

# Classifying four-category visual objects using multiple ERP components in single-trial ERP

Yu Qin<sup>1</sup> · Yu Zhan<sup>1</sup> · Changming Wang<sup>2</sup> · Jiakai Zhang<sup>1</sup> · Li Yao<sup>1</sup> ·  
Xiaojuan Guo<sup>1</sup> · Xia Wu<sup>1</sup> · Bin Hu<sup>1</sup>

Received: 23 July 2015 / Revised: 21 January 2016 / Accepted: 28 January 2016 / Published online: 18 February 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** Object categorization using single-trial electroencephalography (EEG) data measured while participants view images has been studied intensively. In previous studies, multiple event-related potential (ERP) components (e.g., P1, N1, P2, and P3) were used to improve the performance of object categorization of visual stimuli. In this study, we introduce a novel method that uses multiple-kernel support vector machine to fuse multiple ERP component features. We investigate whether fusing the potential complementary information of different ERP components (e.g., P1, N1, P2a, and P2b) can improve the performance of four-category visual object classification in single-trial EEGs. We also compare the classification accuracy of different ERP component fusion methods. Our experimental results indicate that the classification accuracy increases through multiple ERP fusion. Additional comparative analyses indicate that the multiple-kernel fusion method can achieve a mean classification accuracy higher than 72 %, which is substantially better than that achieved with any single ERP component feature (55.07 % for the best single ERP component, N1). We compare the classification results with those of other fusion methods and determine that the accuracy of the multiple-kernel fusion method is 5.47, 4.06, and 16.90 % higher than those of feature concatenation, feature extraction, and decision fusion, respectively. Our study shows that our multiple-kernel fusion method outperforms other fusion methods

and thus provides a means to improve the classification performance of single-trial ERPs in brain–computer interface research.

**Keywords** ERP · Visual object classification · Feature fusion · Decision fusion · Multi-kernel SVM

## Introduction

Cognitive neuroscience research has demonstrated the existence of stimulus-related neural features, such as functional magnetic resonance imaging (fMRI) and event-related potential (ERP) responses to visual stimuli. For example, N170 is a typical ERP component that appears approximately 170 ms after stimulus onset with a greater negative peak in response to a face than a non-face (Thierry et al. 2007). In the domain of language processing, the N400 and P600 reflect lexical–semantic integration and syntactic processing problems, respectively (Graben et al. 2008). In addition, compared with an unfamiliar face, a familiar face can evoke a stronger N400, which is another ERP component (Bentin and Deouell 2000). In contrast to mapping from stimulus to ERP, the development of neuroimaging techniques and brain–computer interface (BCI) research in recent years has enabled the mapping from brain activities to stimulus information, such as deciphering visual information from fMRI or ERP signals recorded when visual stimuli are presented. To characterize noninvasively the relationship between stimulus features and human brain activity, a host of studies have focused on decoding human mental states using brain activities.

In the last decade, researchers have made considerable efforts to predict the contents or categories of images using fMRI responses based on the excellent spatial resolution of

✉ Jiakai Zhang  
jiakai.zhang@bnu.edu.cn

<sup>1</sup> School of Information Science and Technology, Beijing Normal University, Beijing, China

<sup>2</sup> Beijing Anding Hospital, Beijing Institute for Brain Disorders, Capital Medical University, Beijing, China

fMRI (Kay et al. 2008; Miyawaki et al. 2008). Representing responses to objects could be described by a distinct pattern across a broad expanse of areas in the ventral temporal cortex, which suggested that discriminating multiple-category objects from complex and overlapping patterns was feasible (Song et al. 2013; Güçlü and van Gerven 2014). These fMRI studies on object discrimination inspired researchers to investigate classifying object images or prioritizing images in an image pool using ERP signals (Yu et al. 2014; Wang et al. 2012). The main advantage of EEG is its high temporal resolution, which allows studying changes in the brain's electric field over time (Schinkel et al. 2007).

In the literature on ERP-based object discrimination, many studies have focused on category-specific representations of ERP responses, particularly using the N1/N170 component, which is widely used to extract visual stimuli related to electroencephalograph (EEG) features. Shenoy and Tan (2008) successfully used distinct spatial patterns from single-trial N170 components to discriminate between faces and objects. Xu et al. (2012) determined that a category priming effect emerged at the N170 stage in face processing. Li et al. improved N1/N170 classification for face/building categorization through classifier selection. Here, unit classifiers were trained using an EEG trials group with similar pre-stimulus phases (Li et al. 2015). Follow-up studies extended the object discrimination task to animals versus inanimate object categories and mammals versus tools, showing that non-face objects could also be correctly categorized (Murphy et al. 2008b). Besides the N170 component in ERP responses, Talebi et al. (2012) determined that the positive amplitudes of ERP in 300–500 ms were considerably larger for old items than for new. Bigdely-Shamlo et al. (2008) detected satellite images containing target objects (i.e., airplanes) using P3 features in ERP recorded during a rapid serial visual presentation.

In addition to using EEG features from a single period or single ERP component, some studies have utilized EEG signals in several ERP components or several time windows of a single-trial ERP. Philiastides et al. performed a categorization task between faces and cars and extensively investigated discriminative information in two ERP components (early 170-ms and late >300-ms components) (Philiastides and Sajda 2006; Philiastides et al. 2006). Simarova et al. (2010) divided the whole ERP (0–640 ms) into 16 time intervals of 40 ms each and used information from each interval to distinguish two types of images (i.e., animals and tools). Murphy et al. (2008b) also investigated the automatic selection of optimal time intervals from a complete time course to distinguish between three concept categories (e.g., animals, plants, and tools) using a single-trial EEG. Wang et al. (2012) showed that a combination of ERP components (e.g., P1, N1, P2a, and P2b) improved the

classification accuracies of four categories (e.g., faces, buildings, animals, and cars) using the complementarity of discriminative information in different ERP components.

The fusion of different ERP components has been a promising method to improve the classification accuracy of single-trial EEG signals.

The fusion process of ERPs could be divided into three layers: data (i.e., low-level), feature (i.e., intermediate-level), and decision (i.e., high-level). Data fusion combined the data of several original ERP waveforms to produce new raw waveform data. Compared with data fusion, feature fusion managed the selection and combination of spatiotemporal features extracted from different ERPs to remove redundant and irrelevant features. Decision layer fusion provided better and unbiased results by merging a set of unit classifiers that were designed on the features from single ERP components. The feature fusion was an advancement of data fusion. Thus, recent research in ERP fusion has concentrated on decision and feature fusions for better classification accuracy.

