Pers Ubiquit Comput (2018) 22:185–199
https://doi.org/10.1007/s00779-017-1083-4

ORIGINAL ARTICLE

Envisioned speech recognition using EEG sensors

Pradeep Kumar[1] · Rajkumar Saini[1] · Partha Pratim Roy[1] · Pawan Kumar Sahu[1] · Debi Prosad Dogra[2]

Received: 28 July 2017 / Accepted: 20 September 2017 / Published online: 30 September 2017

Abstract Recent advances in EEG technology makes brain-computer-interface (BCI) an exciting field of research. BCI is primarily used to adopt with the paralyzed human body parts. However, BCI in envisioned speech recognition using electroencephalogram (EEG) signals has not been studied in details. Therefore, developing robust speech recognition system using EEG signals was proposed. In this paper, we propose a coarse-to-fine-level envisioned speech recognition framework with the help of EEG signals that can be thought of as a serious contribution in this field of research. Coarse-level classification is used to differentiate/categorize text and non-text classes using random forest (RF) classifier. Next, a finer-level imagined speech recognition of each class has been carried out. EEG data of 30 text and not-text classes including characters, digits, and object images have been imagined by 23 participants in this study. A recognition accuracy of 85.20 and 67.03% has been recorded at coarse- and fine-level classifications, respectively. The proposed framework outperforms the existing research work in terms of accuracy. We also show the robustness in envisioned speech recognition.

Keywords Envisioned speech · Random forest · EEG signals · Assistive technology · Electroencephalography (EEG)

✉ Pradeep Kumar
pradeep.iitr7@gmail.com

Rajkumar Saini
rajkumarsaini.rs@gmail.com

Partha Pratim Roy
proy.fcs@iitr.ac.in

Pawan Kumar Sahu
pawansahu.44@gmail.com

Debi Prosad Dogra
dpdogra@iitbbs.ac.in

[1] Department of Computer Science and Engineering, IIT Roorkee, Roorkee, India

[2] School of Electrical Sciences, IIT Bhubaneswar, Bhubaneswar, India

1 Introduction

Communication is an act through which one can express his/her feelings to others. Normal people could communicate easily with the help of commonly understood words and signs [1]. These words and signs refer to a language or sign language. However, there are people, such as patients suffering from locked-in syndrome condition, who are not able to communicate properly. Though these patients cannot perform any motory actions such as hand or foot movements, including vocalization of words, they are aware of their cognitive aspects. They can understand the nearby ambiance because of proper functioning of their cognitive system [2]. Despite such abnormal conditions, their brain and cognitive systems work properly. With the advancement in sensor technology, it is now possible to develop smart applications that can help us to manage, control, and automate living environment without manual intervention. Internet of things (IoT) is an example where sensors and actuators are embedded in a physical object and linked through wired or wireless networks for communication. IoT techniques are popularly used in security, gaming, home automation, and entertainment applications [3, 4]. Many researchers are working towards various rehabilitation systems to make them communicate and express their feelings through acquisition and

interpretation of brain signals using multiple sensors [5]. This process of acquisition of brain signals and analyzing them to figure out the speech is known as speech imagery [6]. These brain signals could be recorded by various ways, which include invasive and non-invasive methods.

In invasive methods, one has to implant sensors in the subject's brain to record their brain signals. These invasive methods are expensive and difficult to implement as the sensors need to be implanted inside the brain, whereas non-invasive methods do not require implantation of the sensors [7]. In non-invasive ways, signals can be recorded simply by placing the sensors over the scalp of the subject's brain. These methods are economical, handy, and straightforward to implement [8].

With the recent enhancement in cognitive technology and research in brain-computer-interface (BCI), it is possible to study the state of brain through advanced brain sensing techniques [9]. These days, it is possible to capture a particular state of neurons and electromagnetic field created inside the brain. This has been made possible due to availability of electroencephalography (EEG) [10, 11], magnetoencephalography (MEG) [12], and functional magnetic resonance imaging (fMRI) [13]. Among these technologies, EEG has advantages over others due to lower cost, portability, wireless connectivity, and easy handling. Moreover, the technology is popular because of its non-invasive nature and is successfully used in various BCI applications such as emotion recognition [14], stress identification [15, 16], and biometrics [7]. In EEG, small flat metal discs known as electrodes are placed on the scalp to record electrical activities inside the brain as a form of EEG signals. The acquisition of EEG signals could be done with the help of devices such as Emotiv EPOC+ or similar setup.

In this paper, we propose a framework for capturing the thoughts of physically impaired patients with the help of EEG technology and make them able to express their feelings and communicate to others. The work can be helpful in other areas of research such as synthetic telepathy [1], inside library where silent communication is recommended, and in situations where normal communication may not be possible such as fire places, and inside water communication. In literature, some work exist [17–19] that exploit the recognition of thoughts from the human brain. Majority of these work are confined to recognition of syllables such as "ba" and "ku" or alphabets such as "a" and "u" [1, 18, 20, 21]. Therefore, we intend to introduce a new approach to recognize the envisioned speech.

The main contributions of the paper are as follows:

- Firstly, we present envisioned speech recognition framework to predict digits, characters, and images while they were imagined by users in eyes closed resting state using EEG signals.
- Secondly, we propose a coarse-to-finer-level classification approach. Coarse-level is used to predict the category of the envisioned speech, whereas finer-level classification predicts the actual class of the predicted category.
- Thirdly, we demonstrate a detailed analysis of the proposed approach with comparative analysis against existing state-of-the-art techniques.

The rest of the paper is organized as follows. In Section 2, we discuss the related work on envisioned speech systems. Section 3 describes the proposed methodology of recognition of envisioned speech followed by the description of pre-processing, feature extraction techniques, and classification model. Results are explained in Section 4. Finally, we conclude in Section 5 by discussing the future possibilities of the work.

2 Related work

Recent advances in BCI demonstrate the robustness of brain signals in decoding various mental tasks such as imagined speech [18, 19], object understanding, sleep stages [22], and person identification systems [7]. Costa et al. [17] proposed a mental task classification system to discriminate left and right hand movements using EEG signals. The authors have extracted adaptive Gaussian representation (AGR) coefficients from the brain signals of 10 volunteers and fed them in a multilayer perceptron (MLP) neural network classifier for recognition purpose. An average accuracy of 91 and 87% has been recorded for female and male participants, respectively. The authors in [23] have applied common spatial subspace decomposition (CSSD) filter to separate EEG components and eliminate the noise from the acquired signals. Next, spatio-temporal features have been extracted for classification purpose using hidden Markov model (HMM). A BCI system for response error correction has been proposed in [19] to detect error-related negativity (ERN) in a visual discrimination task. ERN detection has been performed using DWT-based decomposition of EEG signals and a Gaussian classifier. The detected ERN has been used to correct subject errors.

