Classification of Brain Signal (EEG) Induced by Shape-Analogous Letter Perception

Rohit Bose[1], Sim Kuan Goh[1], Kian F Wong[2], Nitish Thakor[1], Anastasios Bezerianos[1] and Junhua Li[1, 3,4*]

[1]Singapore Institute for Neurotechnology (SINAPSE), National University of Singapore, Singapore
[2]Centre for Cognitive Neuroscience, Duke-NUS Medical School, Singapore
[3]School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK
[4]Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen, China

Abstract

Visual perception of English letters involves different underlying brain processes including brain activity alteration in multiple frequency bands. However, shape analogous letters elicit brain activities which are not obviously distinct and it is therefore difficult to differentiate those activities. In order to address discriminative feasibility and classification performance of the perception of shape-analogous letters, we performed an experiment in where EEG signals were obtained from 20 subjects while they were perceiving shape analogous letters (i.e., 'p', 'q', 'b', and 'd'). Spectral power densities from five typical frequency bands (i.e., delta, theta, alpha, beta and gamma) were extracted as features, which were then classified by either individual widely-used classifiers, namely k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF) and AdaBoost (ADA), or an ensemble of some of them. The F-score was employed to select most discriminative features so that the dimension of features was reduced. The results showed that the RF achieved the highest accuracy of 74.1% in the case of multi-class classification. In the case of binary classification, the best performance (Accuracy 86.39%) was achieved by the RF classifier in terms of average accuracy across all possible pairs of the letters. In addition, we employed decision fusion strategy to exert complementary strengths of different classifiers. The results demonstrated that the

*Corresponding author
Email=junhua.li@essex.ac.uk

performance was elevated from 74.10% to 76.63% for the multi-class classification and from 86.39% to 88.08% for the binary class classification.

1. Introduction

Perception of letters has been crucial to Brain Computer Interface (BCI) based systems for augmenting communication in patients suffering from Parkinson's disease, Amyotrophic Lateral Sclerosis (ALS), or other similar motor disabilities [1].

Previous studies of letter recognition usually adopted an oddball paradigm of a 6×6 speller [2]. In such design, the matrix comprising of English alphabets and symbols flickers row wise and column wise. When the highlighted row and column matched the desired alphabet which the subjects wanted to select, an evoked P300 potential is elicited in the brain, which can be detected in EEG signals. This evoked potential is time-locked to the display of flickering alphabet and not related to the visual perception of letter shape. Up to now, such perception of letter shape has been less studied and the classification of shape analogous letters should be further addressed.

Several brain imaging modalities like Electroencephalography (EEG)[2], Magnetoencephalography (MEG)[3, 4], and functional Magnetic Resonance Imaging (fMRI)[5], have been used to study the underlying characteristics representing differential perception-induced activities between different English letters. In the case of word perception, it has been observed that the frontal, parietal and the occipito-temporal regions in both hemispheres play a major role. Studies based on fMRI have shown that letter perception was relevant to a couple of brain regions that interact with each other to form a network [6-8]. Such brain regions are mostly located in the occipital regions and left temporal regions [9]. Moreover, repeated presentation of a letter leads to activation bilaterally in the posterior region of the brain [8]. The perception of different letter involves a broad spatial frequency spectrum. For shape analogues letters, this spatial spectrum is almost similar and hence discriminating the

brain perception to these letters becomes a difficult and challenging task [10, 11]. Inter-letter similarities based on Euclidean distance shows that four English letters; 'p', 'q', 'b' and 'd', are the most similar shape analogues letters among all the 26 English letters [6, 12]. These four alphabets fall under a particular visual group consisting of a circle and line. Hence, these four letters were chosen for analysis in this study.

Numerous classifiers have been employed in machine learning studies to differentiate two or more groups or classes. Support Vector Machine (SVM) has been used in numerous fields like text categorization [13], remote sensing [14], neurological disorders [15], power fault diagnosis [16] and also in P300 speller [17-21]. An ensemble of SVM classifiers yielded high performance in P300 speller study [10, 22]. Linear Discriminant Analysis (LDA) has been used in P300 speller studies [23, 24]. A comparative study of different classifiers like SVM and LDA is done in [20] and reported that LDA gives better performance. K-Nearest Neighbor (kNN) has also been used in P300 studies and has yielded satisfactory results[25]. AdaBoost is also a popular classifier and has also been used in P300 speller studies [26, 27]. Random Forest (RF) classifier has also been widely used in P300 speller based study [28]. Therefore, based on the literature (mostly related to P300 speller), in this study, we have selected kNN, SVM, LDA, RF and ADA, to classify the perception of shape analogous letters.

In this study, we investigated the perception of four shape analogous letter using power spectral based features using Power Spectral Density (PSD) from five different frequency bands. Using F-score based feature selection, we selected the relevant features distinguishing the perception of the shape analogous letters. Multi-class classification was then performed based on the selected features to classify the four analogous English letters. To obtain a deeper insight into the classification performance for different letter pairs, we also performed binary class classification for all 6 possible pairwise combination
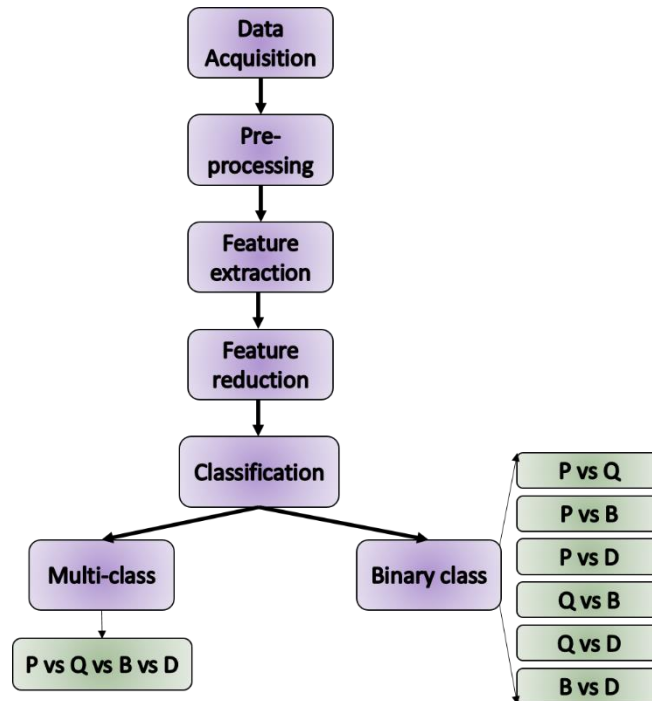
Figure 1. Flowchart of the methodology used in this study. The acquired data is pre-processed and then power spectral features has been extracted. After dimensionality reduction classifiers were trained. The trained classifiers were then used to perform multi-class and binary class classifications

