Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 3342 – 3354

doi:10.3906/elk-1608-130

Classification of EEG signals of familiar and unfamiliar face stimuli exploiting most discriminative channels

Abdurrahman ÖZBEYAZ[1,*], **Sami ARICA**[2]
[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Adıyaman University, Adıyaman, Turkey
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering, Çukurova University, Adana, Turkey

Received: 11.08.2016 • **Accepted/Published Online:** 08.01.2017 • **Final Version:** 30.07.2017

Abstract: The objective of the study is to classify electroencephalogram signals recorded in a familiar and unfamiliar face recognition experiment. Frontal views of familiar and unfamiliar face images were shown to 10 volunteers in different sessions. In contrast to previous studies, no marker button was used during the experiment. Participants had to decide whether the displayed face was familiar or unfamiliar at the instant of stimulus presentation. The signals were analyzed in the preprocessing, channel selection, feature extraction, and classification stages. The novel two-feature extraction and eight-channel selection methods were applied to the analyses. Sixteen classification results were compared and the best performance was investigated. Consequently, the highest average classification accuracy was obtained at 72.67% when piecewise constant modeling feature extraction and relative entropy channel selection methods were used.

Key words: Electroencephalogram, evoked potential, familiar and unfamiliar face stimuli, feature extraction, channel selection, classification

1. Introduction

Electroencephalography (EEG) is a noninvasive brain activity-reading technique. According to this technique, the accumulation of electrical activities within a group of neurons of the brain is recorded with an electrode located on the scalp. Event-related potentials (ERPs) are elicited against the specific stimulus during EEG recordings. Because the behaviors of these potentials vary according to different stimuli, they contain information about the stimuli [1].

The human face contains information about the person's age, sex, identity, and emotional state [2]. Investigations associated with the neuropsychological human face perception system date back to the 1970s [3–6]. The human face perception system in the brain and its correspondence to familiar and unfamiliar stimuli is an important research topic due to its contribution to both mental illness research and forensic inspections. It has been reported in the literature that some brain regions related to face perception are crucially activated [4,7]. Familiar and unfamiliar face stimuli were presented to subjects in a study, and the activated brain regions were compared for two classes. The results showed that some brain regions were activated by familiar face stimuli, including the prefrontal, lateral temporal, and medial temporal (hippocampal and parahippocampal) regions [8]. In another study, the frontal and temporal regions in the skull were related to personal identification. In this study, it was also observed that different activations, associated with familiar and unfamiliar face stimuli,

*Correspondence: aozbeyaz@adiyaman.edu.tr

emerged in the middle occipital gyrus, right inferotemporal cortex, and right posterior fusiform gyrus regions [9]. Activations related to familiar names and external factors were also observed in these regions [10]. In another study, familiar face signals significantly diverged from unfamiliar face signals in the middle lateral fusiform gyrus and inferior occipital gyrus areas [11]. A positive peak occurred at about 600 ms after the onset of visual stimuli and continued for a few hundred milliseconds. This activation emerged in the right central parietal area of the skull [12].

Regions related to face perception in the skull can be localized from EEG and we can provide some information about previously viewed faces by classifying EEG signals responsive to familiar and unfamiliar faces. Thus, the classification process provides information about some disorders originating in the brain and about criminals in a judicial investigation. These are the motivations for classifying EEG signals corresponding to familiar and unfamiliar face stimuli.

There have been many studies on the classification of EEG signals. In several studies, familiar and unfamiliar face stimuli have been classified utilizing the evoked potentials of stimuli and statistical analysis. For instance, Tanaka et al. studied ERPs corresponding to preexisting acquired face familiarity [13]. Sun et al. used a directed lying task and computed ERPs to explore the differentiation between identification and classification of familiar faces [14]. Other studies are relevant to brain–computer or brain–machine interfacing build employing ERPs of stimuli. For example, Jin et al. studied ERP-based brain–computer interface, utilizing facial expression changes and multiple faces [15,16]. Kaufmann et al. used flashing characters accompanied by famous faces to build an ERP-based brain–computer interface [17].

However, these studies do not involve the classification of familiar and unfamiliar face stimuli from the corresponding EEG data. In [18], the authors classified EEG signals recorded in a familiar and unfamiliar face recognition experiment. In this experiment, subjects pressed a button to mark that the displayed face was familiar. A custom wavelet was generated and used to obtain wavelet transform of the EEG data. Features were extracted in the transform domain. The classification and performances were then compared with custom and common wavelets in the paper. To the best of our knowledge, there are no studies related to familiar and unfamiliar classification incorporating machine learning in the literature.

Our study is a further attempt to classify the familiar and unfamiliar face stimuli from EEG. We compute the distance between the ERPs of two channels corresponding to two respective categories. The most distant channels are employed to classify single trial EEG. This approach is different from the method in [18]. In our study, EEG signals are processed in preprocessing, feature extraction, channel selection, and classification stages. To extract features from the data, piecewise constant and piecewise linear approximation (in other words, modeling) of EEG signal are used. Six different channel selection methods are employed to select suitable channels. Two different approaches are chosen to apply the channel selection methods. The first is to identify the channels with the highest variance in the rank for a class while having the lowest variance for the other class in the rank. Variance ratio approach and common spatial pattern technique are used in this category. The other is to measure the similarities and distances among the ERPs of the familiar and unfamiliar classes of channels. Fisher score, r-correlation coefficient, mutual information, and relative entropy methods are employed in this category. A support vector machine is used to classify the familiar and unfamiliar face stimuli. Such an analysis is thought to contribute to the process of classification of EEG signals that are responsive to familiar and unfamiliar classes.