Existing ERP feature fusion for classification could be implemented using two basic strategies: serial and parallel (Yang et al. 2003). In serial ERP fusion, all ERP feature sets were grouped together to form a longer feature. For example, an ERP response was first divided into several intervals, and the ERP waveform in each interval was averaged with the time to obtain ERP features (i.e., spatial features). These ERP features were connected to form a long feature vector (Wang et al. 2012; Blankertz et al. 2011). Wang et al. (2012) used this method to group ERP features (e.g., P1, N1, P2a, and P2b) and explored the classification results of the combined features in single-trial EEGs. Kayser and Tenke (2003) utilized another serial strategy by first grouping all EEG features throughout the ERP responses and then used principal components analysis (PCA) to select several ERP features (e.g., N1, N2, and P3b) from the complete ERP response. In parallel ERP fusion, multiple ERP feature sets were first used to form a complex feature vector space. Then, feature extraction, which included PCA and linear discriminant analysis (LDA), was performed in the complex feature space to produce a new EEG feature vector. Sajda et al. first divided ERP (e.g., 0–1000 ms) into 10 equal intervals. Second, in each interval, the researchers calculated a channel-weighted vector according to Fisher discrimination to merge multiple channels into a single waveform (using spatial fusion). Third, the researchers also averaged waveforms from all intervals according to the weight vector across intervals computed using penalized logistic regression (PLR) methods. Thus, all EEG features within 10 intervals were merged into a single waveform, which was the new EEG feature (using interval fusion) (Sajda et al. 2010). Murphy et al. (2008b) first designed an automatic system to

↳ SOA interessante per il paper

select the optimal time and frequency intervals of ERPs. These intervals were then decomposed into a new time series as features by means of a supervised transformation called common spatial subspace decomposition (CSSD).

In recent years, decision fusion, which is represented by multi-classifier or multi-expert combination strategies, has been rapidly developed and applied to EEG classification. Polikar et al. decomposed ERP into different frequency band components that were used to design multiple classifiers. The classification results were then combined through a modified weighted majority voting procedure (Polikara et al. 2008). Schels et al. partitioned EEG channels into nine overlapping areas containing as many as 18 channels at a time and extracted five time and frequency domain features from each EEG partition. The resulting 45 sets of channels (i.e., five types of features times the nine partitions) were separately trained and classified. The classification results were used jointly by averaging the results of classifiers selected with genetic algorithms (Schels et al. 2011).

In this study, we continued the work of Wang et al. (2012) and further investigated the effect of the level of information fusion in single-trial EEG classification. ERP features were extracted by averaging the waveform in the given time interval of an ERP component to form a spatial distribution of scalp potentials. After extraction, the spatial features were considered feature vectors of each ERP component for classification. We also enriched the concept of information combination and fusion by introducing a novel method called multiple-kernel fusion between feature-level and decision-level fusion (Zhang et al. 2011). Zhang et al. used multi-kernel learning (MIL) to combine measurements from all three biomarkers (i.e., MRI, PET, and CSF) to discriminate between Alzheimer's disease (AD) patients and healthy controls.

In addition to the different descriptions of ERP component fusion methods, we gave considerable attention to classification results and their comparisons. We applied different fusion methods to investigate the spatial features of ERP components (e.g., P1, N1, and P2a, P2b). These methods fused information at the feature, decision, and kernel levels. Here, we compared the four-category visual object classification results of different ERP component fusion methods. All of our experimental results indicated that classification accuracy improved by means of information fusion. In addition, the experimental results demonstrated that multiple-kernel fusion outperforms other fusion methods.

## Materials and methods

### ERP dataset

In our study, we used the dataset from Wang et al. (2012) in which ERP signals were measured while participants

viewed visual images. Eight healthy right-handed participants with normal or corrected-to-normal vision (five females, ages 22–28 years) participated in a four-category object classification task. Prior to the experiment, all participants provided written informed consent to participate.

The participants performed image viewing tasks using stimuli materials downloaded from the Internet. All stimulus images were converted to gray scale and cropped to an identical size (300 × 300 pixels) and centered in a gray background. The luminance of the stimulus images was also manually adjusted to similar values. The mean global luminance values of the images for faces, buildings, cats, and cars were 137.38, 138.32, 134.80, and 135.95, respectively. In each session, the 40 stimulus images for each category were selected randomly without duplication from the image bases.

Each participant viewed 1600 images (i.e., 400 faces, 400 buildings, 400 cats and 400 cars) during 10 sessions. During each session, 160 visual stimuli (40 images for each category) were presented to the participants in random order. The duration of each image presentation was 500 ms, which was followed by blank-screen inter-stimulus intervals (ISI) ranging from 850 to 1450 ms. Each presented stimuli image included a fixation point (a red cross) at the center of the screen. Participants were asked to concentrate on the fixation point and avoid making explicit decisions about the category information of the stimulus.

EEG signals were collected using a Brain Products system including 63 EEG channels and 1 ECG channel. The EEG data were continuously recorded with a sampling rate of 500 Hz and a down-sampling rate of 250 Hz. We selected 50 channels from the 63 EEG channels for further processing. The ECG channel and 11 EEG channels (FP1, FPz, FP2, AF7, AF3, AF4, AF8, F7, F5, F6, and F8) located in the frontal area that were susceptible to contamination from eye blinks were removed. Another two EEG channels located in the lateral temporal areas (TP9 and TP10) were also removed because of difficulty in fastening them to the scalps of some participants. The EEG data were filtered using a band-pass casual finite impulse response filter (45 orders, 0.15–40 Hz, linear phase shift) that was baseline-corrected and epoched by stimulus conditions in the MATLAB-based toolbox EEGLAB.

### Feature extraction of single-trial ERP

The EEG time courses of each trial were segmented into distinct intervals to extract ERP components such as P1 (i.e., P100), N1 (i.e., N170), P2a, and P2b (i.e., the two subcomponents of the ERP components following N1). According to Wang et al., the occipital lobe was primarily responsible for visual processing. Thus, in our study, the time window and width of the aforementioned ERP

components (subcomponents) were determined from the mean ERP across all 1600 trials in the occipital electrodes for each participant. The center of the ERP components was located in the sample time of the ERP peak. The length of the ERP time interval was fixed visual inspection. The lengths of P1, N1 (N170), P2a, and P2b were 40, 60, 70 and 80 ms, respectively.

