 60 trials in all.

Subjects were required to make a button response as soon as they saw a target. Sequences were randomly presented to the subjects.

B. Subjects and Data Preprocessing

Seven adult subjects participated in the experiment. All subjects had normal or corrected to normal vision and reported no history of neurological problems. Informed consent was obtained from all participants in accordance with the guidelines and approval of the Columbia University Institutional Review Board. One subject was later excluded from analysis because of frequent eye blinks and eye movements.

Sixty-channel EEG data as well as the button response and stimulus events were recorded in an electrostatically shielded room. The sampling frequency was set to 1 kHz. Raw data were visually-inspected and trials with large eye movement were excluded. Following data acquisition, DC drift and high-frequency noise were removed by software-based filters. Eye-blink and eye-movement activity was recorded and later removed from the EEG recordings using PCA ([2]).

C. Estimating the Timing of Target Onset

Given a button press in response to a target, if one knows the onset time of the stimulus, one also knows the response time. We estimated the response time using a three-step process. First we identified a set of EEG features, distributed across sensors and time, using a MIFS method. Next a linear discriminator was trained using logistic regression that projected from the feature space found in the first step to a one-dimensional space and maximally discriminated target from non-target responses. These two steps were based on the training data. Lastly, given a button response, the target onset time was estimated based on the mean response time of the training data and the projection result from the linear discriminator.

1) Detecting Neural Signatures of Perceived Target Clips using MIFS: To find a feature space optimal for estimating the stimulus onset time, we first selected several features from high-dimensional spatio-temporal EEG data that best discriminated between responses to target and non-target stimuli.

In [3], Battiti proposed a MIFS method in supervised classification, where the following objective function is used to select a new feature $\hat{f}$,

$$\hat{f} = \arg\max_{f}[I(f,c) - \beta \sum_{i} I(f, f_i')]. \tag{1}$$

The method is iterative, such that a set of features is selected for the discrimination. The first feature is the one that has the maximal mutual information with the class labels. Here f are candidate features, c is the vector of class labels, and f_i's are features that already have been selected. $I(a,b)$ is the mutual information between feature a and b, and it is defined as,

$$I(a,b) = I(b,a) = \sum_{a,b} P(a,b) \log \frac{P(a,b)}{P(a)P(b)}. \tag{2}$$

β is a weighting on the penalty of choosing features which have high mutual information with previously chosen features. i.e., the method searches for the feature that is informative about the class without being predictable from the current set of features. In [3], it was found that a value of β between 0.5 and 1.0 was appropriate for many classification tasks.

Equation (1) can be seen as a simplification of the more generalized form:

$$\hat{f} = \arg\max_{f}[I(f,c) - \beta \cdot I(f, \bigcup_{i} f_i')], \tag{3}$$

where $I(f, \bigcup_i f_i')$ is the mutual information between the new feature f and all selected features. Since computing $I(f, \bigcup_i f_i')$ is expensive and the "curse of dimensionality" appears when the number of selected features is large, instead, in [3] they used $\sum_i (f, f_i')$ (the sum of the mutual information between a new feature and each selected feature). However, as the number of selected features becomes large, this tends to overestimate $I(f, \bigcup_i f_i')$, and the second term in Equation (1), $\beta \cdot \sum_i (f, f_i')$, will dominate the first term, $I(f,c)$. As an alternative, Peng et al.([4]) proposed using $\beta = 1/N$, where N is the number of selected features. Combined with $\sum_i (f, f_i')$, this means they consider the average mutual information between the new feature and each selected feature.

In our method, we combined these two solutions such that,

$$\hat{f} = \arg\max_{f}[I(f,c) - \frac{\beta}{N} \sum_{i} I(f, f_i')]. \tag{4}$$

The value for β and the number of features were optimized by analysis on training data.