In the next section, we describe the collection of EEG data and the experimental procedure. In the third section, we describe the methods, and in the fourth section we discuss the EEG signal-processing stages. In the fifth section, we provide and compare the results obtained using the piecewise constant, linear modeling, and six-channel selection methods. We conclude the study in the final section.

2. Materials

2.1. Subjects

Ten healthy volunteers, 1 female and 9 male, aged from 22 to 52 years with an average age of 34 years and a standard deviation of 9.87, took part in this study. All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders, and all were right-handed. Four subjects wore glasses and all were nonmedicated during the experiment. All participants signed an informed consent form. The experiments were approved by the local ethics committee of the Health Sciences Institute of Adıyaman University.

2.2. Stimuli

One hundred and twenty familiar and unfamiliar face images were used in the study. The familiar group contained 61 well-known personalities collected from the public domain of the World Wide Web and 59 unfamiliar faces obtained from a model agency. We chose faces with frontal views that were alike in general visual appearance and attractiveness. All images were subjectively equalized for luminance and contrast. The stimuli were displayed with custom software (stimuli display program) developed in Borland Delphi 7 on an LCD monitor.

2.3. Experimental procedure

Subjects were seated in a comfortable chair in front of an LCD monitor at about 100 cm from a computer screen controlled by a laptop computer. In contrast to previous studies, no marker button was used in the experiment since it might create a variance in EEG. Participants had to decide whether the displayed face stimulus was familiar or unfamiliar at the instant of the stimulus presentation. All participants declared which face was familiar by marking the printed faces on a form at the end of the experiment. The time course of the experiment was organized as follows: at the beginning of the experiment, a black and white checker was shown for 5 s, and then an 'Experiment begins' warning was displayed for 60 s on the screen to indicate the start of the session. Following the stimulus shown for 800 ms, a gray screen appeared for 1200 ms. Next, a gray fixation was displayed for 2500 ms to enable the subjects to prepare and concentrate on the upcoming trial (Figure 1). During the experiments, familiar and unfamiliar face stimuli images were randomly exhibited. Moreover, because epochs belonging to stimuli were extracted, the display order of the face images was also saved for each subject during the course of the experiment.

2.4. Data

The EEG recordings were achieved with a wireless Emotiv EPOC research EEG neuroheadset. The headset consisted of 14 channels of the extended electrode placement system: AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, and O2, plus common mode sense (CMS)/driven right leg (DRL) references at P3/P4 locations. The 50-Hz power line interference was suppressed with an analog notch filter of the instrument before sampling. A 16-bit analog-to-digital converter with 128-Hz sampling rate equipped in the device was used for converting the analog signal to digital. After the EEG data were acquired for 14 subjects, all data were examined visually to identify corrupted records, and four subjects were removed from the data and excluded from the analyses.

In order to classify EEG signals acquired during the familiar/unfamiliar face recognition experiments, first the EEG signals were enhanced, and then channel selection, feature extraction, and classification processes

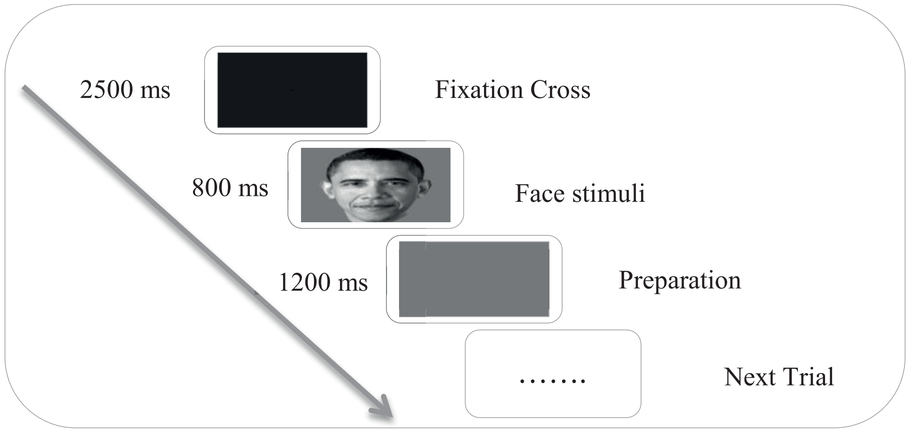

Figure 1. Time course of the experiment on familiar/unfamiliar face recognition.

followed. The algorithms were implemented using MATLAB release R2010a on a personal computer (equipped with an Intel Core i7-3520M CPU@2.90 GHz and 12 GB RAM).

3. Methods

3.1. Preprocessing

A first-order 0.16-Hz digital high-pass Butterworth IIR filter was applied in the forward direction to eliminate the baseline of the EEG. Subsequently, a linear-phase low-pass digital FIR filter was used to remove the background signal (noise) from the EEG signals. The order of the filter was chosen as 10 and the bandwidth was determined as 32 Hz. The 'butter' and 'fir1' functions of MATLAB were implemented to design high-pass and low-pass filters, respectively.

Furthermore, the spatial correlation between channels was reduced by applying a type-II discrete cosine transform to the data in the spatial domain before the feature extraction process.