DaSalla et al. [18] have proposed a BCI system to recognize English vowels using EEG signals of three subjects. The study has been conducted for three classes including two vowels and no-action state. A zero-phase bandpass filter has been applied within a frequency band of 1–45 Hz to remove lower frequencies and electronic noises. Common spatial patterns (CSP) has been applied to the EEG signals to generate new time series. An average accuracy of 71% has been recorded in all three classes using support vector machine (SVM) classifier. In [24], a second-order blind identification (SOBI) algorithm has been used

to remove noises and to decompose the raw EEG signals into mutually orthogonal components. The authors have utilized Hilbert-Huang transformation (HHT) to extract joint temporal and spectral features. Bayesian classifier based on multi-class linear discriminant analysis (LDA) has been used for the classification purpose, where the average accuracy of 58.05% has been recorded. In [25], the authors have proposed an imaginary BCI system to distinguish five primitive shapes using EEG. ICA algorithm has been applied on EEG data to remove the artifacts. Five frequency bands have been extracted from each independent component by applying HHT. The Mann-Whitney-Wilcoxon (MWW) test has been applied for feature selection to rank the EEG channels. An average accuracy of 44.6% has been recorded using LDA classifier. In [21], the authors have proposed an imagined speech classification system for two Chinese characters using EEG signals. CSP has been applied to extract features from the acquired signals and are directly fed to the SVM classifier for recognition purposes. An average accuracy of 66.87% has been recorded on EEG data of eight participants. In [11], the authors have applied CSP and relevance vector machine (RVM) with Gaussian kernel for the classification of imagined Japanese vowels.

Recently, Torres-García et al. [26] have proposed a methodology for channel selection and classification of imagined speech using EEG signals. The selection of channels have been performed using fuzzy inference system (FIS) whose objectives are to minimize the number of channels and the error rate. DWT analysis has been performed on EEG data and five features including four statistical values and the relative wavelet energy (RWE) have been extracted for classification. An accuracy of 68.18% has been recorded on seven channels using random forest classifier. In [27], the authors have applied sonification and textification techniques for the classification of EEG signals recorded while imagining speech. Their dataset consists of five Spanish words that have been imagined by 27 participants and corresponding EEG responses have been recorded simultaneously. The common average reference (CAR) method has been applied to improve signal-to-noise ratio of the EEG. In sonification, EEG frequencies have been scaled down to audible range using fast fourier transform (FFT), whereas EEG textification has been carried out based on higher energy zones and the recognition of imagined speech has been done using SVM and naive Bayes classifiers. Wang et al. [28] have applied convolutional neural network (CNN) to extract spatial and frequency features from the EEG signals for the classification of imagined speech. The authors have also applied recurrent neural network (RNN) to extract temporal features from the CNN feature sequence and connectionist temporal classification (CTC) layer has been used to output the most possible phone sequence of the input imagined speech EEG signal. However, their dataset consists of only two classes of imagined vowels. Similarly, the authors in [29] have proposed the methodology for imagined speech recognition based on covariance matrix descriptors using Riemannian manifold, and RVM classifier has been used for the classification purpose.

In addition, the EEG signals are also widely used to develop various BCI applications including gaming [8], biometrics [7, 9], emotion recognition [14, 30], stress identification [15, 16], or prediction analysis [31]. For example, the authors in [15] have analyzed the spectrum power of EEG bands recorded from prefrontal sites to identify chronic stress between two groups of human subjects. The authors have found that people with chronic stress have higher left prefrontal power. Similarly, Gauba et al. [31] have proposed a multimodal rating prediction framework of advertisement videos using sentiment analysis and EEG signals. Random forest regression-based technique has been used to estimate the rating of an advertisement video.

3 Proposed methodology

In our framework, a wireless neuro-headset known as Emotiv EPOC+ [8] has been used for acquisition of envisioned brain signals. The headset and its associated accessories are shown in Fig. 1.

For recording brain signals, this device incorporates 14 channels namely AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, which are placed over the scalp of the user according to International 10-20 system as shown in Fig. 2, where two reference electrodes, i.e., CMS and DRL, are positioned above the ears.

It captures the brain waves at a frequency of 2048 Hz and later downsamples them to 128 Hz per channel [9]. The captured brain waves are sent to the computing device through Bluetooth technology.

The flow diagram of the proposed envisioned speech recognition framework is depicted in Fig. 3, where a participant is watching an object on the computer screen while wearing the EEG headset.

Next, the participant is asked to imagine the viewed object for 10 s while eyes closed and resting state. A gap of 20 s has been considered before showing the next item to the participant. This gap has been used such that the participant do return to the resting state before thinking of the next item. Similarly, all items have been shown and corresponding envisioned brain waves have been recorded. Next, signal smoothing and feature extraction techniques have been applied to interpret the signals. In this work, three categories of object have been shown to the participants, namely digits, characters, and object images. A coarse-level classification has been applied to recognize one of the three categories using RF classifier. After category classification,

Fig. 1 Details of the Emotiv EPOC+ headset sensor and the associated accessories

a finer-level classification has been applied to each category for recognition of actual object types that are imagined by the user.

3.1 Preprocessing the signals

Since raw EEG signals are usually corrupted with various types of noises, trends, and artifacts that occur due to eye blinks, muscular activities or various electrical noises, we have applied signal smoothing technique using Moving Average (MA) filter to remove the effect of such unwanted artifacts before extracting features [31].

Moving Average (MA) filter It is often used for signal smoothing that simply replaces each data value with the average of neighboring values. The moving average is calculated using (1), where $x[.]$, $\bar{y}[.]$ show the input and output signal, respectively. M represents the number of points in the average. We have applied it assuming $M = 5$. The impact of smoothing on a raw EEG signal is shown in Fig. 4, where Fig. 4a shows the raw EEG signal and the corresponding smoothed signal is shown in Fig. 4b.

$$\bar{y}[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j] \tag{1}$$

3.2 Feature extraction

Next, from each smoothed signal, four different features, namely Standard Deviation (SD), Root Mean Square (RMS), Sum of Values (SUM), and Energy (E) are computed on all 14 electrodes. Hence a new feature vector S(F) of 56 (14X4) dimensions is formed as defined in (2).

$$S(F) = \{SD, RMS, SUM, E\} \tag{2}$$

– **Standard Deviation (SD)** : SD has been successfully used for EEG signals to estimate the variation present in signals [9] and it has been computed using (3), where x, $\bar{x}$ and n represent the raw signal, mean, and number of elements in the signal. Standard deviation corresponds