The difference between the true mutual information $\bar{I}$ and the estimated I can be approximated as ([3]),

$$\Delta I = I - \bar{I} \approx \frac{1}{2M}(K_c K_f - K_c - K_f). \tag{5}$$

Where K_f is the number of quantization levels for the feature data and K_c for the class labels. M is the number of samples. For our two-class discrimination problem, $K_c = 2$. From this equation we see that the smaller the value of K_f, the smaller ΔI will be. However, if the statistical distributions contain substantial structure, using a small value of K_f tends to cancel these details and reduces the estimated mutual information [3]. In [3], the authors recommended $K_f = 10$. Here we used a K_f value of eight and obtained good results.

2) Linear Discrimination (LD): After a set of features were found, linear discrimination using logistic regression was applied on the selected features, where

$$y(t) = \sum_{i} v_i x_i(t) \tag{6}$$

is maximally discriminating between two conditions, with non-target trials corresponding to lower y values and target to higher values. Here i indexes the features and $x_i(t)$s are the feature data. In our application, $x_i(t)$s were EEG data distributed across sensors and time with a window length of T ($t = 1, 2 \ldots, T$), extracted by the MIFS method. We set the window length to be 50ms in our analysis. All $x_i(t)$s were shifted to the stimulus onset for later processing. An optimal set of weights, v_i, were computed to maximally discriminate between two labelled classes of features. Since each feature was a multi-dimensional vector in our case, the discriminating component $\bar{y}$ was the average value of $y(t)$ over t. For more information, refer to [5].

An illustration of this feature selection process is shown in Fig. 1(a).

3) Target Onset Time Estimation: For each subject, the target onset time was first estimated as their mean response time before the button press, denoted by τ_0. Then a modification was made based on their EEG response. We looked for a time window around τ_0 that gave the largest $\bar{y}$ value. The onset of this window, $\hat{\tau}$, was our new estimate of the target onset time, as Equations (7) and (8) show.

$$\hat{\tau} = \arg\max_{\tau} \bar{y}(\tau) = \arg\max_{\tau}(\frac{1}{T}\sum_{t=1}^{T} y(t,\tau)), \tag{7}$$

where

$$y(t,\tau) = \sum_{i} v_i x_i(t,\tau). \tag{8}$$

Here the discriminating component $\bar{y}$ is indexed by τ, which is defined as 400ms before to 200ms after the mean response time estimate of the stimulus onset τ_0, with a step size of 50ms.

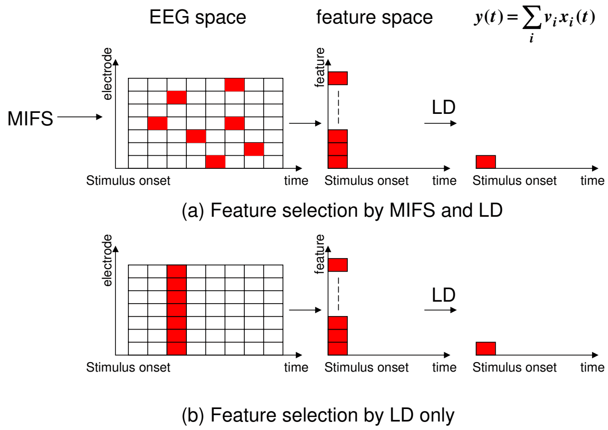

Fig. 1. (a):Feature selection using MIFS and LD. A set of spatio-temporal features was selected by the MIFS method, then these features were shifted to the stimulus onset. After LD, $y(t)$ was computed. (b): Feature selection using LD only. EEG data from all sensors at the same time window were the features for linear discrimination.

III. RESULTS

In our analysis, the first twenty trials were used as the training data, with the rest used for testing (forty minus the excluded trials).

A. Behavioral Response

During the task, the subjects responded as soon as they saw a target clip. Their mean and standard deviation of response time is shown in Table I. In each subject's test trials, the target onset time was first estimated to be the mean response time (calculated from the training data) before the button press. The dashed black line in Fig. 3 shows this estimate for one subject, and the solid black line shows the time when the button response happens. The estimation error for all six subjects is presented in Fig. 5, as the dashed black curve shows.