For each ERP component, we averaged ERP signals across ERP intervals. The mean potential within each interval of component  $k$  for electrode  $i$  was described by the following:

$$x_{ki} = \left( \sum_t x_{kit} \right) / N_k, \quad t \in T_k \quad (1)$$

where  $x_{kit}$  was the electrical potential measured at time  $t$  for electrode  $i$  of the  $k$ th ERP component,  $k \in \{1, 2, 3, 4\}$ , representing P1, N1, P2a, and P2b, respectively;  $T_k$  was the interval of the  $k$ th ERP component; and  $N_k$  was the length of interval  $T_k$ . The feature vectors from a single ERP component  $k$ ,  $x_k \in \mathbb{R}^{N \times 1}$  ( $N = 50$ , which is the number of selected channels), could be considered a spatial feature representing the spatial distribution of scalp potentials within a given time interval. After the aforementioned feature extraction process, we obtained ERP feature vectors for each ERP component, which were fused using different strategies during subsequent data processing.

### Feature concatenation and feature extraction for multiple ERP components

We implemented multiple ERP feature fusions using two basic strategies in the feature layer: feature concatenation and feature extraction. For feature concatenation, we concatenated all ERP features to form long feature vectors, which can be expressed as follows:

$$y = (x_1^T, x_2^T, x_3^T, x_4^T)^T \quad (2)$$

Here,  $x_1 \sim x_4$  represented the ERP features of each component.

For feature extraction, we used class-dependent PCA in which a principal subspace was identified for each class of data independently of other classes. The feature extracted in each principal subspace is then concatenated. We obtained the final feature vectors according to the following:

$$y = (y_1^T, y_2^T, y_3^T, y_4^T)^T \quad (3)$$

Here, we extracted the class-dependent PCA features  $y_c$  from the concatenated ERP features of each category,  $c = \{1, 2, 3, 4\}$ , representing four categories. To obtain the class-dependent PCA features, we calculated the eigenvalue and eigenvector of the covariance matrix of concatenated ERP features in the four categories. The

columns of the eigenvector matrix  $A_c$  were sorted in order of decreasing eigenvalue. The ERP features in each principal subspace were expressed as follows:

$$y_c = A_c^T (x_{1c}^T, x_{2c}^T, x_{3c}^T, x_{4c}^T), \quad c = 1, 2, 3, 4 \quad (4)$$

where  $A_c$  represented the principal subspace projection matrix and  $x_{1c} \sim x_{4c}$  represented the ERP features of each component for each category subspace. To further reduce the dimension of the ERP features, we determined the number of PCs that accounted for more than 95 % of the total variability and obtained a low-dimensional feature  $y_c$ .

### Classifier design

We used SVM to classify the concatenated ERP features by combining four ERP component features, and then fused the ERP features using PCA. We used one-versus-one SVMs to distinguish the four types of objects because SVM was a binary classifier. This approach was used to design an SVM classifier between any two classes of samples and to make a decision regarding classification results.

For a binary classification, we assumed that  $N$  samples existed in the training set, in which  $y_i$  was the ERP feature vector of Sample  $i$  and  $z_i$  was a category,  $z_i \in \{-1, 1\}$ . In this study, each sample for classification corresponded to single-trial EEG data. First, linearly inseparable samples were mapped from their original space to a higher or even infinite dimensional feature space in which they were more likely to be linearly separable than in the original lower-dimensional space. This was accomplished by means of a kernel-induced implicit mapping function. Then, a maximum margin hyperplane was designed in the higher-dimensional space. According to the SVM algorithm,  $w$  and  $b$  were derived such that:

$$\min \phi(w) = w^T w + C \sum_i \xi_i \quad (5)$$

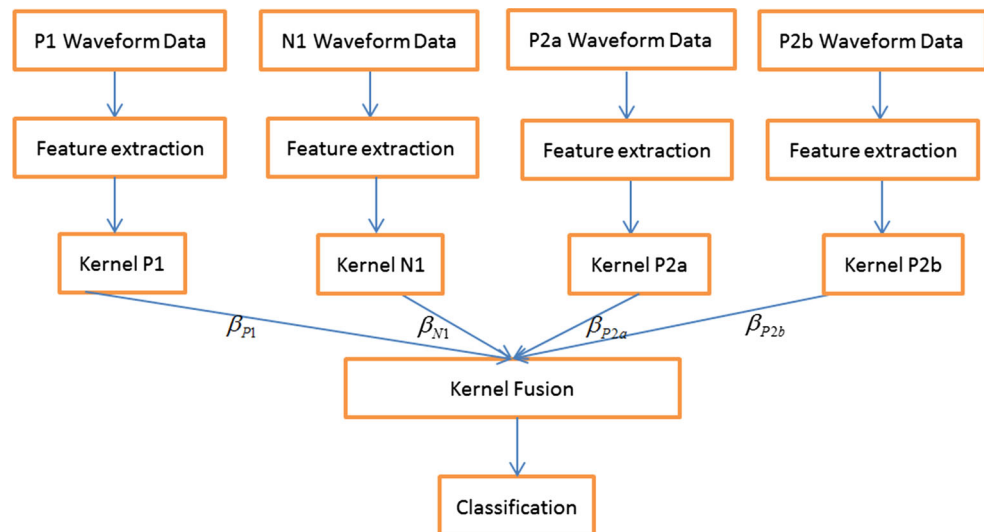
$$\begin{aligned} \text{s.t.} \quad & z_i(w^T \varphi(y_i) + b) \geq 1 - \xi_i \\ & \text{for all } (y_i, z_i), \quad i = 1 \dots N, \quad \xi_i \geq 0 \end{aligned} \quad (6)$$

where  $w$  and  $\varphi$  denote the normal vector of the hyperplane and kernel-induced mapping function, respectively. Using a Lagrange function, the problem could be converted to a simple dual problem in which  $(\alpha_1 \dots \alpha_N)$  were found such that:

$$\max Q(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j z_i z_j K(y_i, y_j) \quad (7)$$

$$\begin{aligned} (1) \quad & \sum_i \alpha_i z_i = 0 \\ (2) \quad & 0 \leq \alpha_i \leq C \quad \text{for all } \alpha_i \end{aligned}$$

$K(y_i, y_j) = \varphi(y_i)^T \varphi(y_j)$  was the kernel function for the two training samples. Thus, we chose a linear kernel. Finally,

**Fig. 1** Schematic of multiple-kernel fusion and classification pipeline

the decision function for the predicted label could be described by:

$$f(y) = \text{sgn} \left( \sum_i \alpha_i z_i K(y_i, y) + b \right) \quad (8)$$

For an unknown label sample, the final label was obtained by majority voting of each classifier.

