35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

ERP Component Analysis for Rapid Image Searching in Finer Categories

Siyuan Xiao, Bangyu Cai, Lei Jiang, Yiweng Wang*, Weidong Chen, Xiaoxiang Zheng

Abstract—**Event-related potentials (ERP)-based image triage (or search) in the context of Rapid Serial Visual Presentation (RSVP) exploits difference in the human brain response to target and distracted stimuli in the form of an image. So far, most paradigms focus on image triage (or search) among rough object categories. In this paper, we explored the possibility and effectiveness of target detection among finer categories like different animals. We analyzed on the difference of ERP components in two image search tasks, a simple-recognition task in which all images of a target are the same and a discriminative-recognition task in which all images are randomly different but belong to the same target category (the same kind of animal). We observed that the P3 amplitude reduced and the P3 latency delayed on the discriminative-recognition condition due to the increased difficulty of identifying different images belonging to the same category. But the average area under ROC curve reached 0.82 which indicated that rapid target detection among finer categories by single-trial ERP were feasible with trivial contribution of N1 and stable contribution of N2 and P3.**

I. Introduction

In recent years, more and more attention has been paid to efficient image search through huge amounts of images in many domains. Systems fully based on computer vision have proven to be infeasible or ineffective for search tasks due to problems of limited accuracy and throughput. In contrast, human vision with its superb recognition capability, which inherently associates with contexts and semantics, makes itself attractive and promising in the application of image search. And it is electroencephalography (EEG) generated from human's scalp during human cognitive process that makes it possible to utilize human vision to develop effective image search systems. In practical use, sets of images which including target images are always displayed to human consecutively in a rapid speed, which is generally called RSVP *i.e.* rapid serial visual presentation, since this paradigm can naturally provide high throughput in a relatively short period of time. And the key point is to exploit the difference of EEG evoked by intended images and distracted images during presentation.

Event-related potentials (ERP) are certain electrical activities of human brain in response to brief visual stimuli [1]. Previous studies have certified that ERP which reflect the underlying human cognitive process can be used to indicate the emergence of certain target images in a RSVP task. For instance, an object categorization system was built to classify images containing animals, faces and inanimate objects based on EEG in [2]. Jun Wang and et al. proposed a BCI-VPM image annotation system in which objects of interest was detected first from a subset of the initial image database which includes 62 object categories (like airplane, anchor, Buddha and etc.)[3]. Huang and et al. designed an ERP-based target detection system to detect scale satellite imagery containing surface-to-air missile sites from those containing no targets[4]. It is easy to notice that the prior works mainly focus on detection among rough categories, *i.e.* categories with large difference in semantics, whereas few works have been found to tell whether it is possible to do effective detection or search among finer categories, *e.g.* categories of different animals, categories of different cars. The similar semantics between finer categories may seriously mislead human and increase the tendency to regard nontargets as targets during rapid presentation of images.

The purpose of this article is to study the possibility of target detection among finer categories and contributions of different ERP components. For the sake of simplicity and typicality, eight common animal categories were chosen for our study for ensuring not so huge difference in difficult of recognizing them. We also designed simple-recognition and discriminative-recognition tasks and the former were for comparative test. We were always interested in common visual evoked ERP components, like N1, N2 and P3, for their underlying possibility of contributing to classification. To this end, we intended to explore the difference of EEG waveforms between targets and nontargets on two different conditions by difference potentials. From the difference potentials we can find out which components can be mainly counted on for effective target detection. Classification for single trial ERP data was performed using stepwise discriminate analysis (SWDA) [5] followed by linear discriminate analysis (LDA) and AdaBoost algorithm [6] to evaluate whether the target detection for finer categories could be feasible.

**This work is supported by grants from National High Tech R&D Program of China (No. 2012AA011602), the National Basic Research Program of China (No. 2013CB329506), the National Natural Science Foundation of China (No. 61031002, 61001172, 61233015), and the Fundamental Research Funds for the Central Universities. All the authors are with Qiushi Academy for Advanced Studies, Zhejiang University.

Lei Jiang, Siyuan Xiao and Weidong Chen are also with College of Computer Science and Technology, Zhejiang University, China (e-mail: fishjianglei@gmail.com, yihan2008026@126.com, chenwd@zju.edu.cn). Bangyu Cai and Xiaoxiang Zheng are also with College of Biomedical Engineering & Instrument Science, Zhejiang University, China (e-mail: cbangyu@gmail.com, zxx667@gmail.com).