of the four letters. Finally, we also employed the decision level fusion strategy to further enhance the classifier performance for both multi-class and binary class classification. A comparative study of five classifiers is also shown in this paper using the post-hoc Tukey-Kramer test to identify the best classifier for both multi-class and binary class classification.

2. Methodology

The flowchart of the methodology adopted in this study is shown in Figure 1. The acquired EEG data were pre-processed and power spectral features from the five frequency bands, namely delta, theta, alpha, beta and gamma, were extracted. F-score based feature reduction technique was employed to select the relevant features with high discriminative power. The multi-class and binary classifications were then performed using five widely-used classifiers. Further, to enhance the classifier

performance, we employed a classifier combination using a decision-based fusion manner.

2.1 Experiment Protocol

A total of 20 subjects (12 females and 8 males; Mean Age: 23.1 and Standard Deviation: 3.1) recruited from the National University of Singapore participated in this study. The subjects were asked to refrain from any consumption of alcohol or caffeine substances for one day before the experiment. The subjects were also instructed to avoid any kinds of strenuous exercise before the experiment. A minimum 7 hours of sleep was instructed to the subjects for the night prior to the experiment. All subjects had a normal or corrected-to-normal vision. All subjects provided informed consent and were monetarily compensated for their participation. The study was approved by the Institutional Review Board of the National University of Singapore in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments.

The whole experiment was divided into 2 sessions with an approximately 1-week gap between the sessions. The fours shape analogous letters: 'p', 'q', 'b' and 'd', were displayed to the subjects visually in a screen. The visual cues were presented for 0.2 sec and then a fixation cross for 1.8 sec. Hence, the total time for each trial was 2 sec. In each session, a total of 120 trials were presented to the subjects. The trial sequences were randomized using optseq (http://surfer.nmr.mgh.harvard.edu/optseq). The stimulus was provided using Psychtoolbox in Matlab 2011a (Mathworks, USA). A detailed description of the experimental protocol can be found in [29].

2.2 Data Acquisition

Concurrent EEG-fMRI data acquisition was conducted at the Clinical Imaging Research Centre (CIRC) using a Siemens 3T Magnetom Prisma MRI scanner (Siemens, Erlangen, Germany) and an ANT amplifier with an MRI-compatible 64 channel Waveguard cap (Waveguard, ANT B. V., Netherlands). The fMRI data were not included in this paper. EEG signals were acquired at a 4KHz sampling rate using 64

130 channels. The impedance was kept below 15 KΩ and the reference was the linked mastoids (M1 and M2).

2.3 Pre-processing

The pre-processing steps for this study have been customized to remove the common
135 EEG artifacts as well as the gradient artifacts caused by a magnetic field [30]. Briefly, EEG signals were first up-sampled to 40 kHz to align the volumes of EEG and fMRI. To remove the MRI related artifacts Canonical Correlation Analysis (CCA) based method has been used [30]. The EEG signals were then down-sampled to 400Hz. The EEG data have then been segmented from the beginning of the visual cue to 1.5 seconds
140 after the cue for each trial. A total of 240 trials were extracted (60 for each class). Movement and ocular artifacts were mitigated using Independent Component Analysis (ICA).

2.4 Feature Extraction

The filtered and epoched EEG signals were first transformed from the time domain
145 to the frequency domain using fast Fourier transform. Next, the power for all the trials in each channel in 5 EEG frequency bands were extracted: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz) and gamma (30–40 Hz). A total of 62 x 5 = 310 features for each epoch were extracted for classification. Dimensionality reduction was done using F-score. All features were first ranked based on F-score. Then a grid search
150 was performed using 10% to 100% of all the features (10% step size) to obtain the optimum number of features that will yield the best performance.

3. Classifiers

In this study, we employed five widely used classifiers: Support Vector Machines (SVM) [31], k-Nearest Neighbors (kNN) [32, 33], Linear Discriminant Analysis (LDA) [34], AdaBoost-SAMME (ADA) [35] and Random Forest (RF)[36].

3.1 k-Nearest Neighbor

k-Nearest Neighbors (kNN) is a supervised classifier that works without a prior assumption about the distribution of the data. To classify an unknown data point, the kNN classifier calculates the distance of the data point with k nearest points and assigns the class based on majority [32, 37, 38]. In this study, we used the Minkowski distance to calculate the distance between the new data point and the k nearest points.

3.2 Support Vector Machine

Support Vector Machine (SVM) is a model-based learning algorithm where it implies a hyperplane to segregate the classes [31]. The hyperplane can be constructed using multiple kernel functions like linear, quadratic, polynomial and Radial Basis Function (RBF) [32, 39]. In this study, we have employed the RBF kernel as it yielded the best performance compared to the other kernel functions. If the data is not completely separable, a soft margin is inserted to increase the tolerance to misclassification and is denoted by parameter C. SVM is a binary classifier but can be extended for multi-class classification using one-versus-rest strategy, where prediction on testing datasets are based on the model built from training datasets.

