Augment Hum Res (2016) 1:3
DOI 10.1007/s41133-016-0001-z

ORIGINAL PAPER

A Communication Paradigm Using Subvocalized Speech: Translating Brain Signals into Speech

Kusuma Mohanchandra[1] · **Snehanshu Saha**[2]

Received: 4 June 2016 / Accepted: 14 September 2016 / Published online: 10 October 2016

Abstract Recent science and technology studies in neuroscience, rehabilitation, and machine learning have focused attention on the EEG-based brain–computer interface (BCI) as an exciting field of research. Though the primary goal of the BCI has been to restore communication in the severely paralyzed, BCI for speech communication has acquired recognition in a variety of non-medical fields. These fields include silent speech communication, cognitive biometrics, and synthetic telepathy, to name a few. Though potentially a very sensitive issue on various counts, it is likely to revolutionize the whole system of communication. Considering the wide range of application, this paper presents innovative research on BCI for speech communication. Since imagined speech suffers from quite a few factors, we have chosen to focus on subvocalized speech for the current work. The current work is considered to be the first to utilize the subvocal verbalization for EEG-based BCI in speech communication. The electrical signals generated by the human brain during subvocalized speech are captured, analyzed, and interpreted as speech. Further, the processed EEG signals are used to drive a speech synthesizer, enabling communication and acoustical feedback for the user. We attempt to demonstrate and justify that the BCI is capable of providing good results. The basis of this effort is the presumption that, whether the speech is overt or covert, it always originates in the mind. The scalp maps provide evidence that subvocal speech prediction, from the neurological signals, is achievable. The statistical results obtained from the current study demonstrate that speech prediction is possible. EEG signals suffer from the curse of dimensionality due to the intrinsic biological and electromagnetic complexities. Therefore, in the current work, the subset selection method, using pairwise cross-correlation, is proposed to reduce the size of the data while minimizing loss of information. The prominent variances obtained from the SSM, based on principal representative features, were deployed to analyze multiclass EEG signals. A multiclass support vector machine is used for the classification of EEG signals of five subvocalized words extracted from scalp electrodes. Though the current work identifies many challenges, the promise of this technology is exhibited.

Keywords Brain–computer interface · Electroencephalography · Subvocalized speech · Speech communication · Subset selection method · Support vector machine

Introduction

The brain–computer interface (BCI) is an evolving technology that facilitates the communication between the human brain and any external device without using the normal output pathways [1]. The BCI is an interface that translates human brain signals into machine control signals to be used where no muscular movements is made. The machine can be a computer, wheelchair, robot, assistive device, or an alternative communication device. The BCI has a broad range of applications, in both the medical and non-medical domains. Using BCI for speech communication is

✉ Kusuma Mohanchandra
kusumalak@gmail.com; kusuma-cs@dayanandasagar.edu

[1] Department of Computer Science and Engineering, Medical Imaging Research Centre, Dayananda Sagar College of Engineering, Bangalore 560078, India

[2] Department of Computer Science and Engineering, The Center for Basic Initiatives in Mathematical Modeling and Computation (CBIMMC), PESIT South Campus, Bangalore 560100, India

one such application; the attempted speech is used to actuate a speech synthesizer, enabling a person to communicate with the external world through his brain signals. An electroencephalography (EEG)-based BCI for speech communication measures the brain electrical activity of an individual during attempted speech through the scalp electrodes. The brain signals picked by the electrodes are sent to the computer, processed, and converted into meaningful words that can be communicated as aural information. Though the primary goal of the speech BCI is to act as an alternative communication device for physically challenged people, it also extends its applications to non-medical fields such as silent speech communication, synthetic telepathy, and cognitive biometrics [2, 3].

A survey of leading-edge literature identifies a gap in the ability to provide speech communication using brain signals to produce meaningful words (such a provision already exists, only for syllables and phonemes). The silent speech can be produced in three ways: (1) talking by moving the speech articulators but without producing any audible sound. The signals are captured mainly by using EMG sensors placed around the neck and mouth; (2) speech imagery—imagine the word to be produced; (3) talking in the mind without moving any speech articulators and without making any audible sound (Subvocalization). Although earlier research has demonstrated that EEG-based BCI for speech communication is possible with imagined speech, the lack of lateralization exhibits a significant challenge in analyzing the neural signals of imagined speech [4]. To date, most of the studies on speech communication are based on invasive approaches. However, a few researchers have decoded only the phonemes and syllables, in a noninvasive way, using EEG signals (shown in Table 4). Therefore, in the current study an effort is made to develop a BCI designed for speech communication using the neural activity of the brain through subvocalized speech. The authors tested the subvocalized speech behavior of the subject, for a selected number of words, measured by the scalp electrodes. Subvocalization is the mental rehearsal of the word without making any audible sound and without moving any speech articulators. Subvocalization refers to the subconscious motor activity that occurs during speech without the presence of a sound wave. Neuroscience studies have shown that the subvocalization of speech plays several roles in auditory imagery. Subvocalization activates motor and auditory pathways, so during subvocal verbalization, additional brain pathways are activated. These induce significantly different activation pattern when compared to the results of imagined speech or visual imagery.