Yiwen Wang and Xiaoxiang Zheng are also with Key Laboratory of Biomedical Engineering of Ministry of Education, Zhejiang University (phone: 86-571-87952339; fax: 86-571-87952865; e-mail: eewangyw@zju.edu.cn)

II. METHOD

A. Experiment and Data acquisition

Ten volunteers including eight men and two women with a mean age of 24 years participated in this study. All participants had normal or correct-to-normal vision. Subjects were instructed to carry out a target detection task during display of a sequence of animal pictures. We had 120 experimental images for eight different animals (*e.g.* cat, frog and fish), 15 images for each. Each animal category may include animals from different subcategories (species). The images were all from Google images. Contents which are unrelated to the animal objects in all our images were got rid of in advance in case it was too difficult for subjects to recognize them. All these images were normalized into the same size (640 x 480) and illumination. In addition, all animal objects were adjusted to be in the center of images so that it would be unnecessary for subjects' eyes to move dramatically to capture the objects; hence, disturbance from electrooculography (EOG) could be decreased.

Subjects were instructed to perform discriminative- and simple-recognition tasks in disparate sessions of runs. In the discriminative-recognition condition, targets of the same category in a run were all different from each other, while in the simple-recognition condition targets in a run were all the same. Every subject will finish two sessions in each condition alternately. Of each trial, 80 images would be randomly displayed with no continuous target images, among which 12.5% contains targets while others did not. Every animal category contributed 10 images which were randomly chosen from each 15 images. Images were presented in series for durations of 500 milliseconds and another 500 milliseconds for nothing but background between every two images, *i.e.* the time interval of every two consecutive images is one second. In practical, subjects were required to click the mouse with their habit side hands after their recognition of targets as soon and exactly as possible. At the beginning of every trial, subjects stared at a fixation cross at the center of the screen.

EEG data was acquired using a NeuroScan SynAmp2 system from a standard electrode cap in which 60 electrodes inlaid at locations according with the International 10-20 system and EOG data from two electrodes placed above and below the left eye. Data was sampled at 1000Hz.

B. Data preprocessing and Data analysis

Before further processing, the recorded EEG data were preprocessed in the following steps: band-pass filtering (0.5-30 Hz), ocular artifacts reduction and baseline correct. After that, the data were broke up into epochs. Each epoch comprised a 1 second segment of EEG, 200 milliseconds before, 800 milliseconds after the onset of stimuli. The 200ms before the stimulus onset were used to correct their baseline. In order to get clean EEG signatures to analyze and compare the difference of ERP components on different conditions, we discarded all epochs if epochs recorded contained peak-to-peak amplitude exceeding 100μV, indicating remain ocular artifacts and false clicked epochs were also discarded. This accounted for 1.25% of the measurements. For both two conditions, averages of 3.41% target epochs (range 8861-9593) and 0.94% nontarget epochs (range 63759-67119) were left with us per subject. Then, the remained target and nontarget epochs were averaged on each condition, each subject and each electrode. Then, all the data were resampled to 100Hz. Finally, the averaged nontargets epochs were subtracted from the averaged target epochs for each subject, each electrode and each condition. Further analysis would be performed on the difference potentials. But when the EEG data were used for offline classification, no epochs were abandoned.

In order to measure the significance of difference ERP components evoked by intended and distracted stimuli, we applied the method of a paired t-test on the difference potentials for each electrode and condition. We used the method of Guthrie and Buchwald [7] to correct for multiple testing that at least four consecutive samples (equivalent to 40ms) [8] should be significantly different from zero in order to be considered as a segment indicating remarkable difference.

C. Offline classification

EEG data from all 60 electrodes were used for single trial classification. Here, each epoch only included 800 milliseconds length of EEG data after the onset of stimuli and was represented as a feature vector of length 4800 (electrodes * EEG samples). Each epoch were depicted by a 16*80*4800 matrix (sessions * trials * feature dimensions) for every subject on each condition. Because of the high dimensions of features, we first applied SWDA [5] on all ten subjects' EEG data to choose most useful features. The dimensions of features of an epoch were at last reduced to about 2100.In this paper , we used the AdaBoost algorithm come up with by Freund and Schapire [6] to improve the performance of the often used LDA in single trial classification [9]. Given a training set $(x_1, y_1),\ldots, (x_m, y_m)$ where x_i represents input samples, m is the size of samples, and $y_i \in \{-1, +1\}$ is the class label of each sample. AdaBoost trains a new certain base classifier h_t on each round $r = 1,\ldots, R$ in series. One of the keys of the algorithm is to preserve an ever-changing weight for every training sample. The weight of training sample i on round r is denoted $D_r(i)$. To begin with, all samples' weights are initialized equally, but wrongly classified samples' weights are increased so that the base learner is guided to pay more attention to the tough samples in the training set.