	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6
mean(ms)	842	705	836	710	877	921
std(ms)	156	94	86	130	240	144

TABLE I
THE SIX SUBJECTS' MEAN AND STANDARD DEVIATION OF THEIR RESPONSE TIME.

B. Target Onset Estimation Based on EEG

In the training data, the spatio-temporal EEG signals from 200ms to 850ms after the onset of each video clip were used for extracting the features and discriminator training. We further divided this 650ms window into 13 small time windows, each 50ms long. As 60 scalp sensors were recorded, the size of our feature pool was 60×13=780 spatio-temporal features. The MIFS method was used to extract features from the feature pool.

We next found the appropriate values for β and the feature number. The linear discriminator's A_z value, computed using ROC analysis [6], was used as the measure of the discrimination method's performance. The result is shown in Fig. 2 (left).

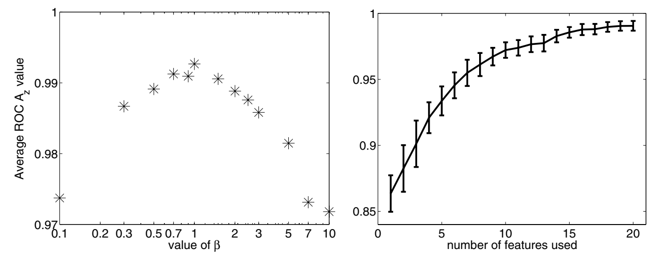

Fig. 2. Average ROC A_z value as a function of β (left) and number of features (right) of the six subjects. Left: The number of features was first set to be 20. Right: $\beta = 1.5$.

From Fig. 2 we see that within a large range of β, the discrimination method was robust and performed well. In our implementation we set $\beta = 1.5$. Then we tested the relationship between the feature number and the discrimination results. Again A_z value from ROC analysis was used as the measure. The result is shown in Fig. 2 (right). We see improvement in the discrimination performance when the number of features increases from one to twenty. When the number of features is above 12, the performance remains more-or-less constant. However as the performance improves, the computational cost also increases. In our implementation, we set the number of features to be 17.

Next a final discriminator was trained and the projection from the original feature space to the new space was obtained for each subject based on their training data using logistic regression. Next the discriminating component $\bar{y}$ values around the original estimated stimulus onset time τ_0 (from 400ms before to 200ms after, in our analysis) were computed for the test data, and the window that gave the optimal $\bar{y}$ result was our new estimation of the target onset, as Equations (7) and (8) show. Fig. 3 shows one subject's response-locked single-trial discriminating component $\bar{y}$ across time, we see that the discriminating component matches the target onset time well.

For a button press, once we locate the target onset, we can estimate the response time to be the time period between the target onset and the button press. Fig. 4 shows the estimated response time as a function of the real response time for all subjects' test data.

Fig. 5 compares the result using our method with that of only using the button response information. The left figure shows the probability density function of the estimation errors (in absolute values), and the right figure shows the cumulative distribution of the estimation errors, both across all six subjects. We see better estimates of the response time when combining EEG and button response data. For example, the estimation errors of 49% of the trials are smaller than 50ms, and 75% are smaller than 100ms using our method, while only 38% of all the trials are smaller than 50ms and 58% are smaller than 100ms using the behavioral

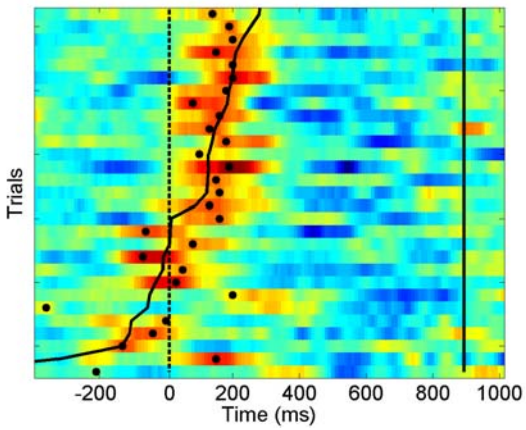

Fig. 3. One subject's response-locked single-trial discriminating component $\bar{y}$ across time. The solid black line shows the button response, the black curve shows the actual stimulus onset time, the dashed black line represents the estimate of stimulus onset time using the mean of the behavioral response time from the training data, and the black dots show the peak values of the discriminating component, which are our new estimates of the stimulus onset.