3.2. Feature extraction

In the feature extraction stage, we used piecewise constant and piecewise linear modeling techniques. Initially, the trials of EEG data were extracted; the time interval of [0.6, 1.6] s from stimulus (in terms of sample indices, the interval was [78, 205] - 128 samples - 1 s) was employed. To obtain the models of a single-trial EEG signal, it was initially partitioned into segments. The length of the segment chosen for four samples amounts to a time interval of 1/32 s (31.25 ms). The window length was decided empirically. In the case of piecewise constant modeling (PCM), the averages of the segments were accepted as features. In the case of piecewise linear modeling (PLM), the slopes of the first-order polynomial approximation of segments were acquired as features.

To describe the procedure, consider that a segment (window) is the sequence $\{y_t\}_{t=0,1,\ldots,N-1}$. Each segment is modeled as follows:

$$y = a, \qquad a = \frac{1}{N}\sum_{t=0}^{N-1} y_t, \tag{1}$$

$$y = b.t + c, \qquad b = \frac{\sum_{t=0}^{N-1} t \cdot y_t}{\sum_{t=0}^{N-1} t^2}, \qquad c = \frac{\sum_{t=0}^{N-1} y_t - b\sum_{t=0}^{N-1} t}{N}, \tag{2}$$

where the first equation is the constant modeling and the second equation is the linear modeling of the segment. Since the segment length was four, each channel provided 32 coefficients. The coefficients coming from each channel (the four most discriminative channels) were concatenated to feed the classifier.

3.3. Channel selection

In the channel selection stage, we employed six different channel selection methods. These were Pearson correlation (r-value), Fisher score (FS), mutual information (MI), relative entropy (KL), variance ratio (VR), and common spatial pattern (CSP). VR and CSP were employed for the channels where variances of EEG were the highest in the rank for a class and consequently lowest in the rank for the other class. The two distances, FS and KL, and the two similarity measures, r-value and MI, were computed between event-related potentials of the classes. The evoked potential of a class is obtained by averaging the training data over trials (ensemble averaging). The channel pairs with the lowest correlation and mutual information were accepted, whereas the channel pairs with the highest Fisher score and KL distance were considered discriminative. The four selected channels (two pairs) were employed for classifying the data, and the results were compared in terms of their performances in classification.

3.3.1. Pearson's correlation (r value)

The most commonly used measure of association is Pearson's product-moment correlation coefficient, often denoted r. This value is a measure of the linear trend between two variables. The value of r will always lie between –1 and 1. In this study, $r = -1$ shows that the two variables are completely distant, and $r = 1$ indicates that the groups are identical (closest). The r value is computed for two quantities X and Y by using Eq. (3) [19]:

$$r = s_{XY}^2 / s_X s_Y, \tag{3}$$

where s_{XY}^2 is the sample covariance, and s_X and s_Y are the sample standard deviations of X and Y, respectively.

3.3.2. Fisher distance

The Fisher distance or FS is the normalized level of discrimination between two classes. Fisher criteria can be calculated as in the following equations:

$$FS = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2} \qquad m_k = \frac{1}{N_k}\sum_{x \in D_k} x \qquad s_k^2 = \frac{1}{N_k - 1}\sum_{x \in D_k} (x - m_k)^2 \ . \tag{4}$$

In the above equations, m_k and s_k are sample average and variance belonging to the kth class ($k = 1, 2$) respectively [20].

3.3.3. Mutual information

Mutual information is a measure of how much information one random variable contains about another. The mutual information of two random variables X and Y is defined by Eq. (5) [21]:

$$I(X;Y) = \sum_{x,y} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}, \tag{5}$$

where $p_X(x)$ and $p_Y(y)$ are the probability densities of the random variables. $p_{XY}(xy)$ is the joint probability density function. To compute the densities, the number of quantization levels is chosen as eight.

3.3.4. Relative entropy

Let $p_X(x)$ and $p_Y(y)$ be two probability density functions of the X and Y variables, respectively. Relative entropy between $p_X(x)$ and $p_Y(y)$ functions is expressed as in Eq. (6) [22]:

$$KL\left(X;Y\right) = \sum_{x} p_X\left(x\right) \log\left(\frac{p_X\left(x\right)}{p_Y\left(x\right)}\right). \tag{6}$$

Alternatively, the time-varying quantized signals are scaled with their sums, and the resultant sequences are accepted as probability densities. The relative entropy computed with these probability densities is named KL2. The number of quantization levels is chosen as eight, as it is for mutual information. Furthermore, the levels are specified to be greater or equal to zero to calculate KL2.

3.3.5. Variance ratio

The variance ratio is the ratio of the average powers of two variables. Suppose that C_X and C_Y denote covariance matrixes belonging to two variables X and Y, respectively. Furthermore, assume that λ_i and τ_j are the ith and jth eigenvalues of covariance matrices of C_X and C_Y, respectively. Then the variance ratio between the ith and jth channels is obtained as in Eq. (7):

$$VR\left(i,j\right) = \frac{\lambda_i}{\tau_j}. \tag{7}$$

Another version of the variance ratio is the ratio of normalized eigenvalues:

$$VR2\left(i,j\right) = \frac{\lambda_i / \left(\lambda_i + \tau_i\right)}{\tau_j / \left(\lambda_j + \tau_j\right)}. \tag{8}$$

The channels with the highest and lowest variance ratio are assumed to be discriminative. This approach is similar to the common spatial pattern filtering method. Both forms of variance ratio have been used in the classification algorithm.