The weights are updated followed by

$$D_{r+1}(i) = \frac{D_r(i)\exp(-\alpha_r y_i h_r(x_i))}{Z_r}, \tag{1}$$

where Z_r is a normalization factor and set as

$$Z_r = \sum_i D_r(i)\exp(-\alpha_r h_r(x_i)) . \tag{2}$$

And α_r is the weight assigned to h_r for the measurement of its importance and typically set as

$$\alpha_r = \frac{1}{2}\ln\left(\frac{1-\varepsilon_r}{\varepsilon_r}\right) . \tag{3}$$

The destination of the base learner is to minimize the error

$$\varepsilon_r = \Pr_{i \sim D_r}[h_r(x_i) \neq y_i]. \quad (4)$$

After the R-th iterator ends, the final classifier we get is

$$H(x) = \text{sign}\left(\sum_{r=1}^{R} \alpha_t h_t(x)\right). \quad (5)$$

Receiver operating characteristic (ROC) analysis [10] is used to illustrate a binary classifier's performance as its discrimination threshold is varied. On the horizontal axis of a ROC curve is the rate of false positives and on the vertical axis of the curve is the rate of true positives. And we calculated the area-under-ROC curves (AUC) to measure the performance of our classification.

III. Results

A. *ERP components*

Spatiotemporal presentation of the amplitudes of the difference potentials for 60 electrodes on discriminative-recognition condition is presented in figure 1(A). Remarkable segments can be observed directly from the highlighted aggregated color blocks. In figure 1(B), spatiotemporal plot reveals that the significant remarkable segments, using the method described in section II.B. The coupled lines in red or blue indicate the start and end of the stable segments which were affirmed to be ERP components. The plot in figure 1(C) shows a glance at the grand average waveforms of the targets and nontargets on both two simple- and discriminative-recognition conditions. Modification of N2 and P3 between targets on discriminative- recognition condition and those on simple-recognition condition was strictly tested by the method described in section II.B.

1) Difference of ERPs between targets and nontargets on discriminative-recognition condition

From figure 1(B), we can easily find out that N2 and P3 are the most remarkable difference between targets and nontargets on discriminative-recognition condition (both $p<0.01$). Though the difference of early segments between targets and nontargets in figure 1(B) can be observed ($p<0.05$), segments vigorously overlapped can hardly be identified as certain kinds of components. In fact, we considered it the result of great imbalance of the number (9600 vs 67200) of the samples between them which led to different increase of SNR. In other word, the difference may reflect the noise level in collected EEG data which could even make classification results worse. This consideration was verified by the results of classification. What's more, the negligible difference of N1 between targets and nontargets on both conditions revealed that N1 also existed in the process of recognition and negation of nontargets which was associated with pattern recognition and stimulus classification [11]. Therefore, a potential trend of our rapid image searching task in finer categories was getting quite tough.