3.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) can be described as a classifier that optimizes the number of features that maximizes the ratio of the between-class variance to the within-class variance for all classes. It works on the assumption that each class has a Gaussian distribution [34, 40].

3.4 Random Forest

Random Forest (RF) is an ensemble classifier. It uses a combination of Classification and Regression Trees or weak classifiers by using a subset of training data using a bagging approach. Hence, similar samples can be used multiple times during classifier training. The prediction of a new sample was done by taking the majority of votes in the ensemble of weak classifiers [41, 42].

Let us consider M is the number of decision trees, x as the observations and C(x) as their class labels. For each decision tree, bootstrap samples X are drawn from the training data. A decision tree T_m is then built recursively by selecting the best division among m variables to obtain two sub-nodes until a leaf node is reached. Subsequently, the ensemble $\{T_m\}_m^M$ is given as an output. This process is iterated M number of times, which is equal to the number of decision trees.

3.5 AdaBoost

AdaBoost (ADA) is also an ensemble classifier using a boosting algorithm (unlike RF) that builds each weak classifier iteratively [35]. The classifier iteratively approximates the Bayes classifier. Initially, using an unweighted training sample, a classifier is built. Based on the error rate, the weights of the training data point are increased. These updated weights are used to build the second classifier. Using this process iteratively, the ensemble of classifiers is built. To extend from binary to multi-class classification, the AdaBoost- Stagewise Additive Modeling using a Multi-class Exponential loss function (ADA-SAMME) is used in this study [35].

Let us consider M as the number of decision trees, x as the data and C(x) as the class labels. For Adaboost, the observation weights are initialized using

$$w_i = \frac{1}{N} ,$$

where i=1,2, N for each training sample, where each sample belongs to the class $\{k \in 1,2, \cdots K\}$ Then an iterative process is run to optimize the decision tree and obtain the decision tree. In the iterative process, firstly, a decision tree T_m (x) is fitted with the

205   training data using the weights w_i. Then, the error is calculated using the following formula,

$$err_m = \frac{\sum_{i=1}^{N} w_i \cdot II \ \{c_i \neq T_m(x_i)\}}{\sum_{i=1}^{N} w_i} \ ,$$

where c_i is the class label of ith data. The parameter a_m is then evaluated using the formula

210

$$a_m = log\left(\frac{1 - err_m}{err_m}\right) + log(K - 1)$$

The previously assigned weights are then updated and re-normalized to

$$w_i \leftarrow w_i \cdot exp \ (a_m \cdot II \ \{c_i \neq T_m(x_i)\}), i \ = \ 1, 2, \cdots, N.$$

This process is iterated to M number of times to obtain the trained classifier.

3.6 Hyperparameter Optimization

215   To optimize the classifier parameters and the number of features used to train the classifiers, we performed a grid-search for a wide range of the parameter values. The extracted features from the EEG signals were divided into 80% training and 20% testing datasets. The classifier parameter values and the percentage of the total number of features were selected one by one to evaluate the performance of the classifier using a

220 5-fold cross-validation technique. For each fold, the training data is further divided into 64% training data to train the model and the remaining 16% of the data for validation. The combination of the respective classifier parameter and the number of features, that yielded the highest performance, were selected for analysis. It should be noted that for RF, ADA, SVM and kNN there is one classifier parameter to be optimized, whereas,

225 LDA has no parameter for optimization. Therefore, for LDA, only the number of features has been varied to obtain the optimum number of features.

     The number of features has been varied from 10% to 100% of all the features with a step size of 10%. In the case of ADA and RF, the number of decision trees was varied based on a percentage of the total number of features (310), i.e. from 10% to 100% with

230 a step size of 10%. For instance, 10% of decision trees, shown in Figure 2, are 31 decision trees (10% of the total number of 310). In other words, the number of decision trees is varied from 31 to 310 with a step size of 31. For SVM, log_2 C parameter was