3.3.6. Common spatial pattern

This method amounts to maximizing the variance of the (spatially) filtered signal under one condition while minimizing it for the opposite condition [23]. Let C_a and C_b express covariance matrixes belonging to two classes a and b, respectively, and C_c is the common covariance matrix:

$$C_a = \frac{1}{K_a} \sum_{k=1}^{K_a} \frac{E_k E_k^T}{\mathrm{trace}\left(E_k E_k^T\right)}, \qquad C_b = \frac{1}{K_b} \sum_{k=1}^{K_b} \frac{E_k E_k^T}{\mathrm{trace}\left(E_k E_k^T\right)}, \qquad C_c = C_a + C_b\ , \tag{9}$$

where E_k is the collection of EEG signals of the kth trial, in which a row holds the EEG signal of a channel and T denotes transposition. The whitening transformation matrix of the common covariance matrix is simply computed by $P = D^{-1/2}U^{-1}$. Diagonal matrix D and unitary matrix U are obtained by eigenvalue decomposition of the common covariance matrix $C_c = UDU^T$. If the common unitary matrix of the eigenvalue decomposition of matrices PC_aP^T and PC_aP^T is B, then the projection matrix (bank of filters; a row is a filter) W and the projected matrix Z_k are given by Eq. (10):

$$W = P^T B, Z_k = W^T E_k. \tag{10}$$

3.4. Classification

We employ a linear support vector machine (SVM) in this study. The linear SVM model is produced based on the training data. To describe it, let a training set of instance label pairs be given as $(x_i y_i)$ with the class labels $y_i \in \{-1, 1\}$. A sample x_i belongs to class $y_i = -1$ if it satisfies $w^T x_i + b \leq -1$, and it belongs to class $y_i = 1$ if it satisfies $w^T x_i + b \geq 1$. The separating function (hyperplane) is $w^T x + b = 0$, where w is the coefficient vector and b is the offset of the separating hyperplane. The linear support vector machine maximizes margin $2/\|w\|$ between support vectors (boundaries of two classes): $w^T x_i + b = 1$ for the class labeled $y_i = 1$ and $w^T x_i + b = -1$ for the class labeled $y_i = -1$. Then the SVM is formulated by the following optimization problem given in Eq. (11) [24]:

$$\max_w (2/\|w\|) \quad \text{subject to,} \quad y_i(w^T x_i + b) \geq 1. \tag{11}$$

The features belonging to the selected four channels were concatenated sequentially before feeding into the classifier. In order to evaluate the classification performance and validate the procedure, EEG data were initially separated into two sets at this stage. Seventy-five percent of the data was designated for training and the remaining 25% was assigned to testing. The testing data were classified using the trained machine and the results were compared to the correct labels. Thus, classification success was calculated. The same process was repeated 20 times for each subject. The training and testing data sets were always specified randomly. An average of 20 classification results were accepted as classification success for each subject. This validation technique is known as Monte Carlo cross-validation [20]. The chance level for each subject was also obtained with 100-fold Monte Carlo cross-validation. Each time the class labels were assigned randomly, an arbitrarily chosen 75% of data trained the classifier, and the accuracy of the classifier was attained from the remaining 25% of data. The average accuracy yielded the chance level.

4. Results

The classification accuracies are summarized in Table 1. The eleventh and twelfth rows are mean value and standard deviation of the results, respectively, and the thirteenth row shows chance levels of classification for each distance/similarity metric. It is observed that the chance levels are almost 50%, as expected for a two-class case, and the average classification accuracies are well above the chance levels. From the eleventh row of the table, it is deduced that the highest average accuracy of the subjects is 72.67% with a standard deviation of 9.71%, achieved by using PCM features and KL2 channel selection measure.

In addition to accuracy, specificity (Qn), which is the measure of the number of detected unfamiliar face stimuli among unfamiliar face stimuli, and sensitivity (Qd), which is the measure of the number of detected

Table 1. SVM classification accuracies (%) with PCM and PLM methods.

	r correlation coefficient		Fisher score		Mutual information		Kullback–Leibler		Kullback–Leibler 2		Variance ratio		Variance ratio 2		Common spatial pattern	
Sbj	PCM	PLM	PCM	PLM	PCM	PLM	PCM	PLM	PCM		PCM	PLM	PCM	PLM	PCM	PLM
1	69.83	70.34	69.31	61.72	70.86	62.41	68.97	62.59	76.72	70.86	71.72	68.79	63.28	59.48	73.10	69.83
2	92.00	88.50	91.17	89.00	90.00	89.50	92.00	91.33	86.17	82.33	83.83	84.50	89.83	89.50	92.00	90.00
3	65.69	65.34	69.83	68.79	68.28	69.48	66.72	67.07	70.69	73.10	57.24	57.24	55.17	56.55	60.86	59.66
4	61.00	60.50	63.50	59.50	62.50	60.67	64.50	60.83	63.00	63.50	62.17	55.00	63.00	58.17	64.67	60.17
5	90.83	79.67	88.33	86.83	88.17	84.17	88.17	83.67	89.33	83.33	82.83	80.67	71.50	70.33	80.17	77.50
6	59.17	68.83	59.00	60.33	57.17	58.50	55.17	59.50	57.50	66.50	54.83	62.50	60.00	57.67	59.50	63.67
7	76.67	75.33	72.67	68.83	71.50	71.00	74.00	72.67	72.83	72.67	65.33	60.67	72.67	68.33	76.00	76.83
8	67.67	67.50	64.00	61.83	64.33	63.67	63.33	61.33	71.33	69.83	78.00	77.33	56.67	53.50	68.50	66.50
9	59.00	58.83	69.50	68.83	66.00	63.50	69.00	66.83	73.00	67.83	59.17	58.67	58.67	59.00	65.83	62.33
10	69.33	68.17	57.67	60.50	63.00	64.67	63.50	62.67	66.17	66.67	67.00	68.00	63.33	57.83	64.83	63.33
Avg.	71.12	70.30	70.50	68.62	70.18	68.76	70.54	68.85	72.67	71.66	68.21	67.34	65.41	63.04	70.55	68.98
Std.	11.99	8.89	11.26	10.85	10.82	10.30	11.44	10.71	9.71	6.58	10.51	10.39	10.31	10.66	9.99	9.73
Chance level	50.80	50.60	50.69	50.21	49.92	49.97	50.69	49.95	50.53	50.01	50.63	50.28	50.40	50.40	50.88	50.35