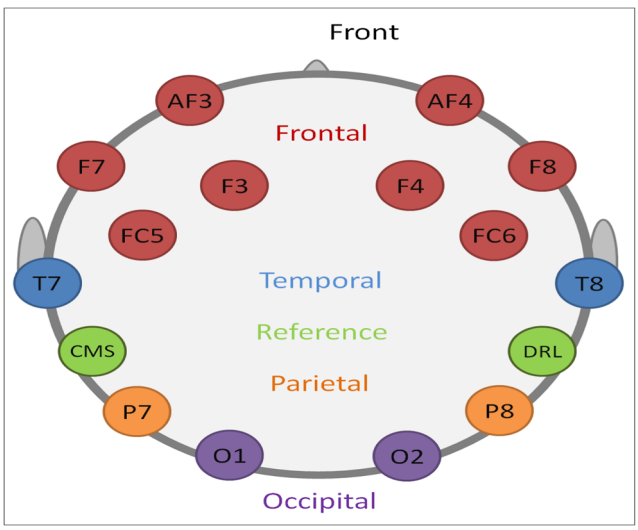


Fig. 2 Pictorial representation of the International 10-20 system

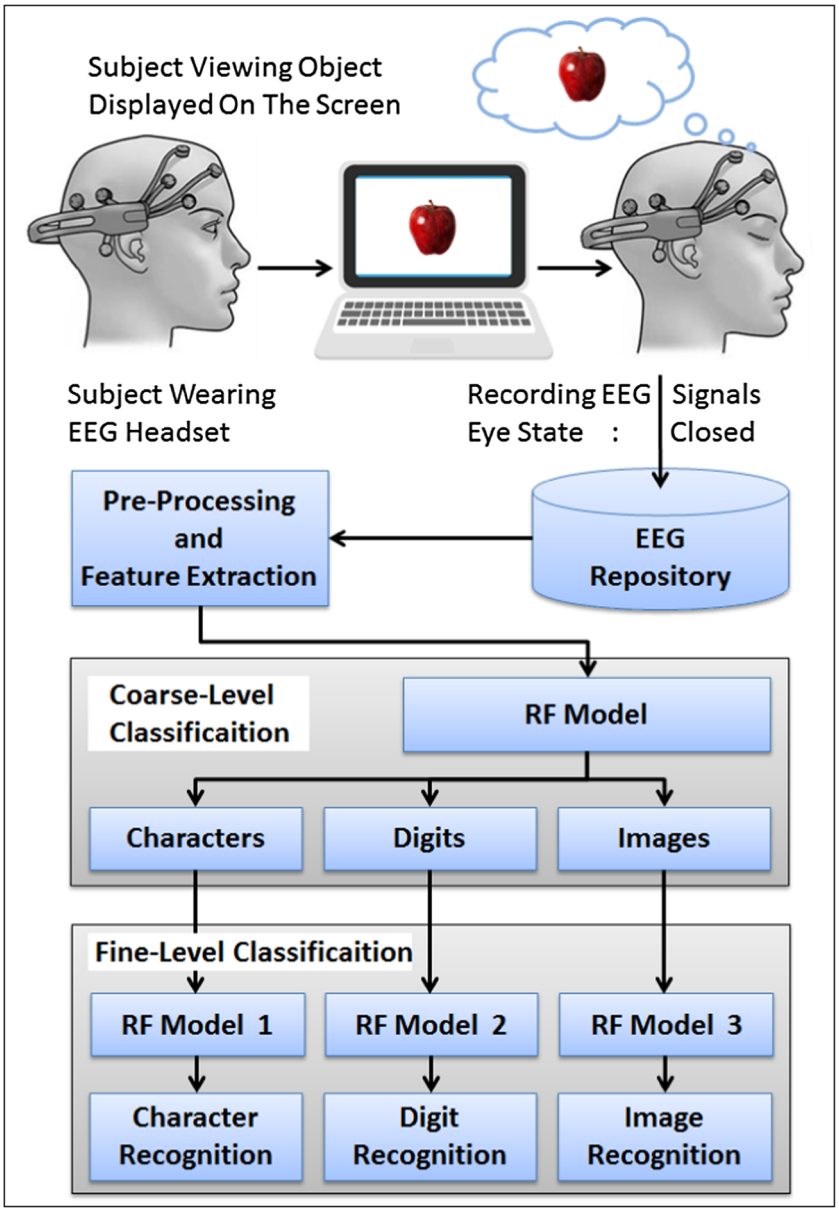


Fig. 3 Flow diagram of the proposed envisioned speech recognition framework

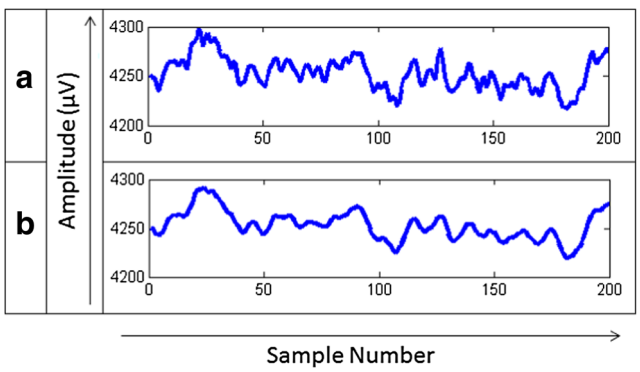


Fig. 4 EEG signal smoothing using MA filter. **a** Raw signals. **b** Smoothed signal

to each frequency band is represented by SD_k with $k = 1, .., 14$, where i represents a signal from one electrode.

$$SD = \left(\frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2 \right)^{1/2} \quad (3)$$

– **Root Mean Square** : RMS is a one-dimensional feature vector for each electrode. It is defined as the square root of the mean of squared values of the quantity taken over an interval [7]. The value can be calculated using (4), where n is the number of samples and x is the input vector for the i^{th} input's amplitude.

$$RMS = \left(\frac{1}{n} \sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \quad (4)$$

– **Sum of Value (SUM):** It is calculated by taking the sum of all values present in a signal as define in (5), where x_i is the i^{th} element of the signal and n denotes the length of the signal.

$$SUM = \sum_{i=1}^{n} x_i \quad (5)$$

– **ENERGY** (E): Energy (E) of a discrete time signal x can be calculated using (6), where x_i is the i^{th} element of the signal and n denotes the length of the signal.

$$E = \sum_{i=1}^{n} (x_i^2) \quad (6)$$

3.3 Envisioned speech recognition using random forest (RF) classifier

Random Forest (RF) has been successfully used in various EEG signal classification problems [32–34]. The classifier basically develops a group of decision trees and grants them to vote for the most suitable class [35]. A set of features are chosen randomly and a classifier is created with a bootstrapped sample of the training data. Each tree casts a unit vote for the most popular class to classify an input vector. Then, it combines the decision of a set of classifiers by weighted or unweighted voting to classify unknown examples. Thus, a large number of trees (classifiers) are generated and finally majority voting technique is used to assign a class to the test sample.