varied from -5 to 17 with a step size of 2. For kNN, the value of k was varied from 1 to 14 with a step size of 1.
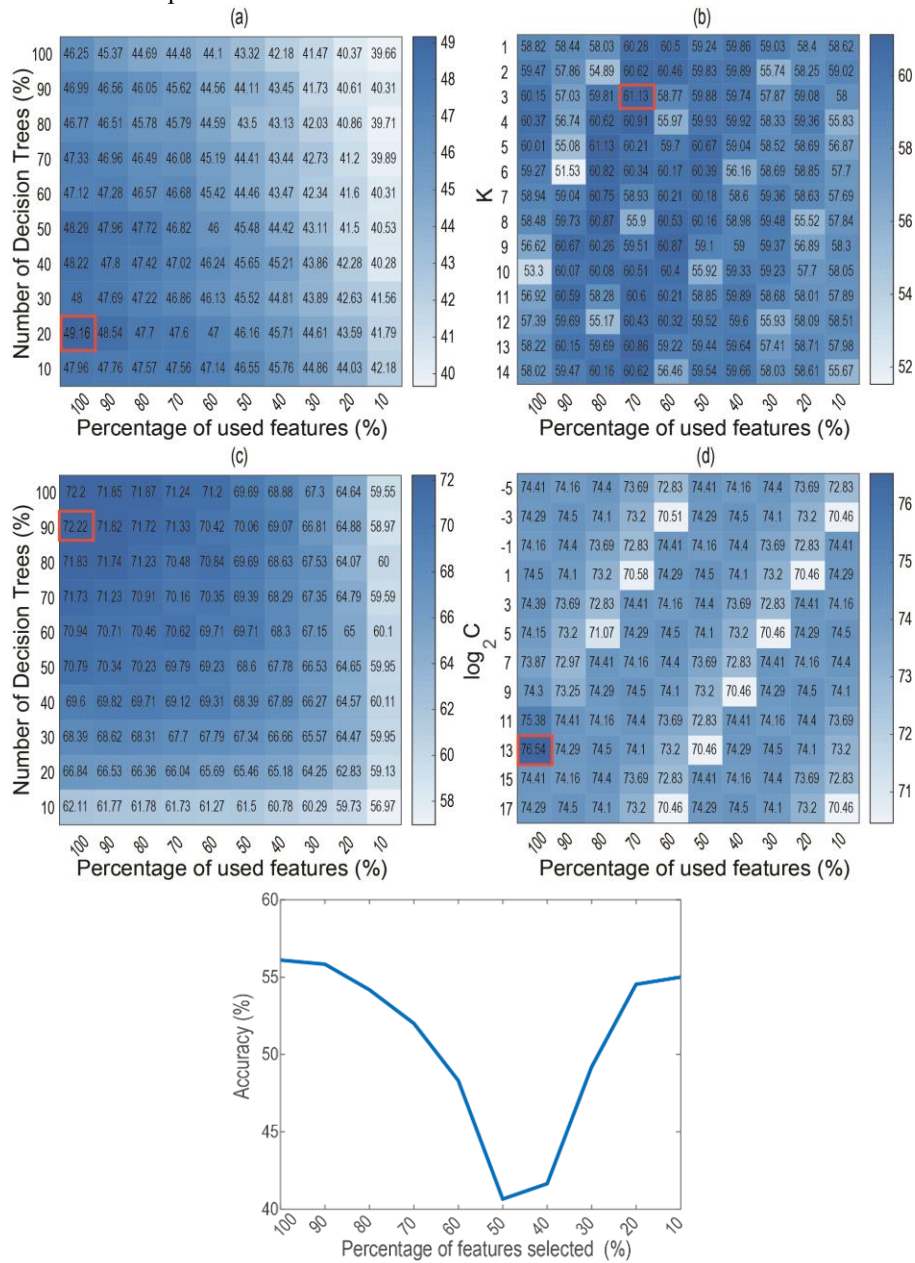

Figure 2. Heatmap of hyperparameters for (a) ADA; (b) KNN; (c) RF; (d) SVM; (e) LDA. The color bar for (a) to (d) indicates the average validation accuracy and the red dashed box indicates the optimal parameter combination for each figure.

Figure 2 shows the grid search results for all the five classifiers. It can be observed that ADA, RF and SVM, the classifier performance increases with the increase in the number of features. Therefore, the best performance was obtained using all the features. In the case of kNN, 70% of all the features gave the highest performance, followed by which the performance decreased. For ADA, 20% of the number of decision trees gave the best results and for RF, 90% of the decision trees gave the highest performance. In the case of kNN, there was no trend for the variation of the k parameter value. The best performance for kNN is obtained when k is three. In the case of SVM, the best performance is obtained for $\log_2 C$ value of 13 (C=8192). In the case of LDA, the highest accuracy is obtained with all the features. These parameter values and the respective number of features that gave the best performance were used for the rest of the analysis in the paper.

4. Results

4.1 Multi-class classification

Five-fold cross-validation was employed to assess classification performance. The trials were divided into training and testing sets. We used 5-fold cross-validation
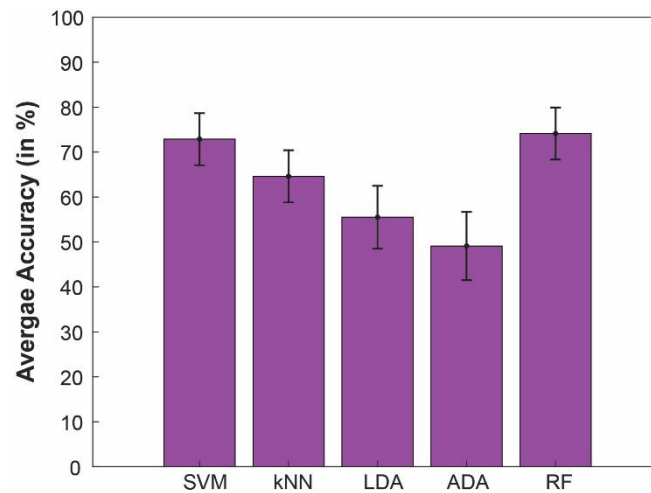
Figure 3. Average classification accuracy across all the subjects for multi-class classification of the shape-analogues letter perception for the five classifiers.

technique to split the data into training and testing sets. All five classifiers were trained using the training set, which was followed by testing the classifier using the testing set. The classification accuracy obtained from the testing dataset is reported in this study. The mean classification accuracies averaging across all subjects for all classifiers are shown in Figure 3.

The RF achieved a classification accuracy of 74.1%, which is the best among the classifiers. The SVM performed a comparable performance (72.8%) compared to the RF. ADA classifier performance the worst with a classification accuracy of 49.1%.

Statistical analysis using one-way ANOVA test shows that there is a significant difference between the performance of the classifiers ($p < 0.001$). To gain insight into the classifier performance, we performed the Post-hoc Tukey-Kramer test (as shown in Figure 4). It can be observed that both SVM and RF performed significantly better than the LDA and ADA ($p < 0.001$). However, there is no significant difference between the RF and SVM. No significant difference was observed between the LDA and ADA. For kNN, a significant difference is observed with RF ($p < 0.05$) but not with SVM.
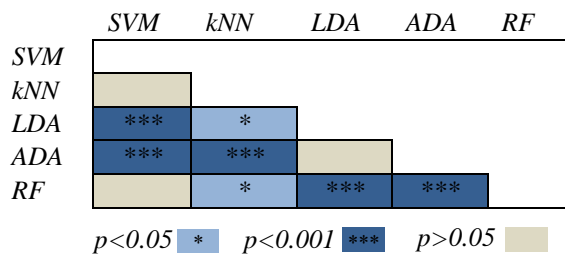

Figure 4. Post-hoc Tukey-Kramer test between all the classifier performances. The dark blue cells indicate significance with p<0.001 and the light blue cells indicate significance with p<0.05. The grey cells indicate no significance was observed.