2) Modification of ERPs between two conditions

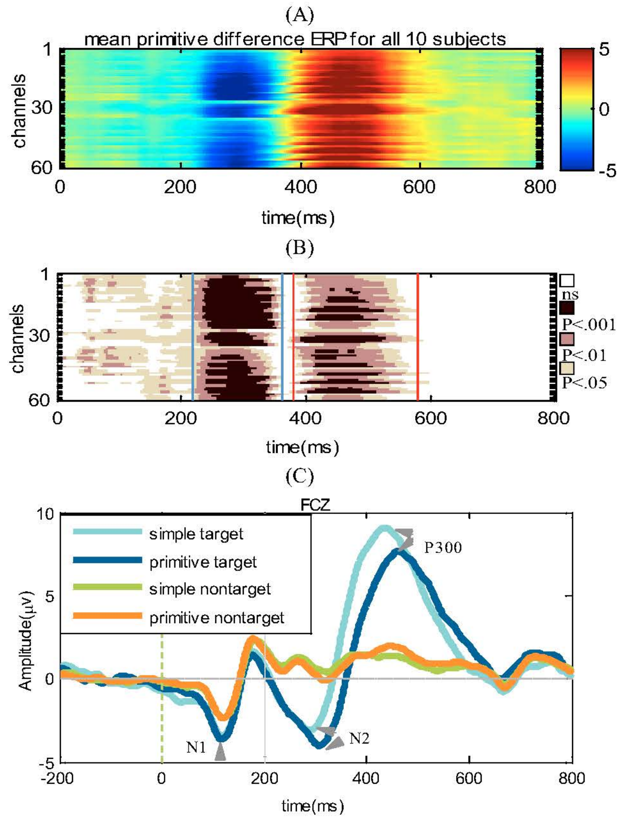

Figure 1: (A) The total average of the amplitude of the difference potentials (μv) for discriminative-recognition condition. (B)The statistical significance (p-values) plot of the difference potentials. (C) Grand average of the target and nontarget waveforms for two conditions in contrast.

Further observation for figure 1(C) implies conceivable modification between ERP components between targets on two conditions. Average amplitude and 50% area latency [12] were used to calculate the amplitude and latency of N2 and P3 of targets for each electrode, each subject and each condition. The time ranges of N2 and P3 components are derived from the above paired t-test results on two conditions while the time of N1 was derived direct from observation on EEG waveforms. Then the time windows for amplitude and latency calculation were gained according to corresponding time ranges. Finally, t-test was performed to amplitude and latency data for each electrode. The t-test analysis revealed that, compared to simple-recognition condition, amplitude of P3 reduces ($p<0.05$ for most electrodes) and latency of P3 delays ($p<0.01$ for all electrodes) in discriminative-recognition condition. These changes can be explained by the increased task difficulty in the discriminative-recognition condition [13]. What's more, amplitude and latency of N1 and amplitude of N2 have no remarkable modification between the two conditions (both $p>0.01$) on all electrodes, while latency of N2 delays on some electrodes, e.g. P3, PZ, P2 and POZ on discriminative-recognition condition compared to simple-recognition condition.

B. *Offline classification*

We performed offline classification between targets and nontargets for different epoch segments after the onset of stimuli, *i.e.* 0-800ms (N1, N2 & P3) and 200-800ms(N2 &P3). The results for 10 subjects are presented in figure 2. Five-fold cross validation was performed for each subject and the average ROC for each subject was then gained using the algorithm from [10]. The accuracy of classification for ten subjects on different epoch segments is shown in table I with precision for nontargets above 70% .The averaged area-under-ROC for 0-800ms epoch segments was 0.8242, and that for 200-800ms epoch segments was 0.8288 which was even a little bit larger. And the p value of the t-test for the two groups of AUCs was 0.4822, so the classification results between the two groups could be considered to be no big different. From figure 2(C), it is easy to notice that classification using epoch segments of 200-800ms is better for some subjects. The results indicate that N1 and the earlier components (200ms after onset) have trivial contributions to classification for finer categories and noise in the early epoch segments can even lead the results worse to some extent.

TABLE I. ACCURACY OF CLASSIFICATION FOR TEN SUBJECTS ON DIFFERENT EPOCH SEGMENTS

	S# 1	S# 2	S# 3	S# 4	S# 5	S# 6	S# 7	S# 8	S# 9	S# 10
N1,N2 & P3	0.78	0.89	0.77	0.82	0.74	0.80	0.79	0.78	0.73	0.73
N2 & P3	0.82	0.86	0.78	0.81	0.77	0.81	0.79	0.78	0.80	0.75

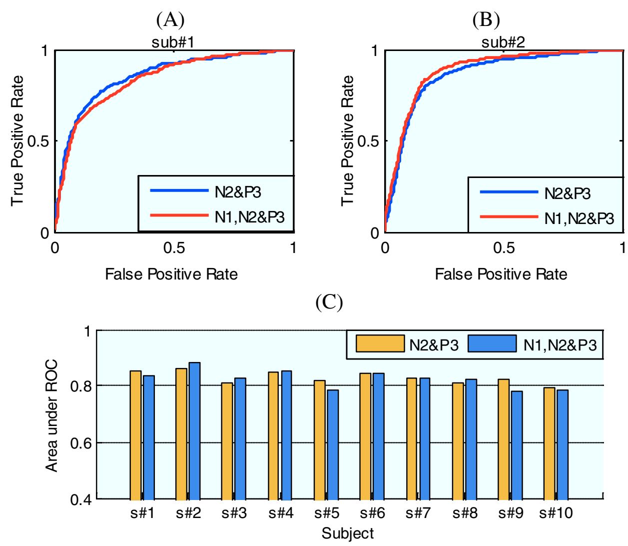

Figure 2: Area-under-ROC of classification results. (A) Subject #1's ROC curves for different epoch segments. (B) Subject #2's ROC curves for different epoch segments. (C) Average area-under-ROC of five-fold cross validation for ten subjects.