In this work, we have used the RF classifier twice to classify EEG signals at coarse level and fine level. The two levels of classification can be seen as a hierarchy as depicted in Fig. 5, where coarse level consist of three classes (i.e., character, image, and digit) and fine level denotes the actual recognition of the classes.

RF consisting of randomly selected features or a combination of features at each node was used to grow a tree. Bagging technique has been used to generate the training dataset by randomly drawing examples with replacement from the original training set. A test sample has been classified by taking the most popular voted class from all tree predictors in the forest. The design of the decision tree requires the choice of an attribute selection measure and a pruning method. Therefore, Gini index (GI) has been used as an attribute selection measure, which measures the impurity of an attribute with respect to the classes [31]. Thus, at coarse level, for a given training set T, selecting one EEG sample at random and saying that it belongs to class C_{coarse} (here $coarse$ denotes one of three classes at coarse level), GI can be calculated using (7), where $\frac{f(C_{coarse},T)}{|T|}$ is the probability that the selected case belongs to class C_{coarse}.

$$\sum \sum_{j \neq coarse} \left(\frac{f(C_{coarse},T)}{|T|} \right) \left(\frac{f(C_j,T)}{|T|} \right) \quad (7)$$

Similarly, at fine level, there are ten classes for each coarse-level class. Thus, for training set M, selecting one EEG sample at random and saying that it belongs to class C^f_{fine} (here $fine$ represents one of ten classes at fine level and f denotes one of three coarse-level classes), the GI can

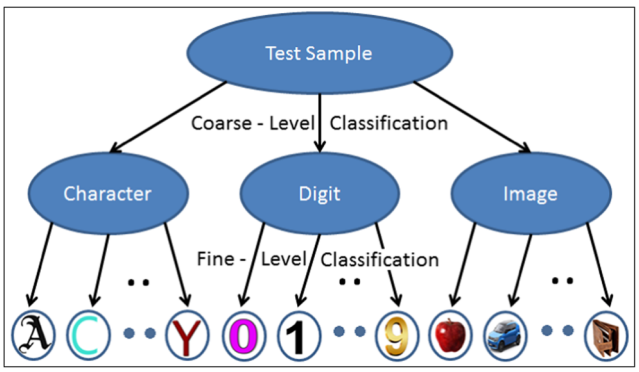


Fig. 5 Pictorial representation of coarse and fine-level classification of imagined speech using RF

be calculated using (8), where $\frac{f(C^f_{fine},M)}{|M|}$ is the probability that the selected case belongs to class C^f_{fine}.

$$\sum \sum_{g \neq fine^f} \left(\frac{f(C^f_{fine},M)}{|M|} \right) \left(\frac{f(C_g,M)}{|M|} \right) \tag{8}$$

4 Experiments and discussions

This section presents the description of the dataset and experimental results. Results have been obtained in two phases. First, a coarse-level classification has been performed to recognize the three categories of the envisioned items. Next, finer-level recognition results have been computed to recognize the actual items of each category. Final results have been obtained using a 10-fold cross validation scheme. According to this scheme, the dataset is divided into 10 equal parts. Out of which, nine parts are used in training and the rest are used in testing. Similarly, all the parts are tested one-by-one and the average results are reported.

4.1 Dataset collection

In this study, 23 participants aged between 15 to 40 years have been enrolled to collect data. Our aim is to include a variety of participants with varying age. EEG recordings have been performed using Emotiv EPOC+ sensor. All participants are university students and have been requested to remain calm during the whole process with clear thoughts. Moreover, all of them have been requested not to consume caffeine or alcohol and not to smoke prior to the recording process to avoid any effects of these substances on the nervous system. A slide presentation was prepared that consisted of 20 text and 10 non-text items. Out of 20 text classes, 10 slides are of digits from 0–9, whereas the rest of the slides consist of 10 character images. A pictorial representation of 20 text classes is depicted in Fig. 6.

Similarly, slides of non-text classes consists of different images that are commonly used in daily life routine as shown in Fig. 6, have been prepared. From the dataset, each slide has been shown to every participant for 10 s. Next, the participant has been asked to envision the shown item for 10 s in eyes closed resting state. Between two successive recordings, a gap of 20 s has been introduced to clear the previous imaginary thoughts of participant. Using this protocol, 690 (i.e. 23×30) EEG recordings have been collected. To analyze the EEG signals, each recording has been split into multiple signals for coarse and fine-level classification with different time duration of 250 ms and 50 ms, respectively. Therefore, a total of 138,000 and 27,600 recording of EEG signals have been analyzed in this study at coarse and fine-level classification, respectively. Variations in the EEG recordings of a user can be seen in Fig. 7 when imagined different class types.

Fig. 6 Pictorial representation of all text and non-text classes involved in the dataset. **a** Character classes. **b** Digit classes. **c** Object images

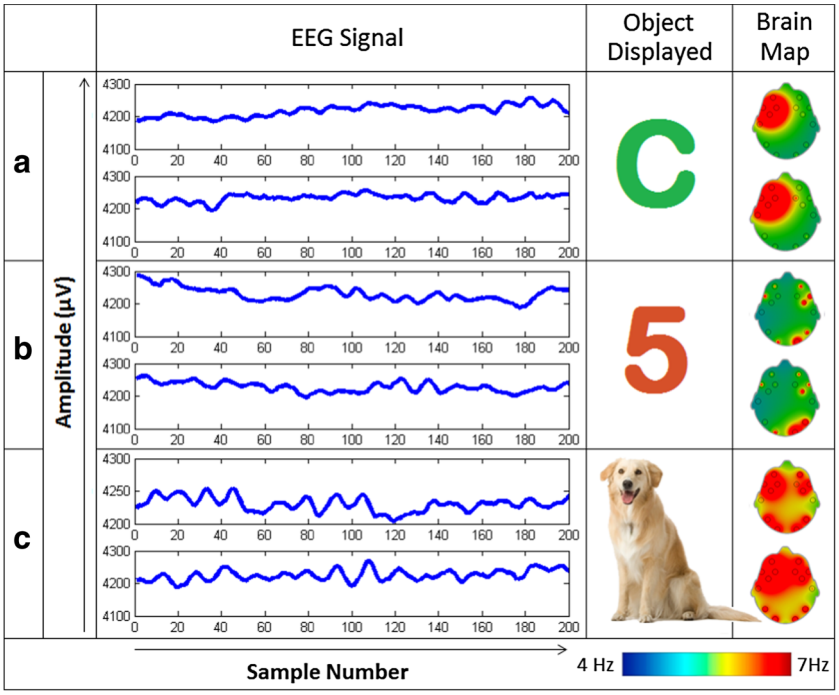


Fig. 7 Variations in the EEG signals of a user when imagined different class types