familiar face stimuli among familiar face stimuli, are also reported in Table 2. Since the performance of PCM is better than that of PLM, sensitivity and specificity are only given for PCM features in this table. Selected channels contributing to classification success are different for each subject.

Table 3 shows the channels decided by using PCM features. When this table is examined, it is observed that the number of chosen channels differs with respect to the channel selection methods. It is seen from Table 1 that Subject 5 provides the best classification performance (89.33%) when Kullback–Leibler distance (KL2) for channel selection with PCM features is used. The two pairs of highest counts in the selected channel histogram are visualized with a 14 $\times$ 14 grid (matrix) in Figure 2a for Subject 5. The black-colored grids (elements) show the decided pairs. The brain map of the determined channels for this subject is also shown in Figure 2b. This brain map shows the locations of the selected channels on the scalp for familiar and unfamiliar face stimuli. From the figure, it is observed that F3–FC5 and F3–T7 are the most significant pairs of channels. It can be noted that F3 is chosen two times. The evoked potentials of these channels are shown in Figure 4 to examine their categorization capability. These potentials are obtained by averaging the trials of the whole data for each class. The stimulus time is accepted as a reference for these plots.

The average classification success of the subjects, which is obtained using KL2 for channel selection with PCM features, is the highest average performance. In order to obtain the most selected channels by subjects, the most distant channel pairs of the subjects are counted and the two pairs with the highest polls are considered to be the selected channels on average. The common matrix of the pairs of the most distant channels is displayed in Figure 3a. The brain map of these selected channels on the skull is shown in Figure 3b. From Figure 3, it is deduced that F3–FC5 and F3–O1 are the most selected pairs of channels. As seen, F3 is chosen two times. When these figures are examined, it is understood that the frontal and occipital regions in the brain are active

Table 2. Specificity (Qn) and sensitivity (Qd) of the classification with PCM (%).

	r correlation coefficient		Fisher score		Mutual information		Kullback–Leibler		Kullback–Leibler 2		Variance ratio		Variance ratio 2		Common spatial pattern	
Sbj	Qn	Qd	Qn	Qd	Qn	Qd	Qn	Qd	Qn	Qd	Qn	Qd	Qn	Qd	Qn	Qd
1	66.00	73.93	69.00	69.64	72.00	69.64	66.00	72.14	65.00	76.07	75.00	68.21	75.00	61.43	75.00	71.07
2	91.67	92.33	93.00	89.33	89.33	90.67	92.67	91.33	93.33	87.00	93.33	84.00	93.33	86.33	93.33	90.67
3	70.67	60.36	73.00	66.43	70.00	66.43	69.00	64.29	54.33	68.93	63.33	56.43	63.33	56.07	63.33	58.21
4	71.33	50.67	71.00	56.00	71.33	53.67	71.00	58.00	73.33	56.00	72.33	52.67	72.33	52.67	72.33	57.00
5	91.33	90.33	86.67	90.00	86.33	90.00	87.00	89.33	68.00	91.00	76.67	85.67	76.67	75.00	76.67	83.67
6	56.67	61.67	55.33	62.67	53.33	61.00	52.00	58.33	55.00	63.00	57.00	58.00	57.00	65.00	57.00	62.00
7	78.33	75.00	76.67	68.67	72.67	70.33	76.67	71.33	73.67	72.67	75.67	64.00	75.67	71.67	75.67	76.33
8	68.67	66.67	62.67	65.33	65.67	63.00	65.33	61.33	53.33	68.33	71.67	76.67	71.67	60.00	71.67	65.33
9	62.67	55.33	72.33	66.67	69.67	62.33	68.33	69.67	62.67	72.00	66.33	56.67	66.33	54.67	66.33	65.33
10	78.53	57.31	70.00	41.54	74.12	48.46	72.65	51.54	76.18	56.15	75.88	50.77	75.88	46.54	75.88	50.38
Avg.	73.59	68.36	72.97	67.63	72.45	67.55	72.06	68.73	67.48	71.12	72.72	65.31	72.72	62.94	72.72	68.00
Std.	11.50	14.33	10.79	14.24	10.04	13.74	11.44	13.11	12.36	11.58	9.67	12.83	9.67	11.90	9.67	12.54