### Multiple-kernel fusion

Multiple-kernel fusion based on the SVM could be derived from the single kernel SVM. In the training set, we defined

$$K(y_i, y_j) = \sum_k \beta_k K^{(k)}(y_i^{(k)}, y_j^{(k)}) \quad (9)$$

as a mixed kernel between the features of training samples  $y_i$  and  $y_j$ . The superscript  $k$  represented the number of kernels ( $k \in \{1, 2, 3, 4\}$ ), and  $\beta_k$  was the combined weight of each kernel, whereby:

$$\sum_k \beta_k = 1 \quad (10)$$

The decision function for the predicted label could then be obtained according to the following:

$$f(y) = \text{sgn} \left( \sum_i \alpha_i z_i \sum_k \beta_k K^{(k)}(y_i^{(k)}, y_j^{(k)}) + b \right) \quad (11)$$

The multiple-kernel fusion method combined multiple kernels into a single kernel, which differed from traditional multiple-kernel learning (Hinrichs et al. 2009). We used a coarse-grid search method from 0 to 1 with a step size of 0.1 through a fivefold cross-validation on the training samples to determine the optimal weights  $\beta_k$ . After  $\beta_k$  were obtained, we used them to combine multiple kernels into a mixed kernel, and then performed standard SVM using the mixed kernel. Figure 1 showed the classifying process using the method of multiple-kernel fusion.

As explained above, various features from different single ERP components were fused in the kernel feature layer using this kernel fusion method.

### Decision fusion

We applied multi-classifier results fusion in the layer of the decision fusion. For each of the single ERP component features,  $x_k$  and  $k \in \{1, 2, 3, 4\}$ , we designed SVM classifiers to obtain the classification results for single ERP components. For each component, we obtained the classification result  $f_k(y)$  from the respective classifier. We adopted weighted majority voting to obtain final classification results.

$$f(y) = \sum_k a_k f_k(y) \quad (12)$$

where  $a_k$  represented the weight of classification result of single-trial ERP component  $k$ . Note that  $a_k$  was determined by the error rate on the training set. For every classifier, the error rate was calculated as follows:

$$e_k = \Pr[f_k(y_n) \neq z_n] \quad (13)$$

$$a_k = \frac{1}{2} \ln \left( \frac{3/2 - e_k}{e_k} \right) \quad (14)$$

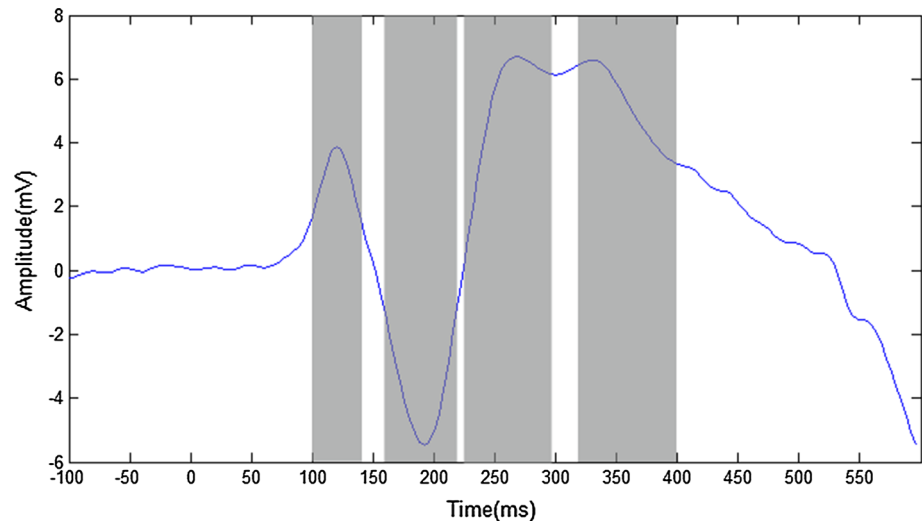
## Results

### Segmentation of ERP components

The durations of ERP components were determined from EEG responses of channels in the occipital lobe (PO8). For all eight participants, we selected the same window length for identical ERP components. The window lengths P1 and



**Fig. 2** Durations of different ERP components (*gray blocks* represent the durations of P1, N1, P2a, and P2b)



N1 were 40 and 60 ms, respectively. P2 was divided into two internally connected subcomponents, P2a and P2b, and their lengths were 70 and 80 ms, respectively. Figure 2 showed the average ERP of all eight participants and the durations of different ERP components.

### Classification results using features from single ERP components

We employed single ERP features (P1, N1, P2a, and P2b) in single-trial EEGs of all eight participants to discriminate four categories of visual objects. To obtain stable classification results at the individual level, we used leave-one-subset-out cross-validation. Each subset contained the EEG trials from every two consecutive experimental sessions and each session contained 40 trials for each class and participant. Each subset contained 80 trials for each category. For each participant, the trials from the four subsets were used for training and the rest were used as subsets for testing.

The spatial features of ERP component features obtained by averaging temporal samples within the range of the single ERP component were fed into the SVM classifiers. The classification accuracies of the trials used for testing were then averaged across participants. The left portion of Table 1 showed the classification results using features from single ERP components.

**Table 1** Mean accuracy and standard deviation across participants

Single ERP component (%)		Multiple ERP components (%)	
P1	38.90 ± 7.99	Multiple-kernel fusion	72.57 ± 6.78
N1	55.07 ± 9.45	Feature concatenation	67.10 ± 6.99
P2a	45.31 ± 5.32	Feature extraction	68.51 ± 7.30
P2b	41.76 ± 5.31	Decision fusion	55.67 ± 8.01

Table 1 showed that all four ERP components (P1, N1, P2a, and P2b) performed better than chance (i.e., 25 % for four category classifications). Paired Wilcoxon sign rank tests indicated that the difference between the classification results using each single ERP feature and chance (25 %) was significant ( $p < 0.01$ , eight pairs). N1 had a significantly higher classification accuracy than did the other components, followed by P2a and P2b. This suggested that classifiers effectively used discriminative information contained in N1.

### Classification results using features from multiple ERP components

Tables 1 and 2 showed the classification results at the individual level obtained after cross-validations using features from four ERPs (P1, N1, P2a, and P2b). The classification results were obtained by means of feature fusion using concatenation and extraction, and by means of the multiple-kernel fusion method. To further compare the effect of different layer fusion methods, we analyzed fusion results at the decision level. The decision fusion results are based on the weighted majority vote of four classifiers, each of which is designed for a single ERP component.