4.2 Binary class classification

For binary classification, we evaluated all the possible pairs of the shape-analogues letters (totally six pairs: 'p' vs 'q', 'p' vs 'b', 'p' vs 'd', 'q' vs 'b', 'q' vs 'd' and 'b' vs 'd'). The average classification accuracies for each binary classification are shown in Figure 5. It can be observed that for binary classification LDA, ADA and RF give better performance than SVM and kNN. One-way ANOVA test shows that there is a significant difference between the performance of the classifiers ($p < 0.001$), for all the

6 cases. RF gives the highest classification accuracy for all the 6 classification cases (average classification accuracy=86.41%). Post-hoc test shows that RF gives a significantly high performance (p<0.01) from kNN for all the cases. Compared to SVM, RF gives a significantly higher performance (p<0.05) for all the cases except 'p' vs 'd'. No significant difference was obtained between LDA, ADA and RF classifier. The worst performance is obtained by kNN classifier (average classification accuracy is 64.04%).
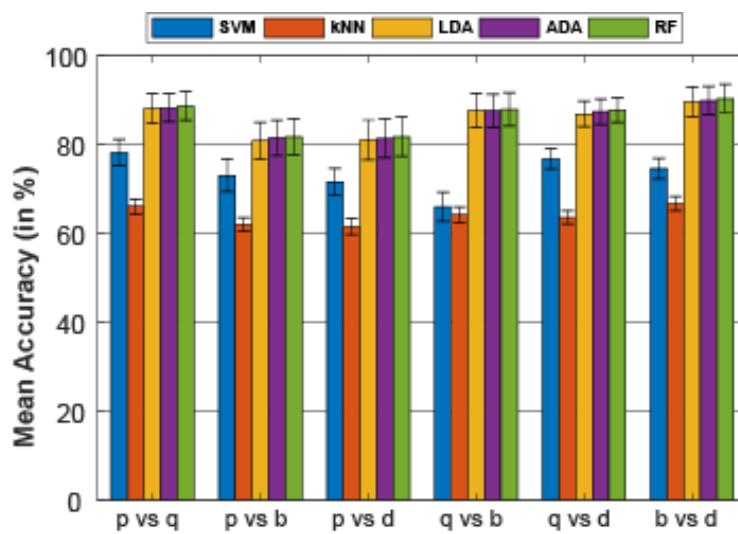

Figure 5. Average classification accuracy across all the subjects for each binary class classification of the shape-analogues letter perception for the five classifiers.

4.3 Decision Level Fusion based Classification

To enhance the performance of the classification, we employed the decision level fusion strategy. Among the five classifiers, three best classifiers were included for decision fusion (LDA, ADA and RF). The weighted consensus voting was utilized to fuse the outputs of individual classifiers[43]. A weight corresponding to the respective classification accuracy was assigned to each classifier. This manner was adopted for decision level fusion of both binary and multi-class classification conditions. Table 1 shows the classification performance for both binary and multi-class classification obtained using decision level fusion strategy. Using bootstrapping technique, we

290 obtained the 95% confidence interval based on 5000 times of resampling. The lower and upper values of the interval are shown in Table 1 along with the mean accuracies. For comparison, the accuracy obtained using RF classifier (since RF achieved the highest classification accuracy for both binary and multi-class classification cases) is also shown.

Table 1. Average accuracy [confidence interval] for multi class and binary class classification using Random Forest classifier and decision level fusion strategy. The confidence interval is calculated

		Random Forest (Mean [Confidence Interval]) (in %)	With Fusion Strategy (Mean [Confidence Interval]) (in %)
Multi-class Classification		74.10 [69 79]	76.63 [73 80]
Binary Classification	p vs q	88.72 [82 94]	90.56 [86 95]
	p vs b	81.80 [73 89]	86.17 [81 91
	p vs d	81.71 [72 89]	86.28 [78 91]
	q vs b	87.94 [79 94]	88.10 [82 93]
	q vs d	87.83 [82 93]	86.33 [77 93]
	b vs d	90.38 [83 95]	91.08 [87 94]

295

It can be observed from Table 1, for multi-class classification, decision level fusion strategy has improved the classification performance to 76.63% from 74.10%. For binary class classification, decision level fusion has been observed to improve the classification performance from 85.6% to 87.49% (average classification accuracy across all the 6 300 cases). Except for 'q' vs 'd', the classification accuracy has improved using the decision level fusion strategy. However, we did not observe a statistically significant difference in performance between the RF and the fusion-based classifier for both binary and multi-class classification cases. One possible explanation could be that individual classifiers made quite similar judgements in sample classification.

4.4 Common Selected Electrodes and Frequency Bands

We further analyzed the common electrodes and frequency bands that were selected after the F-score based feature reduction step. This will allow us to understand the specific electrodes and the frequency bands, that contributes to the classification performance.