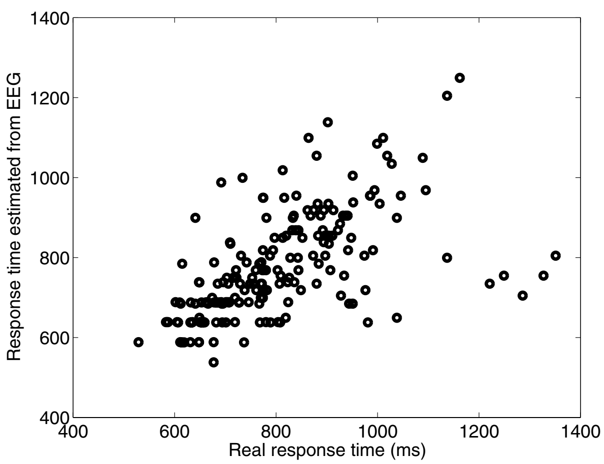

Fig. 4. The estimated response time versus the actual response time.

response as the estimate. In Fig. 5 (right) we also present the result using LD directly, with all sixty channels being the original feature space, and the time window set at 300ms, 400ms, 500ms, and 600ms, respectively. An illustration of this process is shown in Fig. 1(b). We see that EEG from 300ms time window is not a good indicator of the target onset, but 400ms, 500ms and 600ms windows can all improve the estimation performance. However, our method using MIFS, which extracts spatio-temporal features, gives the best estimate of the response time among all methods.

IV. CONCLUSION

In this paper we show that using MIFS and LD one can estimate the visual stimulus onset time during an RSVP task. Such methods have applications to brain computer interfaces, specifically image triage applications which have been termed "cortically coupled computer vision" [7].

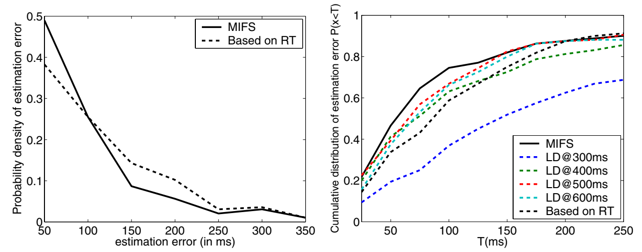

Fig. 5. Left: Probability density function of the estimation error (in ms) across subjects for our method and using response time information only. Right: Cumulative distribution of the estimation error. Cumulative distribution shows the probability of being less than or equal to T as a function of the estimation error.

REFERENCES

[1] A.D. Gerson, L. Parra, and P. Sajda, "Cortical origins of response time variability during rapid discrimination of visual objects", *NeuroImage*, Vol. 28, 2005, pp 342-353.

[2] L. Parra, C. Spence, A. Gerson, P. Sajda, "Recipes for the linear analysis of EEG", *Neuroimage*, Vol. 28, 2005 , pp 326-341.

[3] R. Battiti, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning", *IEEE Trans. Neural Networks*, Vol. 5, No. 4, 1994, pp 537-550.

[4] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, 2005, pp 1226-1238.

[5] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda, "Linear spatial integration for single-trial detection in encephalography", *NeuroImage*, Vol. 17, 2002, pp 223-230.

[6] J. Swets "ROC analysis applied to the evaluation of medical imaging techniques", *Investigative Radiology*, Vol. 14, 1979, pp 109-121.

[7] A.D. Gerson, L. Parra and P. Sajda "Cortically-coupled computer vision for rapid image search", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, in press.