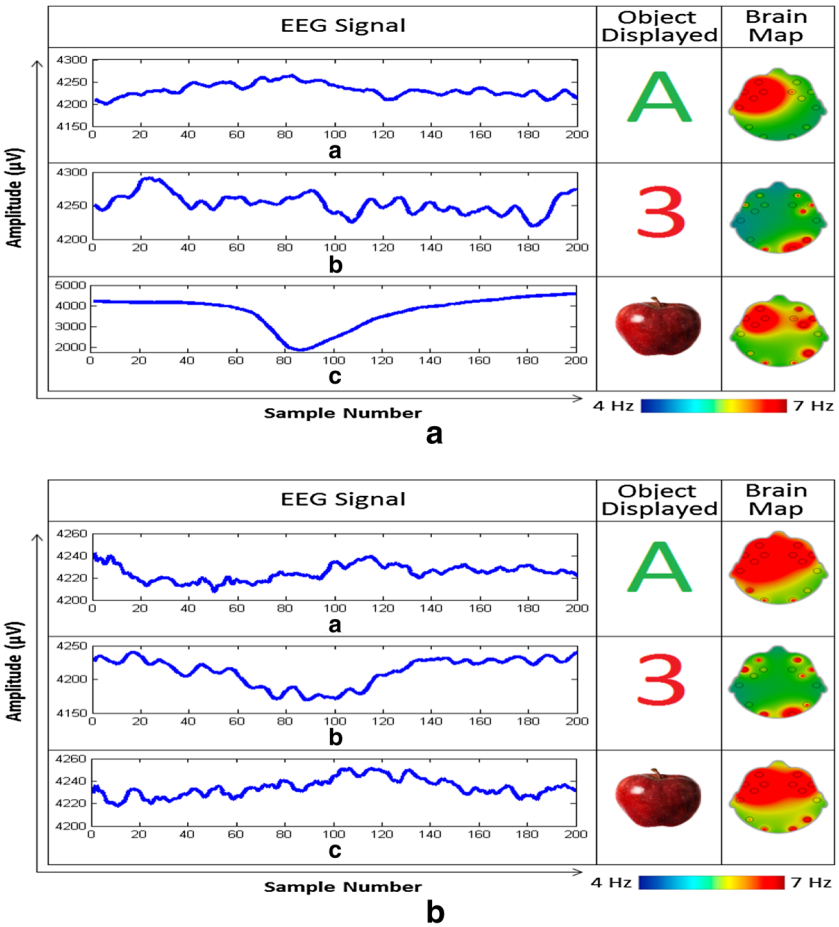


Fig. 8 Variations in the EEG signals when imagined same set of classes by two different users: (**a**) imagined by user U1 (**b**) imagined by user U2

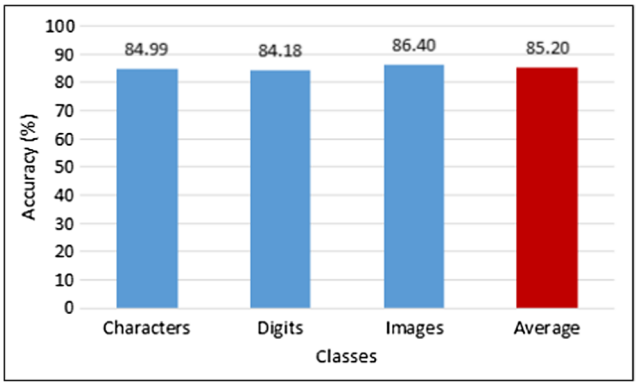
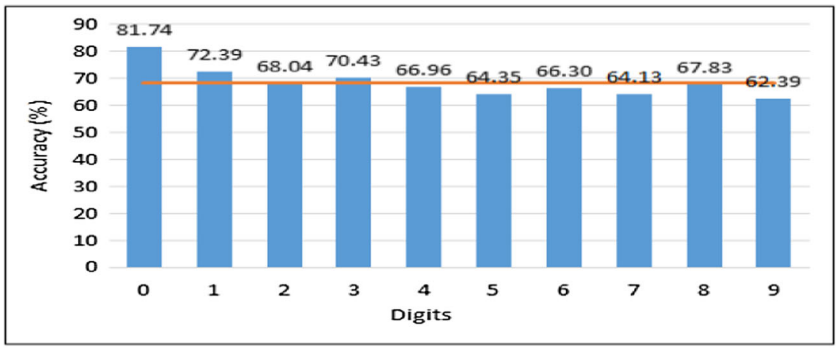
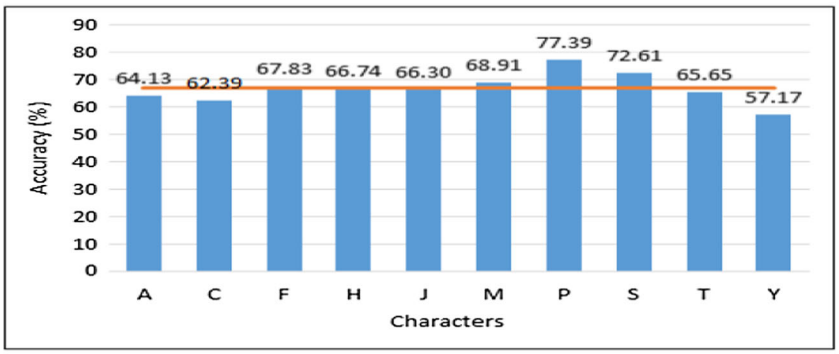


Fig. 9 Coarse-level classification results of three different type of imagined classes

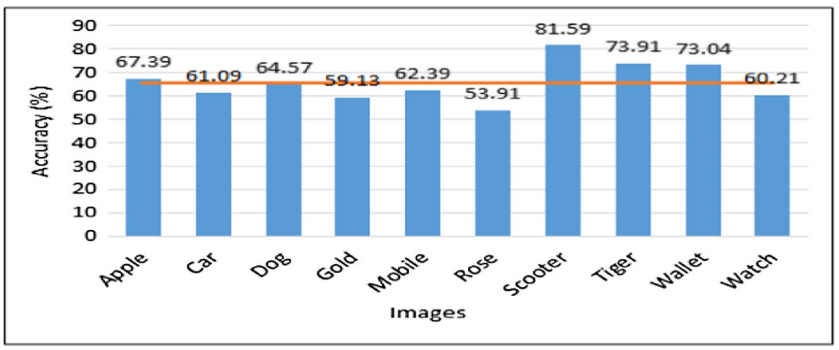
Fig. 10 Fine-level recognition accuracies for each imagined class items. **a** Imagined digit. **b** Imagined characters. **c** Imagined images



a



b



c

Fig. 11 An average recognition rates of fine-level classification of all three classes

Similarly, the recordings of two users (U1 and U2) are depicted in Fig. 8 when imagined same set of test and non-text types.