To identify the consistent electrode and frequency band, we selected the top 20% of the features after the F-score based feature reduction step. From these top 20 of the features, we calculated the number of times it has been common across the subjects. If a particular feature or band is common across more than 7 subjects, we selected it as a consistent electrode or frequency band. Figure 6 shows the consistent electrodes (Figure 6(a)) and frequency bands (Figure 6(b)). It can be observed that the selected consistent electrodes are mainly located in the pre-frontal, lateral and parieto-occipital areas. In terms of the frequency band, the most consistent band is gamma, followed by theta and alpha.
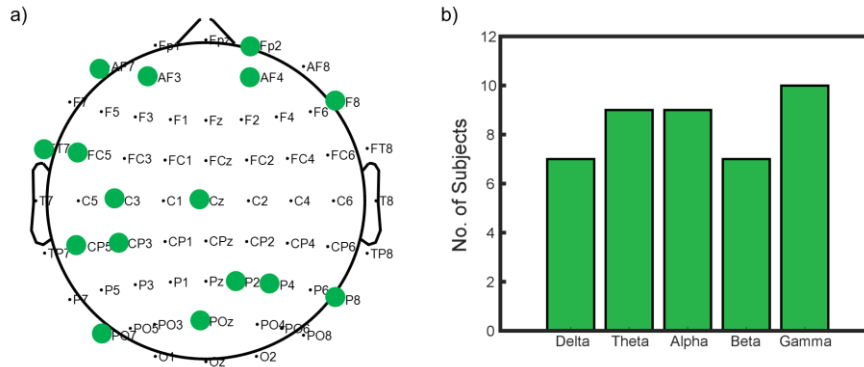
Figure 6. a) Common channels (marked with green) that are common for more than 7 subjects are shown. The common electrodes can be observed in the pre-frontal, frontal, left central and parietal regions. b) The frequency band that is common across the number of subjects is shown. The gamma band is most common across the subjects (10 subjects) followed by theta and alpha band.

5. Discussions

In this study, we performed both multi-class and binary class classification. The results clearly show that RF achieved the best performance in both the binary and multi-class classification compared to the other classifiers. Therefore, RF proves to be a reliable classifier that excels in both multiple and binary class classifications. A similar conclusion was drawn in other machine learning studies. For example, a detailed comparative study of classifiers using 121 datasets showed that RF classifier performs the best [44]. Similar results were observed in an image recognition study where RF performed better than SVM [45]. Other studies in protein-localization pattern classification [46] and motor fault diagnosis [47] also observed similar results. The ensemble of classifiers comprising the RF classifier can be represented as a nonlinear function which allows better learning of the data[48]. Further, the bagging technique of the RF classifier helps to reduce variance and overfitting of the data. These properties enable the RF classifier to perform better than other classifiers.

Letter perception depends on mapping the letters in a spatial frequency spectrum and then matching this template when that letter reappears. Even if the letter is combined with noises, the human brain performs low-pass and high-pass filtering to extract the spatial spectrum template matching with a particular letter [6, 49]. Such perception

activates different brain regions namely the fusiform gyrus, pre-central areas, inferior frontal gyrus located at the central and frontal brain regions. These activations are reflected in the spectral power of the EEG signals. In our study, such differences in brain activation were also observed for shape analogous English letters. Different activation levels in the brain allow segregating the spatial spectrum of the respective letters in the brain. The classification performance also was similar for all the letter pairs in case of binary classification. Therefore, our results clearly indicate that EEG spectral power features prove to be an effective feature to distinguish the perception of shape-analogous letters.

We observed that the classification performance enhanced for decision level fusion strategy compared to the best individual classifier performance (RF in our case) for both binary and multi-class classifications. Due to different properties of different classifiers, the misclassification trials vary from one classifier to another. For instance, although the RF gave high performance, the trials misclassified by RF might be correctly classified by other classifiers. Decision-based fusion strategy provides the solution as it takes the advantages of different classifiers and improves the classification accuracy through the majority voting [50-53].

Apart from the decision fusion strategy, fundamental properties of the classifiers (like the optimization of the classifiers), also play a major role in the performance. For instance, multiple trees in RF classifier enhances the performance and decreases the chance of over-fitting of the data [41, 54]. A similar observation has been made in our study. In Figure 2(c), we can observe that for RF, the classification performance increases with the increase in the number of trees. Such observations were not made for kNN classifier. In the case of kNN, the testing samples are assigned based on the nearest neighbours. Hence, increasing the number of nearest neighbours will increase the chance to be misguided by outliers.

The common electrodes that were selected after the feature reduction step across the subjects are mainly located in the pre-frontal, lateral and parieto-occipital areas. The visual complexity of the letters is processed in the occipital regions [4]. In addition to the visual areas, the sensorimotor areas, mainly in the left hemisphere, has been

previously reported to be engaged in letter perception [8]. This left sensorimotor region is also useful to discriminate the perception of shape analogous letters in our study. The visual processing capability of the parietal region is a well-known fact and has been found to be involved in letter perception [55, 56]. For letter identification, it has been found that stored motor program of a letter in the brain which determines the trace of writing the alphabet, is important for letter perception [8, 57]. In case of similar letters, we can also infer that beyond the shape and visual features of the letter, the subjects rely on the stored motor pattern in the brain that is required to write that specific letter.

The gamma frequency band was found to contribute most to the accuracy followed by theta and alpha bands. Gamma band has been observed in retrieving and forming object representation [58, 59]. Both gamma and alpha bands have been observed to be dominant during letter perception [60]. The activity in the motor cortex for letter perception is observed in the alpha band [61]. The theta band has been observed to govern several cognitive processes [62]. Theta band oscillations are also responsible for visual perception [63]. Theta stimulation in the right parietal has shown to enhance global visual processing in the brain [64]. These three bands also show high discriminating power across the subjects.

6. Conclusions

In this study, we aimed at classifying the perception of shape analogues English letters based on EEG signals. Our approach provides an alternative method for classifying letter perception in the brain using features based on actual brain activation during the letter perception compared to the traditional P300 speller, where the flickering of the desired letter in a speller is identified. Furthermore, in contrast to the large and distinct P300 signal, the focus of this study is on the more challenging task of classifying the perception of four lowercase English letters with highly similar visual characteristics (consisting of a line and circle). We used spectral power features from all the five frequency bands and then F-score to select relevant features for classification. We obtained 74.1% classification accuracy across all the subjects. In the case of binary classification, we obtained 85.60% classification accuracy (averaged across all the six

classification problems). Based on post-hoc Tukey-Kramer test, RF gave the best performance for both the binary and multi-class classification. Using the decision level fusion strategy, we further enhanced the classification performance to obtain 76.63% and 87.49% for multi-class and binary classifications, respectively. These results show that using power spectral features, the perception of shape-analogous English letters can be classified using the RF classifier. In future, we aim to investigate the real-time classification of perception of similar shape English letters.