Table 2 showed the differences between the averaged classification accuracies across participants using different fusion methods. The classification results of a multi-kernel feature fusion boosted the classification accuracy of N1 by approximately 16 %. The feature fusion by concatenation and extraction improved N1 accuracy by 11 and 13 %, respectively. The resulting decision fusion achieved similar classification performance compared with N1. As shown in Table 2, feature fusion (feature concatenation and extraction) and multiple-kernel fusion generally yielded better classification results than did decision fusion. Another

**Table 2** Differences between classification accuracies of diverse fusion methods to combine four ERP components

	N1	Multiple-kernel fusion (%)	Feature concatenation (%)	Feature extraction (%)	Decision fusion (%)
N1	*	−17.50 <sup>a</sup>	−12.03 <sup>a</sup>	−13.44 <sup>a</sup>	−0.60
Multiple-kernel fusion	−	*	5.47 <sup>a</sup>	4.06 <sup>a</sup>	16.90 <sup>a</sup>
Feature concatenation	−	−	*	−1.41 <sup>b</sup>	11.43 <sup>a</sup>
Feature extraction	−	−	−	*	12.84 <sup>a</sup>
Decision fusion	−	−	−	−	*

Values in the table represent increases in classification accuracies corresponding to items in the first column after subtracting those corresponding to items in the first row

<sup>b</sup>  $p < 0.05$ , <sup>a</sup>  $p < 0.01$ , ‘\*’ indicates that no comparison applies, and ‘−’ indicates a repeated comparison

meaningful finding shown in Table 1 was that multiple-kernel fusion had the highest classification accuracy.

To confirm the difference between diverse fusion methods, we analyzed the differences between the aforementioned classification accuracies. Table 2 showed the results of statistical testing on improved classification accuracies. As shown, the difference between the classification results from the multiple-kernel fusion method and those of other fusion methods across all participants was significant ( $p < 0.01$ ). The multiple-kernel fusion method, which had an accuracy of 72.57 %, consistently outperformed the other methods derived from multiple components. The accuracy of the multiple-kernel fusion method was 5.47, 4.06, and 16.90 % higher than those of feature concatenation, feature extraction, and decision fusion, respectively. In addition, the difference between the serial and parallel strategies of feature layer fusion was significant ( $p < 0.05$ ). The accuracy classification of decision fusion was only 0.60 % higher than that of the best component classification (N1). However, the difference was not significant ( $p > 0.05$ ). These results confirm that kernel fusion is more effective in utilizing complementary classification information contained in different ERP components.

In addition to classification accuracy, we also compared the time consumption of different ERP fusion methods. Table 3 showed the average time (ms) across participants used to classify single-trial ERPs evoked by images in the test set. Multiple-kernel fusion consumed the least time (0.0517 ms average per visual image). The average times

used to classify single-trial ERPs for feature concatenation, feature extraction, and decision fusion were six times, two times, and fourteen times longer than that of multiple-kernel fusion, respectively. For the same type of classifier (linear SVM), the feature dimension was dramatically related to the computation involved in the classification. Here, we used the feature fusion of four ERP components as examples. The feature dimension of multiple-kernel fusion was 50. For feature extraction, the mean principal components feature number across participants was 95.04 (with a standard deviation of 7.4), which was less than the 200 components of feature concatenation.

### Kernels of multiple-kernel fusion

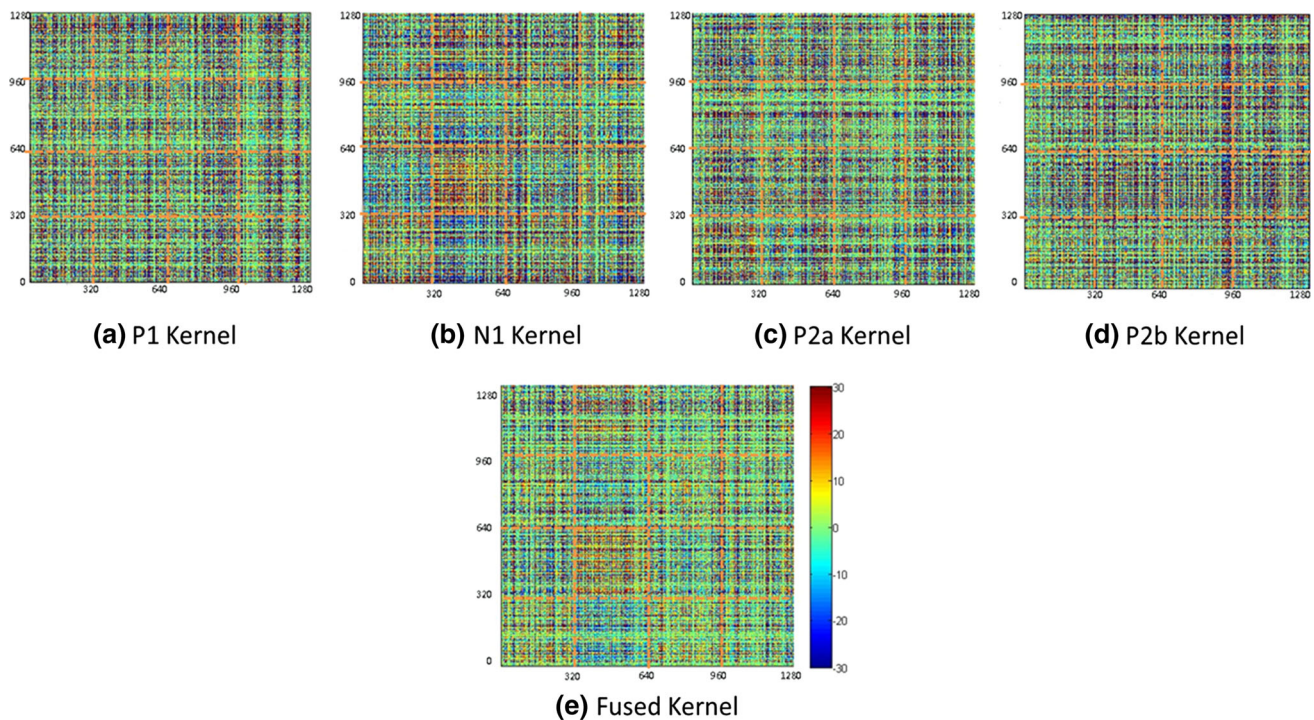
To show the classification performance of a single ERP (P1, N1, P2a, and P2b) kernel intuitively, we plotted the kernels. Figure 3 showed individual kernels of ERP components and the fused kernel trained from one participant. The kernels of the other seven participants showed similar results. Figure 3b revealed that the kernel of N1 could be used to distinguish object categories. The category difference of the kernel value was obvious and could be identified by the naked eye. The kernel corresponding to P1, P2a, and P2b did not display category information as clear as that of N1. After implementing a multiple-kernel method, we obtained the fused kernel shown in Fig. 3e. Category information was strengthened for the fused kernel. Thus, the classification effect improved while running SVM classification with a fused kernel. The complementary information that integrated different ERP components was utilized more effectively with multiple-kernel fusion.