4.2 Coarse-level classification results

In coarse-level classification, text and non-text type imagination has been recognized which includes character, digits, and object images. This classification ease the process of recognition at fine-level when the actual recognition of envision speech takes place. Classification has been performed using EEG signals. The results are computed using RF classifier by varying the number of trees in the forest. Classification results for the three classes are depicted in Fig. 9, where an accuracy of 85.20% has been recorded with 32 number of trees. The accuracy for each text and non-text class has also been evaluated as shown in Fig. 9. It can be observed from the figure that the maximum accuracy is recorded for object images class with 86.40%. Next, these results are used to compute the recognition at fine-level classification.

4.3 Fine-level classification results

Here, the actual recognition of envision speech takes place. Recognition has been performed on the EEG signals classified at coarse-level in one of the three classes. Recognition has been performed using three parallel RF classifiers by varying the number of trees in the forest from 1 to 50. Recognition accuracies of the envision speech for each item of all the three classes is shown in Fig. 10, where Fig. 10a depicts the performance of imagined digits, Fig. 10b shows the accuracy of imagined characters, and Fig. 10c depicts the recognition rate of imagined images of various objects. A horizontal line has been drawn in each figure to show the average performance of the system. It can be seen from the Fig. 10a that a maximum accuracy of 81.74% has been recorded for "zero" ("0") in digit class whereas an accuracies of 77.39 and 81.59% have been recorded for "P" and "Scooter" classes in characters and images categories, respectively.

A comparative performance analysis between the average recognition rates of all the three classes is shown in Fig. 11, where a maximum accuracy of 68.46% has been recorded with the EEG signals recorded for imagined digits at 40 number of trees, whereas an accuracy of 66.91 and 65.72% has been recorded on characters and object images with 23 and 36 number of trees, respectively. Thus, an average recognition performance of 67.03% for all the three categories has been recorded at this level of classification as shown in the Fig. 11.

Fig. 12 Recognition accuracies with different top choices

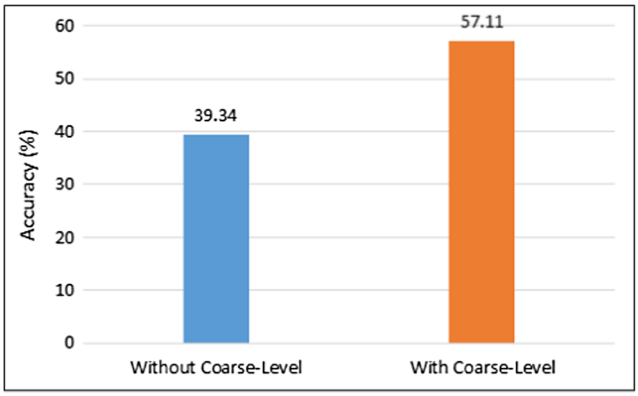


Fig. 13 Comparative performance analysis with and without performing coarse-level classification

Accuracy with different top choices of confidence has also been computed for all the three class types for top five choices. According to this scheme, top five ranked results as per their confidence have been analyzed for each test sample and when the correct label of test sample is found in the top rank, it is considered as duly classified. The scheme increases the classification performance and reduces the chance of errors. Recognition results for the top five choices are depicted in Fig. 12.

In order to show the effectiveness of the proposed framework, performance of the system has been computed with and without performing coarse-level classification. Total accuracy of the system with coarse-level classification is calculated using (9), where CLA, DA, CA, and IA are the accuracy of coarse level, digits, characters, and images, respectively. In this work, the values of CLA, DA, CA and IA are 85.20, 68.46, 66.91, and 65.72%, respectively. Total accuracy of the system is depicted in Fig. 13, where the proposed method outperforms the recognition performance without coarse-level classification by a margin of 17.77%.

$$Total-accuracy = CLA * \left(\frac{DA + CA + IA}{3} \right) * \frac{1}{100} \quad (9)$$

Experiments have also been conducted to find the dominating EEG channels corresponding to different brain lobes for the recognition of imagined speech. Performance has been measured according to four different brain lobes, namely frontal (AF3,AF4,F3,F4,F7,F8,FC5,FC6), parietal (P7,P8), occipital (O1,O2), and temporal (T7,T8). Recognition performance for all brain lobes corresponding to each class are presented in Table 1, where the maximum accuracy of 85.20 and 67.03% have been recorded for coarse- and fine-level classification, respectively, considering all 14 channels. However, when comparing the performance

Table 1 Recognition accuracies at coarse and fine-level of the system with electrodes located at different brain lobes

Brain lobe	Electrodes	Coarse-level accuracy (%)	Fine-level accuracy (%)
Frontal	AF3, AF4, F3, F4, F7, F8, FC5, FC6	72.8632	55.9130 digit 57.7609 character 52.4783 image
Temporal	T7, T8	61.4011	15.4348 digit 18.5435 character 14.00 image
Parietal	P7, P8	52.0299	17.6957 digit 21.1739 character 16.1739 image
Occipital	O1, O2	53.1746	15.8696 digit 18.3478 character 15.9783 image
All	**AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, O2**	**85.20**	**68.4565 digit** **66.9130 character** **65.6304 image**

Bold emphasis has been used to make a difference between existing techniques and the proposed methodology

Table 2 Details of the age groups

Age group	Years	Number of participants
G1	15–20	8
G2	22–27	10
G3	30–40	5

at different brain lobes, maximum accuracies have been recorded on the frontal lobe in both coarse- and fine-level classifications.

4.3.1 Impact of aging on envisioned speech recognition

To find out the impact of aging on the proposed framework, we have divided the dataset into three subsets. The details of each subset is presented in Table 2.

Recognition accuracy of all the three classes for each age group have been computed separately. Results are depicted in Fig. 14, where average accuracy of 63, 74.25, and 73.23% are recorded with G1, G2, and G3 age groups, respectively.