References

[1] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on Batch Normalization for P300 signal detection," *Neurocomputing,* 2017.

[2] F. Akram, S. M. Han, and T.-S. Kim, "An efficient word typing P300-BCI system using a modified T9 interface and random forest classifier," *Computers in biology and medicine,* vol. 56, pp. 30-36, 2015.

[3] A. Tarkiainen, P. Helenius, P. C. Hansen, P. Cornelissen, and R. Salmelin, "Dynamics of letter string perception in the human occipitotemporal cortex," *Brain,* vol. 122, pp. 2119-2132, 1999.

[4] A. Tarkiainen, P. L. Cornelissen, and R. Salmelin, "Dynamics of visual feature analysis and object‐level processing in face versus letter‐string perception," *Brain,* vol. 125, pp. 1125-1136, 2002.

[5] L. Cohen, S. Dehaene, L. Naccache, S. Lehéricy, G. Dehaene-Lambertz, M.-A. Hénaff, *et al.*, "The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients," *Brain,* vol. 123, pp. 291-307, 2000.

[6] J. Grainger, A. Rey, and S. Dufau, "Letter perception: from pixels to pandemonium," *Trends in cognitive sciences,* vol. 12, pp. 381-387, 2008.

[7] T. Parviainen, P. Helenius, E. Poskiparta, P. Niemi, and R. Salmelin, "Cortical sequence of word perception in beginning readers," *Journal of Neuroscience,* vol. 26, pp. 6052-6061, 2006.

[8] K. H. James and I. Gauthier, "Letter processing automatically recruits a sensory–motor brain network," *Neuropsychologia,* vol. 44, pp. 2937-2949, 2006.

[9] H. Gros, K. Boulanouar, G. Viallard, E. Cassol, and P. Celsis, "Event-related functional magnetic resonance imaging study of the extrastriate cortex response to a categorically ambiguous stimulus primed by letters and familiar geometric figures," *Journal of Cerebral Blood Flow & Metabolism,* vol. 21, pp. 1330-1341, 2001.

[10] D. Krusienski and G. Schalk, "BCI Competition III Challenge 2004," 2004.

[11] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and

the t-value scalogram," *IEEE Transactions on Biomedical engineering,* vol. 51, pp. 1057-1061, 2004.

[12]   P. Courrieu, F. Farioli, and J. Grainger, "Inverse discrimination time as a perceptual distance for alphabetic characters," *Visual Cognition,* vol. 11, pp. 901-919, 2004.

[13]   T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, 1998, pp. 137-142.

[14]   G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 66, pp. 247-259, 2011.

[15]   G. Orru, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review," *Neuroscience & Biobehavioral Reviews,* vol. 36, pp. 1140-1152, 2012.

[16]   F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements," *IEEE Transactions on Power Systems,* vol. 26, pp. 1474-1483, 2011.

[17]   M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 1073-1076, 2004.

[18]   P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter, "Improving transfer rates in brain computer interfacing: a case study," in *Advances in Neural Information Processing Systems*, 2003, pp. 1131-1138.

[19]   B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl*, et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE transactions on neural systems and rehabilitation engineering,* vol. 14, pp. 153-159, 2006.

[20]   D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan*, et al.*, "A comparison of classification techniques for the P300 Speller," *Journal of neural engineering,* vol. 3, p. 299, 2006.

[21]   N. V. Manyakov, N. Chumerin, A. Combaz, and M. M. Van Hulle, "Comparison of classification methods for P300 brain-computer interface on disabled subjects," *Computational intelligence and neuroscience,* vol. 2011, p. 2, 2011.

[22]   A. Rakotomamonjy and V. Guigue, "BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller," *IEEE transactions on biomedical engineering,* vol. 55, pp. 1147-1154, 2008.

[23]   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, pp. 174-179, 2000.

[24]   L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology,* vol. 70, pp. 510-523, 1988.

[25] C. S. Syan and R. E. Harnarinesingh, "Comparison of pre-processing and classification techniques for single-trial and multi-trial P300-based brain computer interfaces," *American journal of applied sciences,* vol. 7, p. 1219, 2010.

[26] A. Yildirim and U. Halici, "Analysis of dimension reduction by PCA and AdaBoost on spelling paradigm EEG data," in *Biomedical Engineering and Informatics (BMEI), 2013 6th International Conference on*, 2013, pp. 192-196.

[27] Q. Hongzhi, M. D. Xu Minpeng, W. Baikun, L. Zhipeng, and Y. Tao, "Study of channel selection based on AdaBoost SVM in P300 Speller," *Chinese Journal of Scientific Instrument,* vol. 26, pp. 53-57, 2012.

[28] F. Farooq and P. Kidmose, "Random forest classification for p300 based brain computer interface applications," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, 2013, pp. 1-5.

[29] J. Li, J. Lim, Y. Chen, K. Wong, N. Thakor, A. Bezerianos*, et al.*, "Mid-Task Break Improves Global Integration of Functional Connectivity in Lower Alpha Band," *Frontiers in human neuroscience,* vol. 10, 2016.

[30] J. Li, Y. Chen, F. Taya, J. Lim, K. Wong, Y. Sun*, et al.*, "A unified canonical correlation analysis-based framework for removing gradient artifact in concurrent EEG/fMRI recording and motion artifact in walking recording from EEG signal," *Medical & biological engineering & computing,* vol. 55, pp. 1669-1681, 2017.