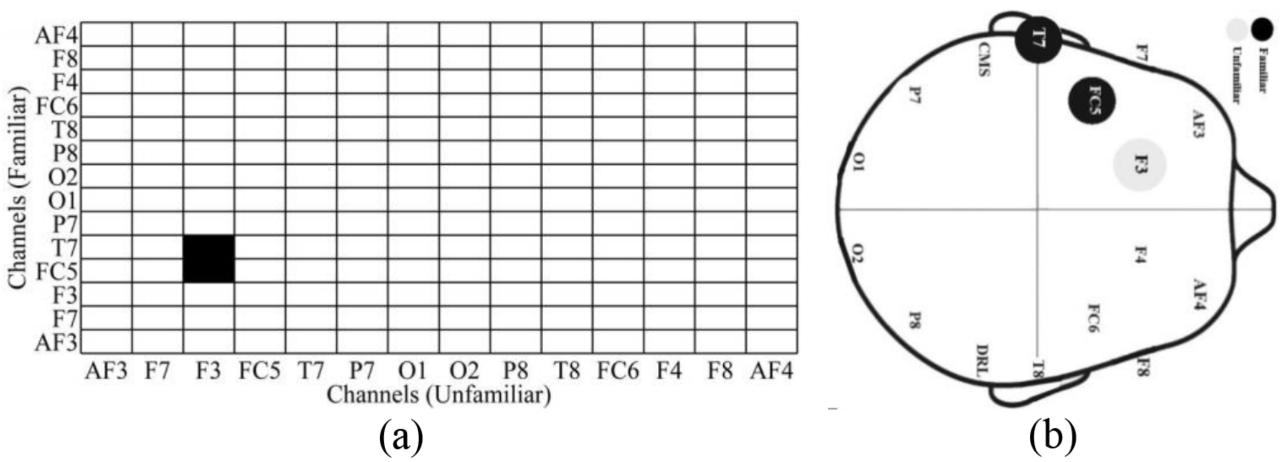

Figure 2. Selected channels for Subject 5: a) two most distant channel pairs pictured on the grid, b) locations of the channels on the scalp.

for familiar and unfamiliar face selectivity. In other words, it is observed that (F3, FC5) and (F3, O1) are the most distant channel pairs for familiar and unfamiliar classes.

5. Discussion and conclusion

The EEG data of the familiar/unfamiliar face recognition experiment were classified. Results were obtained for 10 subjects who participated in the experiment. In the experiment, the interstimulus interval was not randomized; instead, it was constant and the same for all subjects. No button was used to mark displayed portraits that were known to participants. Any undesired event during an EEG experiment could cause

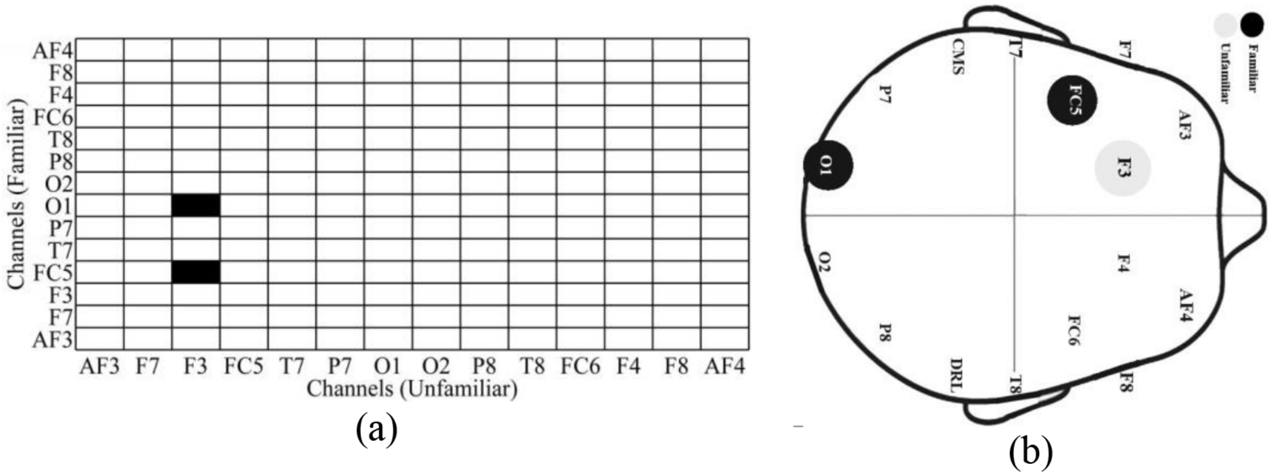

Figure 3. Preferred channels in average: a) the chosen channel pairs visualized on the matrix, b) locations of the channels on the scalp.

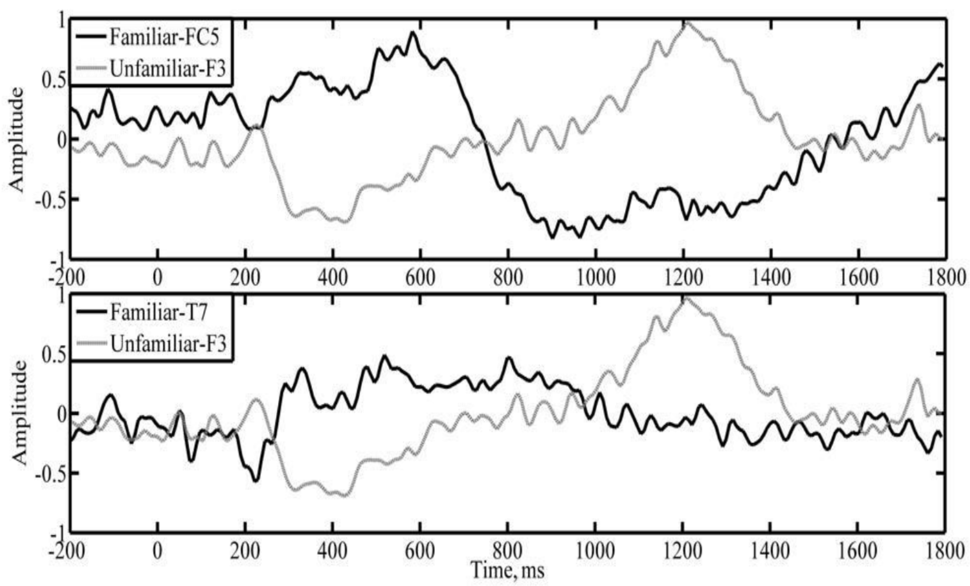

Figure 4. Evoked potentials (EPs) of channels F3, FC5, and T7 for Subject 5.