It can be noted that lower recognition rates have been recorded for the G1 group. It may be because of very young participants with age between 15–20 years who are engaged with multiple activities inside the University campus. Therefore, such users may not have envisioned the displayed items in a well defined manner.

4.3.2 Analysis of EEG duration for envision speech recognition

In this section, we analyze the optimized duration of EEG signals that is necessary for recognition of envisioned speech. For this, the performance of the system has been computed at coarse as well as finer levels by varying EEG signals with different time durations. Average recognition rates for both types of classification are depicted in Fig. 15, where highest accuracy of 85.20 and 67.03% have been recorded for coarse- and finer-level classification on EEG signals with durations of 250 and 50 ms, respectively.

4.4 Comparative analysis

A comparative analysis has also been performed with classifiers such as artificial neural network (ANN) [36] and SVM [37]. These classifiers are widely used by the researchers for modeling EEG signals in various BCI applications [7, 9, 38].

SVM is a kernel-based classifier and is used to perform both linear and non-linear classifications [37]. The main function of the classifier is to map the data into feature space where a hyperplane separates the classes. The general solution is presented in (10), where k is the kernel function.

$$f(x) = \sum_i a_i y_i k(x_i, x) \tag{10}$$

Similarly, ANN is a computational model based on the structure and functions of biological neural networks [38, 39]. Here, a feed-forward neural network with two hidden and one output layer has been implemented for the recognition of imagined speech. The training has been performed with error back-propagation algorithm and Sigmoid function as the activation function for all units. All weights have been initialized with small random values and a gradient-descent search in the networks weight space has been used for a minimum of squared error function of the networks output. The recognition performance of both SVM and ANN for all 30 classes are depicted in Fig. 16, where the proposed RF classifier based recognition accuracy outperforms the other two classifiers. It is because RF classifier

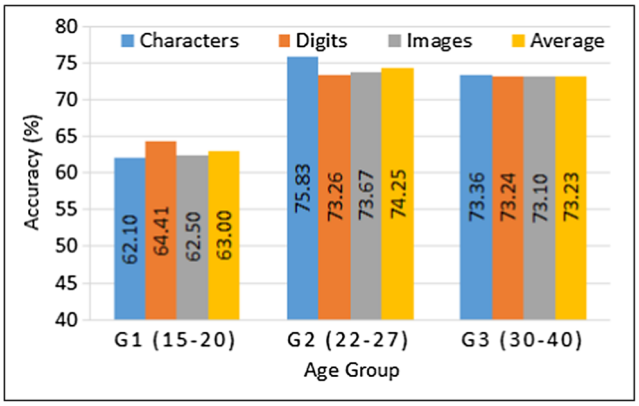


Fig. 14 Recognition performance of the system with participants of different ages

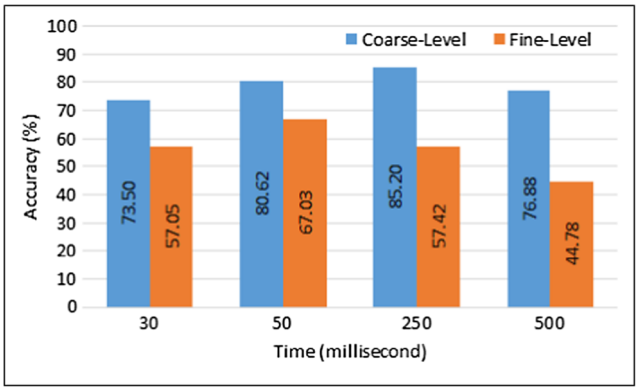


Fig. 15 Analysis for the optimized duration of EEG signals for imagined speech recognition

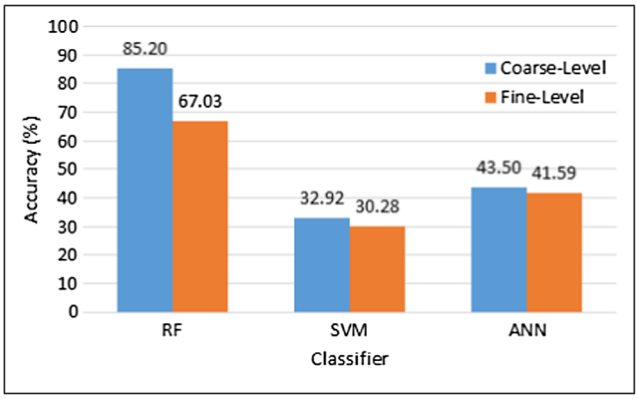


Fig. 16 Comparative performance analysis of coarse and fine-level classification with other classifiers

Table 3 Comparative analysis with state-of-the-art techniques

Methods	Approach	Number of Classes	Number of Subjects	Accuracy (%)
Esfahani et al. [25], 2012	EEG, ICA, HHT, MWW, LDA	5 3D shapes	10	44.6
Matsumoto et al. [11], 2014	EEG, RVM-G, CSP, AC	5 Japanese vowels	5	79
Proposed Methodology	**EEG, MA filter, SD, RMS, ENERGY, Coarse-to-fine level using RF classifier**	**10 digits 10 characters 10 objects**	**23**	**85.20 (coarse-level) 67.03 (fine-level)**

Bold emphasis has been used to make a difference between existing techniques and the proposed methodology

uses bootstrap aggregation or bagging ensemble technique to create multiple models and then combine them to produce improved results. Usually, ensemble methods results in more accurate solutions than a single model would; therefore, in this work, we have recorded higher accuracy using RF in comparison to other classification models.

In addition, we have compared the proposed framework of imagined speech recognition with existing technique [25]. Esfahani et al. [25] have proposed a imaginary BCI system to distinguish five 3D objects using EEG signals. They have applied ICA algorithm on EEG data for removing the artifacts and five frequency bands have been extracted from each independent component by applying HHT. A Mann-Whitney-Wilcoxon (MWW) test has been conducted for feature selection to rank the EEG channels. Similarly, the authors in [11] have proposed classification of five Japanese language vowels using imagined EEG signals. They have applied the relevance vector machine with Gaussian kernel (RVM-G) for classification with common special patterns (CSP) filtering and adaptive collection (AC) to divide the data into small segments. The comparison is shown in Table 3, where the proposed work has a significant improvement of performance though more classes of objects from large number of users are considered.

5 Conclusion

In this paper, we have proposed a new coarse-to-finer-level framework for envisioned speech recognition to assist the speech impaired people using EEG signals. A dataset of EEG signals has been recorded using 30 text and non-text class objects being imagined by multiple users. Recognition of both coarse and finer levels has been performed using RF classifier, where an accuracy of 85.20 and 67.03% have been recorded, respectively. The overall accuracy of the system has been recorded as high as 57.11% when computed using the RF classifier. In the future, more robust features can be extracted from the EEG signals to improve the recognition performance of the system. Moreover, the framework can be extended to develop various BCI applications including synthetic telepathy and rehabilitation systems.