### Weight coefficients in multiple-kernel fusion

Previous results demonstrated that P1 + N1 + P2a had good classification accuracy, and further analysis showed no significant differences between P1 + N1 + P2a and P1 + N1 + P2a + P2b. Here, we analyzed the weight

**Table 3** Mean time consumption for classification of four ERP components fusion

Methods	Time (ms)
Multiple-kernel fusion	0.0517
Feature concatenation	0.3414
Feature extraction	0.1180
Decisive fusion	0.7321

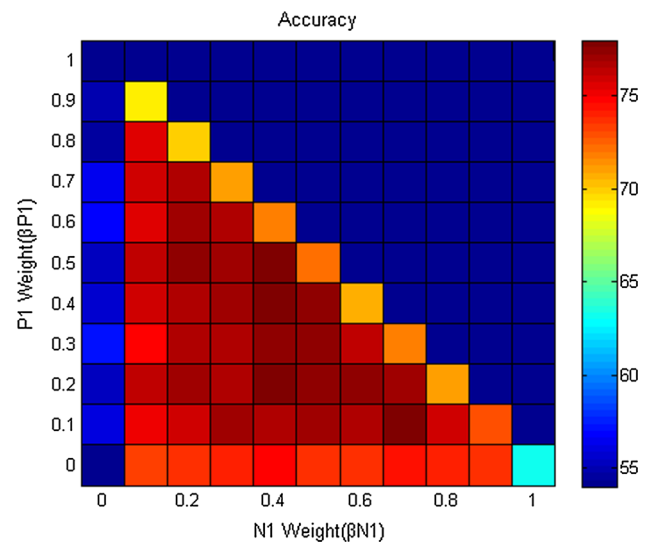


**Fig. 3** Kernels of ERP components (P1, N1, P2a, and P2b) and the fused kernel. The *horizontal* and *vertical* axes represent the samples sorted by categories in Training Set 1. Colors represent the kernel values of the pairs of samples

coefficients in multiple-kernel fusion with the ERP combination of P1 + N1 + P2a.

To investigate the effects of different combining weights (i.e.,  $\beta_{P1}$ ,  $\beta_{N1}$ , and  $\beta_{P2a}$ ) on the performance of the multiple-kernel fusion classification method, we tested all of their possible values ranging from 0 to 1 with a step interval of 0.1 under the constraint of  $\beta_{P1} + \beta_{N1} + \beta_{P2a} = 1$ . Figure 4 showed the classification results with respect to different kernel fusion weights of P1, N1, and P2a for Participant 1. The two axes represented the weights of kernels P1 and N1. The weight of kernel P2a is  $\beta_{P2a} = 1 - \beta_{P1} - \beta_{N1}$ . Note that the valid coefficients were located in the lower-left triangle matrix. Colors of the grids represented the classification accuracies using the mixed kernels with the grid coefficients. The three grids in the vertices of the triangle matrix denoted single-component classification results. The grids on the three edges (with the exception of those on the vertices) indicated two-component fusion classification results. All other grids represented the accuracies of three-component fusions.

As shown in Fig. 4, the three vertices of the lower triangle of the bottom right square (single component N1) achieved the best accuracy. The colors of the squares on the left edge of the lower triangle approximated blue, which indicated that P1 + P2a had a low accuracy that approached chance. The accuracies in the bottom edge were better than those shown at the bottom, indicating that



**Fig. 4** Classification results for Participant 1 with respect to different kernel coefficients for three components fusion (N1, P1, and P2a)

N1 and P2a had better accuracy. Nearly all squares with component N1 had better classification accuracy than the squares without N1. In addition, fusing N1 with other components resulted in higher accuracies than the N1 component alone (bottom-right vertex). The results confirmed that combining multiple ERP components with multiple-kernel fusion improved classification accuracy.



**Table 4** Results from multiple-kernel fusion and the improvement of classification accuracies by combining features from multiple ERP components

	P1 (%)	N1 (%)	P2a (%)	P2b (%)	N1 + P2a (%)	P1 + N1 + P2a (%)
P1 + N1	25.28 <sup>a</sup>	9.12 <sup>a</sup>	*	*	−2.86	−7.23 <sup>a</sup>
P1 + P2a	14.72 <sup>a</sup>	*	8.32 <sup>a</sup>	*	−13.42 <sup>a</sup>	−17.79 <sup>a</sup>
P1 + P2b	10.75 <sup>a</sup>	*	*	7.90 <sup>a</sup>	−17.39 <sup>a</sup>	−21.76 <sup>a</sup>
P1 + N1 + P2a	32.51 <sup>a</sup>	16.35 <sup>a</sup>	26.11 <sup>a</sup>	*	4.37 <sup>a</sup>	*
P1 + N1 + P2b	28.24 <sup>a</sup>	12.08 <sup>a</sup>	*	25.38 <sup>a</sup>	0.10	−4.27 <sup>a</sup>
P1 + P2a + P2b	19.92 <sup>a</sup>	*	13.52 <sup>a</sup>	17.06 <sup>a</sup>	−8.22 <sup>a</sup>	−12.59 <sup>a</sup>
P1 + N1 + P2a + P2b	33.66 <sup>a</sup>	17.50 <sup>a</sup>	27.26 <sup>a</sup>	30.80 <sup>a</sup>	5.52 <sup>a</sup>	1.15
N1 + P2a	*	11.98 <sup>a</sup>	21.74 <sup>a</sup>	*	*	−4.37 <sup>a</sup>
N1 + P2b	*	6.25 <sup>a</sup>	*	19.55 <sup>a</sup>	−5.73 <sup>a</sup>	−10.10 <sup>a</sup>
N1 + P2a + P2b	*	13.96 <sup>a</sup>	23.72 <sup>a</sup>	27.26 <sup>a</sup>	1.98 <sup>a</sup>	−2.39
P2a + P2b (P2)	*	*	7.47 <sup>a</sup>	11.01 <sup>a</sup>	−14.27 <sup>a</sup>	−18.64 <sup>a</sup>

Values represent the increases in classification accuracies corresponding to the items in the first column after subtracting those corresponding to the items in the first row

<sup>a</sup>  $p < 0.01$ , ‘\*’ means no comparison applies

Additional observations indicated that the grids with higher accuracy were clustered in the inner region of the triangle, implying that fusing complementary information was necessary for achieving effective classification. When  $\beta_{P1} = 0.2$ ,  $\beta_{N1} = 0.4$ , and  $\beta_{P2a} = 0.4$ , the highest accuracy was obtained (77.87 %).