unwanted changes in the EEG records. These unacceptable variations are artifacts for the records. Button-pressing is a controlled event; however, preparation and planning might also alter the EEG before button pressing actions. It was not guaranteed that the response for face recognition and button pressing would never overlap. We also required subjects to focus on the face recognition event. Consequently, we did not include the marking of familiar and unfamiliar faces during the experiment. Furthermore, it is not certain that the faces recognized by a subject during and after the experiment coincide. On the other hand, we expect that the faces recognized by a subject during and after the experiment mostly coincide.

The type-II discrete cosine transform reduced the spatial correlation of the channels. The 1-s epochs and stimulus onset of 0.6 s to 1 s, were analyzed in the study. The 1-s epochs with 0.6 s to 1.6 s stimulus onset were also analyzed in the study. In the feature extraction process, the trial was divided into 31.25-ms

Table 3. The most distinctive pairs of channels obtained with the channel selection methods using PCM features.

	r correlation coefficient				Fisher score				Mutual information				Kullback–Leibler				Kullback–Leibler 2				Variance ratio				Variance ratio 2				Common spatial pattern			
Sbj	I		II		I		II		I		II		I		II		I		II		I		II		I		II		I		II	
1	F7	F8	F7	FC6	F4	O1	F4	P8	F3	P8	O2	T7	F8	O1	F8	T7	F8	F7	F8	AF3	AF4	AF3	AF3	AF4	O1	O1	AF4	AF4	AF3	FC5	AF3	F7
2	F7	F7	AF3	P8	F3	O1	F3	T7	AF3	T7	AF4	O1	F3	O1	FC5	F7	FC5	AF3	FC5	F3	AF4	AF3	AF3	AF4	O1	O1	F3	F3	AF3	FC5	F7	O2
3	T7	FC6	T7	P7	P8	F4	F8	FC 5	FC5	F3	P8	F4	P8	F3	F3	P8	FC5	F3	AF3	P8	AF4	AF3	AF3	AF4	O1	O1	F7	F7	AF3	P7	F7	O2
4	F7	F8	F7	T8	T8	F3	T8	O2	F7	O2	AF3	T7	O2	O1	T8	O1	FC6	AF3	F7	AF4	AF4	AF3	AF3	AF4	AF3	AF3	AF4	AF4	AF3	O2	AF3	O1
5	F3	F4	F3	FC5	F3	O1	F3	P8	F3	P8	F4	O1	F3	P8	FC5	F3	FC5	F3	T7	F3	AF4	AF3	AF3	AF4	O1	O1	T7	T7	AF3	FC5	F7	T7
6	F7	P8	F7	F8	T7	F3	FC5	F3	O1	F3	O1	AF4	O1	F4	P8	FC6	O1	F3	P8	F7	AF4	AF3	AF3	AF4	T7	T7	F7	F7	AF3	F3	AF3	FC5
7	T7	O2	P8	O1	F4	FC6	F4	O1	O1	F3	FC5	F3	FC5	P7	F4	FC5	O1	F3	O2	T7	AF4	AF3	AF3	AF4	P7	P7	F3	F3	AF3	T7	F7	T7
8	AF3	O2	AF3	P8	AF4	P7	O2	F7	O2	F4	O2	T7	FC6	P7	FC5	FC6	AF4	O1	AF4	T7	AF4	AF3	AF3	AF4	F8	F8	P8	P8	AF3	T7	F7	P7
9	F7	O1	F7	P8	F4	AF4	F4	T8	T8	F7	T8	O1	F4	T8	F8	P8	F7	AF4	F3	P7	AF4	AF3	AF3	AF4	P7	P7	AF4	AF4	AF3	FC5	F7	O2
10	F3	F4	F3	O2	FC6	P8	FC6	T7	AF3	P8	T8	FC6	F3	FC6	FC5	P8	T8	T7	F4	F3	AF4	AF3	AF3	AF4	O2	O2	P8	P8	F7	T8	AF3	O1

nonoverlapping segments. The average of a segment (PCM) and the slope of the first-order polynomial fitted to the segment (PLM) were employed as features. The four most discriminative channels (two channel pairs) were chosen by applying the similarity (correlation coefficient, mutual information) and distance (Fisher score, relative entropies) measures and the common spatial pattern filtering and variance ratio techniques. To the best of our knowledge, variance ratio as a channel selection method was first proposed in this study. The classification performances of the models and channel selection methods were examined and reported.

The classification accuracies were obtained with 20-fold Monte Carlo cross-validation. PCM for feature extraction and the Kullback–Leibler distance (KL2) as a channel selection method with SVM classifier achieved 72.67% average accuracy for the data set of this paper and yielded the highest performance. PCM yielded almost 2% more accuracy than PLM for the same channel selection method in the feature extraction stage. KL2 as a channel selection method performed almost 2%–3% better than other methods. KL2 is a measure of difference between time changes of two signals, whereas KL is a measure of distance between two probability distributions (scaled histograms). This does not mean that if these two signals have similar histograms the time changes of the two signals are similar. This is possibly why KL2 is better. In the case of two classes, the chance level was $100/2 = 50\%$, and this level was almost attained by 100-fold Monte Carlo cross-validation, each time randomly assigning class labels to data (Table 1). This shows that the results are not obtained by chance and are not accidental.