References

1. Brigham Katharine, Vijaya Kumar BVK (2010) Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In: 4th international conference on bioinformatics and biomedical engineering, pp 1–4
2. Pineda JA, Allison BZ, Vankov A (2000) The effects of self-movement, observation, and imagination on/spl mu/rhythms and readiness potentials (RP's): toward a brain-computer interface (BCI). IEEE Trans Rehabil Eng 8(2):219–222
3. Jara AJ, Lopez P, Fernandez D, Castillo JF, Zamora MA, Skarmeta AF (2014) Mobile discovery: discovering and interacting with the world through the internet of things. Pers Ubiquit Comput 18(2):323–338
4. Han K, Kim J, Shon T, Ko D (2013) A novel secure key paring protocol for rf4ce ubiquitous smart home systems. Pers Ubiquit Comput 17(5):945–949
5. Metsis V, Kosmopoulos D, Athitsos V, Makedon F (2014) Non-invasive analysis of sleep patterns via multimodal sensor input. Pers Ubiquit Comput 18(1):19–26
6. Pei X, Hill J, Schalk G (2012) Silent communication: toward using brain signals. IEEE Pulse 3(1):43–46
7. Kaur B, Singh D, Roy PP A novel framework of EEG-based user identification by analyzing music-listening behavior. Multimedia Tools and Applications 1–22. https://doi.org/10.1007/s11042-016-4232-2
8. Badcock NA, Mousikou P, Mahajan Y, de Lissa P, Thie J, McArthur G (2013) Validation of the emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. PeerJ 1:e38
9. Kumar P, Saini R, Roy PP, Dogra DP (2017) A bio-signal based framework to secure mobile devices. J Netw Comput Appl 89:62–71
10. Gandhi T, Panigrahi BK, Anand S (2011) A comparative study of wavelet families for EEG signal classification. Neurocomputing 74(17):3051–3057
11. Matsumoto M, Hori J (2014) Classification of silent speech using support vector machine and relevance vector machine. Appl Soft Comput 20:95–102
12. Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. J Cogn Neurosci 14(8):1125–1138
13. Price CJ (2012) A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62(2):816–847
14. Kanjo E, Al-Husain L, Chamberlain A (2015) Emotions in context: examining pervasive affective sensing systems, applications, and analyses. Pers Ubiquit Comput 19(7):1197–1212
15. Peng H, Bin H, Zheng F, Fan D, Zhao W, Chen X, Yang Y, Cai Q (2013) A method of identifying chronic stress by EEG. Pers Ubiquit Comput 17(7):1341–1347
16. Menezes MLR, Samara A, Galway L, SantAnna A, Verikas A, Alonso-Fernandez F, Wang H, Bond R (2017) Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset. Pers Ubiquit Comput 1–11. https://doi.org/10.1007/s00779-017-1072-7
17. Costa EJX, Cabral EF (2000) EEG-based discrimination between imagination of left and right hand movements using adaptive gaussian representation. Med Eng Phys 22(5):345–348
18. DaSalla CS, Kambara H, Sato M, Koike Y (2009) Single-trial classification of vowel speech imagery using common spatial patterns. Neural Netw 22(9):1334–1339
19. Parra LC, Spence CD, Gerson AD, Sajda P (2003) Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring. IEEE Trans Neural Syst Rehabil Eng 11(2):173–177
20. D'Zmura M, Deng S, Lappas T, Thorpe S, Srinivasan R (2009) Toward EEG sensing of imagined speech. In: In International Conference on Human-Computer Interaction, pp 40–48
21. Li W, Zhang X, Zhong X, Zhang Y (2013) Analysis and classification of speech imagery EEG for BCI. Biomed Signal Process Control 8(6):901–908
22. Hsu Y-L, Yang Y-T, Wang J-S, Hsu C-Y (2013) Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing 104:105–114

23. He SL, Gao X, Yang F, Gao S (2003) Imagined hand movement identification based on spatio-temporal pattern recognition of EEG. In: 1st EMBS conference on neural engineering, pp 599–602
24. Deng S, Srinivasan R, Lappas T, D'Zmura M (2010) EEG classification of imagined syllable rhythm using hilbert spectrum methods. J Neural Eng 7(4). https://doi.org/10.1088/1741-2560/7/4/046006
25. Esfahani ET, Sundararajan V (2012) Classification of primitive shapes using brain–computer interfaces. Comput Aided Des 44(10):1011–1019
26. Torres-García AA, Reyes-García CA, Villaseñor-Pineda L, García-Aguilar G (2016) Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. Expert Systems with Applications 59:1–12
27. González-Castañeda EF, Torres-García AA, Reyes-García CA, Villaseñor-Pineda L (2017) Sonification and textification: Proposing methods for classifying unspoken words from EEG signals. Biomed Signal Process Control 37:82–91
28. Wang K, Wang X, Li G (2017) Simulation experiment of bci based on imagined speech EEG decoding. arXiv:1705.07771
29. Nguyen CH, Karavas G, Artemiadis P (2017) Inferring imagined speech using EEG signals: a new approach using riemannian manifold features J Neural Eng https://doi.org/10.1088/1741-2552/aa8235
30. Soleymani M, Pantic M, Pun T (2012) Multimodal emotion recognition in response to videos. IEEE Trans Affect Comput 3(2):211–223
31. Gauba H, Kumar P, Roy PP, Singh P, Dogra DP, Raman B (2017) Prediction of advertisement preference by fusing EEG response and sentiment analysis. Neural Netw 92:77–88
32. Donos C, Dümpelmann M, Schulze-Bonhage A (2015) Early seizure detection algorithm based on intracranial EEG and random forest classification. Int J Neural Syst 25(05):1550023
33. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H (2012) Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. Comput Methods Prog Biomed 108(1):10–19
34. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA (2009) A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinf 10(1):213
35. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
36. Übeyli ED (2009) Combined neural network model employing wavelet coefficients for EEG signals classification. Digital Signal Process 19(2):297–308
37. Chapelle O, Vapnik V, Bousquet O, Mukherjee S (2002) Choosing multiple parameters for support vector machines. Mach Learn 46(1–3):131–159
38. Yadava M, Kumar P, Saini R, Roy PP, Dogra DP (2017) Analysis of EEG signals and its application to neuromarketing. Multimedia Tools and Applications 76(18):19087–19111
39. Kumar P, Gauba H, Roy PP, Dogra DP (2017) A multimodal framework for sensor based sign language recognition. Neurocomputing 259:21–38

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com