[31] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *International Conference on Artificial Neural Networks*, 1997, pp. 999-1004.

[32] R. Bose, K. Samanta, and S. Chatterjee, "Cross-correlation based feature extraction from EMG signals for classification of neuro-muscular diseases," in *Intelligent Control Power and Instrumentation (ICICPI), International Conference on*, 2016, pp. 241-245.

[33] L. E. Peterson, "K-nearest neighbor," *Scholarpedia,* vol. 4, p. 1883, 2009.

[34] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of human genetics,* vol. 7, pp. 179-188, 1936.

[35] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and its Interface,* vol. 2, pp. 349-360, 2009.

[36] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* vol. 2, pp. 18-22, 2002.

[37] E. Fix and J. L. Hodges Jr, "Discriminatory analysis-nonparametric discrimination: consistency properties," California Univ Berkeley1951.

[38] M. J. Islam, Q. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed, "Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers," in *Convergence Information Technology, 2007. International Conference on*, 2007, pp. 1541-1546.

[39] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST),* vol. 2, p. 27, 2011.

[40] H. Rajaguru and S. K. Prabhakar, "Time frequency analysis (dB2 and dB4) for Epilepsy classification with LDA classifier," in *Communication and*

530   *Electronics Systems (ICCES), 2017 2nd International Conference on*, 2017, pp. 708-711.

[41]   L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[42]   M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and* 535   *Remote Sensing,* vol. 114, pp. 24-31, 2016.

[43]   B. Gokberk and L. Akarun, "Comparative analysis of decision-level fusion algorithms for 3D face recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 1018-1021.

[44]   M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need 540   hundreds of classifiers to solve real world classification problems?," *The Journal of Machine Learning Research,* vol. 15, pp. 3133-3181, 2014.

[45]   A. Kouzani, "Faceparts for recognition," in *TENCON 2006. 2006 IEEE Region 10 Conference*, 2006, pp. 1-4.

[46]   A. Z. Kouzani, "Subcellular localisation of proteins in fluorescent microscope 545   images using a random forest," in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, 2008, pp. 3926-3932.

[47]   B.-S. Yang, X. Di, and T. Han, "Random forests classifier for machine fault diagnosis," *Journal of mechanical science and technology,* vol. 22, pp. 1716- 550   1725, 2008.

[48]   K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall, "A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers," *Physiological measurement,* vol. 35, p. 2191, 2014.

555 [49]   O. G. Selfridge, "Pandemonium: a paradigm for learning in mechanisation of thought processes," 1958.

[50]   F. Castanedo, "A review of data fusion techniques," *The Scientific World Journal,* vol. 2013, 2013.

[51]   K. Veeramachaneni, L. Osadciw, A. Ross, and N. Srinivas, "Decision-level 560   fusion strategies for correlated biometric classifiers," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, 2008, pp. 1-6.

[52]   B. Jiang, B. Martinez, M. F. Valstar, and M. Pantic, "Decision level fusion of domain specific regions for facial action recognition," in *Pattern Recognition* 565   *(ICPR), 2014 22nd International Conference on*, 2014, pp. 1776-1781.

[53]   J. Harvy, E. Sigalas, N. Thakor, A. Bezerianos, and J. Li, "Performance Improvement of Driving Fatigue Identification Based on Power Spectra and Connectivity Using Feature Level and Decision Level Fusions," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and* 570   *Biology Society (EMBC)*, 2018, pp. 102-105.

[54]   C. R. Patti, S. S. Shahrbabaki, C. Dissanayaka, and D. Cvetkovic, "Application of random forest classifier for automatic sleep spindle detection," in *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*, 2015, pp. 1-4.

575 [55]   D. Milner and M. Goodale, *The visual brain in action*: Oxford University Press, 2006.

[56] K. H. James and L. Engelhardt, "The effects of handwriting experience on functional brain development in pre-literate children," *Trends in neuroscience and education,* vol. 1, pp. 32-42, 2012.

[57] D. Flowers, K. Jones, K. Noble, J. VanMeter, T. Zeffiro, F. Wood*, et al.*, "Attention to single letters activates left extrastriate cortex," *Neuroimage,* vol. 21, pp. 829-839, 2004.

[58] N. A. Busch, S. Debener, C. Kranczioch, A. K. Engel, and C. S. Herrmann, "Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response," *Clinical Neurophysiology,* vol. 115, pp. 1810-1820, 2004.

[59] C. Tallon-Baudry, O. Bertrand, C. Wienbruch, B. Ross, and C. Pantev, "Combined EEG and MEG recordings of visual 40 Hz responses to illusory triangles in human," *Neuroreport,* vol. 8, pp. 1103-1107, 1997.

[60] A. Ihara and R. Kakigi, "Oscillatory activity in the occipitotemporal area related to the visual perception of letters of a first/second language and pseudoletters," *Neuroimage,* vol. 29, pp. 789-796, 2006.

[61] K. Heimann, M. A. Umilta, and V. Gallese, "How the motor-cortex distinguishes among letters, unknown symbols and scribbles. A high density EEG study," *Neuropsychologia,* vol. 51, pp. 2833-2840, 2013.

[62] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, "Gamma, alpha, delta, and theta oscillations govern cognitive processes," *International journal of psychophysiology,* vol. 39, pp. 241-248, 2001.

[63] T. Demiralp, Z. Bayraktaroglu, D. Lenz, S. Junge, N. A. Busch, B. Maess*, et al.*, "Gamma amplitudes are coupled to theta phase in human EEG during visual perception," *International journal of psychophysiology,* vol. 64, pp. 24-30, 2007.

[64] V. Romei, J. Driver, P. G. Schyns, and G. Thut, "Rhythmic TMS over parietal cortex links distinct brain frequencies to global versus local visual processing," *Current Biology,* vol. 21, pp. 334-337, 2011.