IV. Conclusion

We studied the possibility and effectiveness of target detection for finer categories in the context of RSVP. We designed the simple-recognition and discriminative-recognition tasks for evaluating the change of different ERP components. We used a paired t-test method to explore those ERP components of which difference was remarkable. We found out that N1 components existed in EEG signatures of both targets and nontargets which revealed similar semantics-associated processes of recognition and stimulus classification in early phase and led to tough image searching tasks in finer categories. In fact, the difference between them was so trivial that N1 components could be considered to be limited to contribute to the detection of target stimuli as the classification results revealed. Average area-under-ROC for 0-800ms epoch segments was 0.8242 which was slightly smaller than 0.8288 of that for 200-800ms epoch segments. The fact showed that speed of target detection for finer categories would be delayed after the latency of N1 components.

From our study, N2 and P3 were still important and stable components for the classification between targets and nontargets, though the reduction of P3 amplitude in the discriminative-recognition condition may make the classification more difficult. The results of offline classification on single trial ERP detection indicated that it was feasible to do target detection for finer categories. The future work will focus on experiments of target detection for extended more kinds of finer categories and algorithms should be improved to gain better results of target detection.

References

[1] S. Makeig, M. Weserfield, T. -P. Jung, S. Enghoff, J. Townsend, E. Courchesne and T. J. Sejnowski, "Dynamic Brain Sources of Visual Evoked Responses," Science, vol. 295, pp. 690-694, 2002.

[2] A. Kaoor, P. Shenoy and D.Tan, "Combining Brain Interfaces with Vision for Object Categorization," Computer Vision and Pattern Recognition, 2008, CVPR 2008, IEEE Conference on,pp. 1-8, IEEE, 2008.

[3] J. Wang, E. Pohlmeyer, B. Hanna, Y. –G. Jiang, P. Sajda and S. –F, Chang "Brain State Decoding for Rapid Image Retrieval," in Proceedings of the 17th ACM international conference on Multimedia, pp. 945-954, ACM, 2009.

[4] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan and K. E. Hild., "A framework for rapid visual image search using single-trial brain evoked responses," Neurocomputing 74(2011), pp. 2041-2051.

[5] E. Donchin, K. M. Spencer and R. Wijesinghe, "The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface," IEEE TRANSACTIONS ON REHABILITATION ENGINEERING, vol. 8, no. 2, JUNE 2000

[6] E. Robert and Schapire, "The boosting approach to machine learning: An overview, LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-(2003), pp. 149-172.

[7] D. Guthrie and J. S. Buchwald, "Significance Testing of Difference Potentials," Psychophysiology 28, vol. 28, no. 2, pp. 240-244.

[8] M. E Thurlings, A. -M. Brouwer, J. B. F. Van Erp, B. Blankertz and P. J. Werkhoven, "Does bimodal stimulus presentation increase ERP components usable in BCIs?" Journal of Neural Engineering, vol. 4, no. 9, 2012.

[9] Y. Huang, D. Erdogmus, S.Mathan and M. Pavel "Boosting Linear Logistic Regression for Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks", Proceeding of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3,2006.

[10] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers", Technical Report, Copyright Hewlett-Packard Company, 2003.

[11] R. Simson and H. G. Vaughan, "Event-Related Potential Correlates of Two Stages of Information Processing in Physical and Semantic Discrimination Tasks", PSYCHOPHYSIOLOGY, vol. 20, no. 2, 1983.

[12] S. J. Luck, *An Introduction to the Event-Related Potential Technique*, Cambridge, MA:MIT Press, 2005.

[13] J. Polich, "Updating P300: An integrative theory of P3a and P3b", Clinical Neurophysiology: official journal of the international Federation of Clinical Neurophysiology 118. 10 (2007): 2128.