### Classification with different ERP combinations

The classification results derived from using multiple-kernel fusion for multiple ERP (P1, N1, P2a, and P2b) features confirmed the complementarity of discriminative information from different ERP components. Our results confirm that the discriminative information contained in multiple ERP components should be integrated and fully utilized to improve object discrimination.

To verify that the discriminative information from different ERP components was complementary, we selected four components (P1, N1, P2a, and P2b) and compared the classification results with every possible ERP component combination. Table 4 shows the results of multiple-kernel fusion using different ERP combinations.

As shown in Table 4, the fusion of two ERP components (i.e., P1 + N1, P1 + P2a, P1 + P2b, N1 + P2a, N1 + P2b, and P2a + P2b) with the multiple-kernel fusion method yielded significantly higher classification accuracies than when using single feature alone (P1, N1, P2a, and P2b;  $p < 0.01$ ). Among the classification results derived from combining features from two ERP components, N1 + P2a had the highest accuracies and was considerably better than all other combinations with the exception of P1 + N1. This improvement in classification performance may be because of better complementarity for P1 and N1. In addition, three component combinations were considered. The combination of P1 + N1 + P2a or

N1 + P2a + P2b generally yielded better performance than N1 + P2a, suggesting that P1 and P2b could still provide complementary discriminative information to N1 + P2a. The combination of P1 + N1 + P2a + P2b had the highest accuracy. However, the difference was not significant compared with P1 + N1 + P2a, and P1 + N1 + P2a + P2b consumed a larger data volume.

### Discussion

#### Further discussions on classification results

Our study compared the classification accuracies of four-category objects using EEG features from single ERP components (P1, N1, P2a, and P2b) and their fusion at the feature and decision levels. Single-trial ERPs provided sufficient information to discriminate between categories, and the discrimination information in different ERP components was complementary.

In Wang et al. spatial features extracted from single ERP components were classified using Fisher discrimination and LDA. Here, we used SVM to design classifiers for single ERP features and concatenated ERP features. To fuse multiple ERP components at the feature level, we also employed multiple-kernel fusion methods that extended the application of SVM for the purpose of integrating the discrimination information of different ERP components. For four-category object classification, the results from the multiple-kernel fusion method were nearly 6 % better than those for ERP features concatenation using Fisher linear discrimination (Wang et al. 2012). Here, multiple-kernel fusion employed SVM, Wang et al. employed Fisher LDA. To eliminate the effect of different classifiers, we designed an SVM classifier using the same fusion method in Wang’s

research (feature concatenation), as previously mentioned. The classification accuracy of multiple-kernel fusion was also 4.06 % higher than that for feature concatenation.

Our results were in line with Wang et al.'s (2012) finding that N1 had the highest classification accuracies. The N1 component could discriminate between face and non-face objects more effectively in the binary classification (faces-buildings, faces-cars, faces-cats) (Thierry et al. 2007; Han et al. 2015). In addition, no other components could be used to distinguish between particular visual stimuli objects with greater accuracy. Another notable finding was that the classification accuracy of multiple ERPs was superior to that of single ERP components (Wang et al. 2012), which was also in line with previous studies (Yang et al. 2003; Blankertz et al. 2011; Kayser and Tenke 2003; Sajda et al. 2010; Murphy et al. 2008a). Our results further validated the complementarity of discriminative information from different ERP components.

We sought an effective method to fuse the complementary information of different ERP components. The two fusion methods (feature concatenation and extraction) achieved higher classification accuracy at the feature level compared with single ERP components and decision fusion. Feature extraction reduced feature dimensions and had slightly better classification accuracy while requiring less computation.

Multiple-kernel fusion outperformed the other methods (feature concatenation, feature extraction, and decision fusion). Complementary information was utilized more effectively by kernel fusion than fusion at the feature (feature concatenation and extraction) and decision levels.

The decision fusion results were less favorable than those derived from feature fusion. This might be because the classification accuracy for the single P1 and P2a components was worse than for N1. Thus, classification using P1 and P2a could not provide sufficient supplementary information for N1 in decision fusion.

### Diversity of single ERP components in classification

Complementary discrimination information between different ERP components has been studied extensively (Wang et al. 2012; Yang et al. 2003; Blankertz et al. 2011; Kayser and Tenke 2003; Sajda et al. 2010; Murphy et al. 2008a). In addition to the complementarity of discriminative information from different ERPs, another question arose regarding information redundancy (i.e., whether information from a specific ERP component can be derived from other ERP components). If redundancy was present among the ERP components P1, N1, P2a, and P2b, the discriminative information contained in these ERPs was not independent. Thus, the best classification accuracy

**Table 5** Jaccard similarity coefficient and kappa index of each component by comparing their single classification results through fivefold cross-validation

	Kappa index	Jaccard coefficient
P1 vs. N1	0.1921	0.3117
P1 vs. P2a	0.1534	0.3124
P1 vs. P2b	0.1384	0.3087
N1 vs. P2a	0.2671	0.3428
N1 vs. P2b	0.2630	0.3781
P2a vs. P2b	0.3558	0.5010

could be achieved with only a subset of the ERP components investigated. Therefore, we used the Jaccard similarity coefficient and Kappa index to measure similarities and differences between each component by comparing their classification results.

As shown in Table 5, the Jaccard similarity coefficient and Kappa index of P2a and P2b (0.3558 and 0.5010) was larger than other combinations, which indicated strong similarity between P2a and P2b for four-category classification. These results explained the insignificant difference between the four-component combination (P1 + N1 + P2a + P2b) and the best three component combination (P1 + N1 + P2a). Although P2b alone had effective classification accuracy, it contained redundant information from P2a, thus contributing little complementary discrimination information for P2a.