The statistical difference among channel selection methods accompanied with feature type (each column in Table 1) was investigated by a one-way ANOVA test. The test returned $P = 0.942$, which shows that there is no significant difference among channel selection approaches. Statistically, there is no prominent channel selection technique (together with feature) in the set of channel selection methods used. The interaction between features and channel selection methods was also investigated by applying two-way ANOVA. The test resulted in $P = 1$ for the interaction of channel selection methods and features. This outcome shows that there is no connection (relation) between channel selection and feature types.

Because the highest accuracy was obtained with zero-order polynomial fitting of nonoverlapping windows of EEG and the KL2 channel selection technique, we explored the most decided discriminative pairs of channels among subjects for this feature and the channel selection method. It was seen that F3–FC5 and F3–O1 were the most commonly selected pairs of channels. As F3 was chosen two times, the number of chosen channels was three. It was discovered that the frontal and occipital regions in the brain were decisive for familiar and unfamiliar face categorization.

In this study, the EEG channels were limited to 14 channels. From the standard international 10/20 electrode placement system, central and parietal electrodes were missing [25]. These electrodes were not included in the analysis. It was not possible to investigate the association of missing electrodes with categorization. It is known that the parts of the brain relevant for cognition and vision are mainly the frontal and occipital regions, respectively [4,7–12,26]. Most of the occipital and frontal regions were covered with the used electrodes in this research. Nevertheless, this is a shortcoming of the study, and future studies will use electrodes of the 10/20 system.

In conclusion, the findings of the proposed approach show that the study is suitable for face recognition for the purposes of medical applications such as diagnosis of brain diseases and criminal identification.

Acknowledgments

This study was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) and Adıyaman University's Scientific Research Fund (Project No. TIPBAP/2012-0008).

References

[1] Woodman GF. A brief introduction to the use of event-related potentials in studies of perception and attention. Atten Percept Psycho 2010; 72: 2031-2046.

[2] Gazzaniga MS. The Cognitive Neurosciences. 4th ed. Cambridge, MA, USA: MIT Press, 2009.

[3] Bruce V, Young A. Understanding face recognition. Brit J Psychol 1986; 77: 305-327.

[4] Jeffreys DA. Evoked potential studies of face and object processing. Vis Cogn 1996; 3: 1-38.

[5] Ewbank MP, Andrews TJ. Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. Neuroimage 2008; 40: 1857-1870.

[6] Eimer M. Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. Cognitive Brain Res 2000; 10: 145-158.

[7] Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 1997; 17: 4302-4311.

[8] Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM. Neural systems underlying the recognition of familiar and newly learned faces. J Neurosci 2000; 20: 878-886.

[9] Rossion B, Schiltz C, Robaye L, Pirenne D, Crommelinck M. How does the brain discriminate familiar and unfamiliar faces?: a PET study of face categorical perception. J Cognitive Neurosci 2001; 13: 1019-1034.

[10] Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci 2000; 4: 223-233.

[11] Rossion B, Caldara R, Seghier M, Schuller AM, Lazeyras F, Mayer E. A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain 2003; 126: 2381-2395.

[12] Kaan E, Swaab TY. Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. J Cognitive Neurosci 2003; 15: 98-110.

[13] Tanaka JW, Curran T, Porterfield AL, Collins D. Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. J Cognitive Neurosci 2006; 18: 1488-1497.

[14] Sun D, Chan CCH, Lee TMC. Identification and classification of facial familiarity in directed lying: an ERP study. PLoS One 2012; 7: e31250.

[15] Jin J, Daly I, Zhang Y, Wang X, Cichocki A. An optimized ERP brain–computer interface based on facial expression changes. J Neural Eng 2014; 11: 36004.

[16] Jin J, Allison BZ, Zhang Y, Wang X, Cichocki A. An ERP-based BCI using an oddball paradigm with different faces and reduced errors in critical functions. Int J Neural Syst 2014; 24: 1450027.

[17] Kaufmann T, Schulz SM, Grünzinger C, Kübler A. Flashing characters with famous faces improves ERP-based brain–computer interface performance. J Neural Eng 2011; 8: 56016.

[18] Çelik U, Arica S. Classification of evoked potentials of familiar and unfamiliar face stimuli using multi-resolution approximation based on excitatory post-synaptic potential waveform. Comput Electr Eng 2013; 39: 1571-1584.

[19] Puth MT, Neuhäuser M, Ruxton GD. Effective use of Pearson's product–moment correlation coefficient. Anim Behav 2014; 93: 183-189.

[20] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. New York, NY, USA: Wiley, 2000.

[21] Cover TM, Thomas JA. Elements of Information Theory. 2nd ed. Hoboken, New Jersey, USA: Wiley-Blackwell, 2006.

[22] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951; 22: 79-86.

[23] Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Proc Mag 2008; 25: 41-56.

[24] Alpaydın E. Introduction to Machine Learning. 2nd ed. Cambridge, MA, USA: MIT Press, 2004.

[25] Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. Int J Clin Neuropsyc 2001; 112: 713-719.

[26] Guyton AC, Jhon E. Cerebral cortex, intellectual functions of the brain, learning and memory. In: Schmitt W, Gruliow R, editors. Textbook of Medical Physiology. 11th ed. Philadelphia, PA, USA: Elsevier, 2006. pp. 714-727.