Zhang et al. (2011) noted that for two components, a smaller Jaccard coefficient and Kappa index indicated greater complementary information. However, we found that the classification of multiple ERP components relied not only on differences between ERP components but also on their independent discrimination ability. For example, for two weak ERP components (P1 and P2b), the low classification accuracies of P1 and P2b resulted in a low classification for P1 + P2b despite the fact that the Kappa indices of P1 and P2b were small.

### Advantages of multiple-kernel fusion

We used a new ERP component fusion method called multiple-kernel fusion to discriminate four-category visual objects. The multiple-kernel fusion method improved the classification accuracy of single-trial ERPs. Compared with other fusion methods, including feature and decision fusion, the multiple-kernel fusion method could most effectively utilize complimentary classification information among ERP components.

Compared with the feature-level and decision-level fusion methods, the kernel-level fusion method had the following advantages. First, different ERP components

could contribute unequally to classification by using different weights. Second, this method better utilized implicit complementary information by fusing the ERP components. Multiple-kernel fusion fused information in the middle level between feature and decision levels. The feature produced from different ERPs was input into a fusion kernel. This differed from feature fusion by concatenation and extraction. The multiple-kernel fusion method differed from decision fusion because each ERP component feature vector was not used alone to classify a single-trial ERP. All kernels with different ERP component features were used to complete the classification. In addition, feature dimensions in each kernel were reduced by kernel fusion, and the time consumption was also reduced. Our results indicated that multiple-kernel fusion is a promising method for real-time BCI classification system.

Because the optimal weight of each kernel was determined by traversal searching during the training of the SVM using multiple-kernel fusion, the multiple-kernel fusion method was somewhat more time-consuming than the other methods. An increased number of ERP components substantially increased the time required. Future research should be directed toward the development of a more effective approach with the aim of identifying an effective weight for each kernel.

**Acknowledgments** This work is supported by the National High-Tech R&D Program (863 Program) under Grant No. 2012AA011603. This work is also supported by the NSFC Key Program (91320201), General Program (61375116, 61473044), Young Scientist Fund of NSFC (31300924), and the Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (SKLNST-2013-1-03).

## References

- Bentin S, Deouell LY (2000) Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognit Neuropsychol* 17:35–54
- Bigdely-Shamlo N, Vankov A, Ramirez RR, Makeig S (2008) Brain activity-based image classification from rapid serial visual presentation. *IEEE Trans Neural Syst Rehabil Eng* 16(5):432–441
- Blankertz B, Lemm S, Treder M, Haufe S, Müller K-R (2011) Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56(2):814–825
- Graben P, Gerth S, Vasishth S (2008) Towards dynamical system models of language-related brain potentials. *Cogn Neurodyn* 2(3):229–255
- Güçlü U, van Gerven MAJ (2014) Unsupervised feature learning improves prediction of human brain activity in response to natural images. *PLoS Comput Biol* 10(8):e1003724
- Hinrichs C, Singh V, Xu G, Johnson S (2009) MKL for robust multimodality AD classification. *Med Image Comput Comput Interv* 12:786–794
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452:352–355
- Kayser J, Tenke CE (2003) Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clin Neurophysiol* 114:2307–2325
- Li H, Zhang L, Zhang J, Wang C, Yao L, Wu X, Guo X (2015) Improving N1 classification by grouping EEG trials with phases of pre-stimulus EEG oscillations. *Cogn Neurodyn* 9(2):103–112
- Miyawaki Y, Uchida H, Yamashita O, Sato M-A, Morito Y (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60:915–929
- Murphy B, Dalponte M, Poesio M, Bruzzone L (2008a) Distinguishing concept categories from single-trial electrophysiological activity. *CogSci*. <https://clic.cimcc.unim.it>
- Murphy B, Dalponte M, Poesio M, Bruzzone L (2008b) Distinguishing concept categories from single-trial electrophysiological activity. In: *Proceedings on Annual Meeting of the Cognitive Science Society*, pp 403–408
- Philiastides M, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb Cortex* 16:509–518
- Philiastides M, Ratcliff R, Sajda P (2006) Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J Neurosci* 26(35):8965–8975
- Polikara R, Topalisa A, Parikha D et al (2008) An ensemble based data fusion approach for early diagnosis of Alzheimer's disease. *Inf Fusion* 9(1):83–95
- Sajda P, Pohlmeier E, Wang J et al (2010) In a blink of an eye and a switch of a transistor: cortically coupled computer vision. *Proc IEEE* 98(3):462–478
- Schels M, Scherer S, Glodek M, Kestler HA, Palm G, Schwenker F (2011) On the discovery of events in EEG data utilizing information fusion. *Comput Stat* 28:1–14
- Schinkel S, Marwan N, Kurths J (2007) Order patterns recurrence plots in the analysis of ERP data. *Cogn Neurodyn* 1(4):317–325
- Shenoy P, Tan D (2008) Human-aided computing: utilizing implicit human processing to classify images. In: *Proceedings of the Conference on Human Factors in Computing System (ACM SIGCHI 2008)*, pp 845–854
- Simanova I, van Gerven M, Oostenveld R, Hagoort P (2010) Identifying object categories from event-related EEG toward decoding of conceptual representations. *PLoS ONE* 5:e14465
- Song S, Ma X, Zhan Y, Zhan Z, Yao L, Zhang J (2013) Bayesian reconstruction of multiscale local contrast images from brain activity. *J Neurosci Methods* 220:39–45
- Talebi N, Nasrabadi AM, Curran T (2012) Investigation of changes in EEG complexity during memory retrieval: the effect of midazolam. *Cogn Neurodyn* 6(6):537–546
- Thierry G, Martin CD, Downing P, Pegna AJ (2007) Controlling for inter stimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci* 10:505–511
- Wang C, Xiong S, Hu X, Yao L, Zhang J (2012) Combining features from ERP components in single-trial EEG for discriminating four-category visual objects. *J Neural Eng* 9(5):056013
- Xu M, Lauwereyns J, Iramina K (2012) Dissociation of category versus item priming in face processing: an event-related potential study. *Cogn Neurodyn* 6(2):155–167
- Yang J, Yang JY, Zhang D, Lu JF (2003) Feature fusion: parallel strategy vs. serial strategy. *Pattern Recogn* 36:1369–1381
- Yu K, Al-Nashash H, Thakor N, Li X (2014) The analytic bilinear discrimination of single-trial EEG signals in rapid image triage. *PLoS ONE* 9(6):e100097
- Zhang D, Wang Y, Zhou L, Yuan H, Shen D (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55(3):856–867