One of the main techniques for studying subvocalization is electromyography, which detects minute muscle potentials in speech organs. The procedure records diffuse muscle activities and hence show only the overall activity level, but not the exact words or sounds being subvocalized. Identifying the articulatory pattern is not possible accurately. Hence, in the current work, EEG is used to acquire the brain signals during subvocal verbalization of the words. The basis of this effort is the presumption that, whether the speech is overt or covert, it always originates in the mind. The human act of talking involves a complex set of phonatory and articulatory mechanisms. But even when the acoustic aspects of phonetics are removed, we are still "speaking" in our head. This introduces brain activation and changes in the power dynamics of the brain. So, in the current work, EEG is used to measure changes in voltage in the brain during subvocalized speech production. As a preliminary investigation, in this study the subvocalized speech behavior of three normal subjects was tested for later comparison to speech disabled subjects. The experiment was conducted for a selected number of words measured from the scalp EEG electrodes. The model used in developing a BCI for speech communication is presented in module 2. Limitations of the methods used and future enhancements are also discussed.

Methods

The architecture used in the BCI speech communication system is shown in Fig. 1. The data acquisition system captures the EEG data from the electrodes at specific locations on the scalp for input to the BCI system. The preprocessing module involves amplification, analog filtering, A/D conversion, and artifact removal, thereby improving the signal-to-noise ratio of the EEG signal. Next, in the feature selection stage the dependent discriminatory features of each subvocalized word are extracted from the preprocessed signals. These features form a feature vector, upon which the classification is done. In the classification stage, feature patterns are identified to determine the subvocalized words spoken by the user. Once the classification into such categories is done, the words are quickly recognized and the speech sounds are produced by the speech synthesizer.

To accept the electrical activity of the brain from scalp recordings, the signals must be of sufficient strength and duration from a considerable number of trials. The EEG signals are extracted from a large number of channels due to the brain's voluminous conductive output, yielding a large dataset and a significant computational challenge. Selecting relevant features from such large datasets is a fundamental challenge in EEG-based BCI. Much of the data extracted from electrodes, placed in various regions, may be extraneous, irrelevant, or even "noise" for the classification problem at hand.

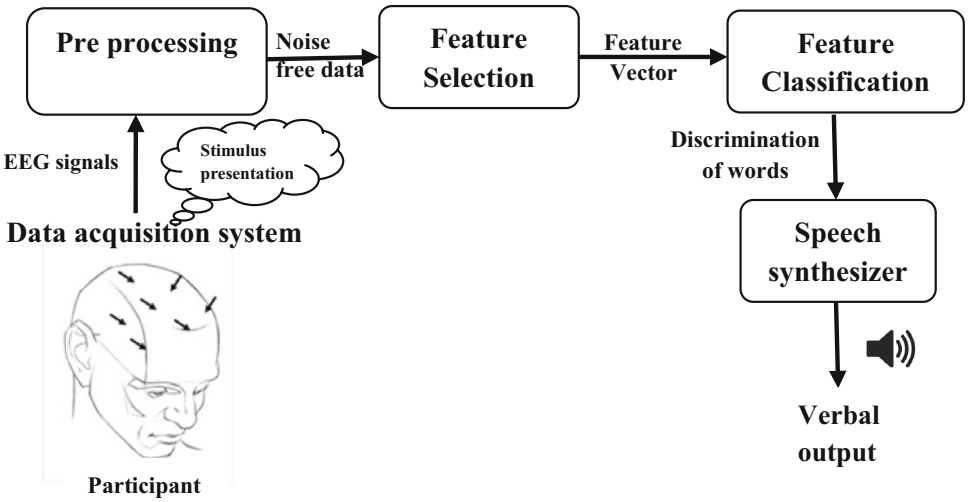

Fig. 1 Functional model of the EEG-based BCI system

Hence, the objective of this study is to investigate data reduction and classification methods to minimize the computational complexity of analyzing the EEG signals. EEG signals suffer from the curse of dimensionality due to the intrinsic biological and electromagnetic complexities. In this context, the subset selection method (SSM), based on a focused set of principal representative features (PRF), is used to select data and reduce the dimensionality. The respective variance contributions, computed for an optimal number of channels, are considered as principal features. The prominent variances obtained from the SSM, based on principal features, are selected for multiclass EEG signal analysis. At this point, a multiclass support vector machine (SVM) is used to classify the EEG signals for five subvocalized words extracted from the scalp electrodes.

Data Acquisition Paradigm

The EEG data are recorded using a Neuroscan 64-channel cap-based EEG device with the standard 10–20 montage placements. Vertical eye movements were detected using separate channels, placed above and below the subject's right eye. Horizontal eye movements were recorded by different electrodes put on either side of the eyes (temple region—temporal). In this study, meaningful words, catering to the basic needs of a person, are considered. The EEG data are extracted during subvocalization of the word, i.e., when the subject talks silently in his mind without any overt physical manifestation of speech. The data acquisition paradigm is shown in Fig. 2. The experiment involved three volunteer participants referred to as subject1 through subject3. The five words selected are "water," "help," "thanks," "food," and "stop"—referred as word1, word2, word3, word4, and word5, respectively, in subsequent modules. Subject1 had been trained in the BCI experiments of subvocalized speech; the other subjects had never participated before in BCI experiments. All volunteer subjects were right-handed male students between the ages of 20 and 25. All subjects are otherwise normal and underwent the EEG process without any neurological antecedents.

While the participant subvocalized the word in his mind, the brain electrical activity was recorded by the EEG system. The experimental paradigm was presented with E-Prime 2.0 software. In each trial, a word is presented on the computer screen at time zero. The display of word is followed by three beeps in a particular rhythm. After the third beep, the participant has to subvocalize the word in his mind five times in the same rhythm as the beeps. During this period, no audio stimulus is presented. Approximately 2 s after the last beep, the subject starts to subvocalize the word shown on the monitor at the given rhythm. The participant is instructed to avoid blinking or move any muscles and to concentrate on the word shown. Each trial has five instances of a particular word, and the duration of a single trial was 17 s followed by a short break. Then the next word would be displayed for the subject to subvocalize. The time interval for rest between each trial was randomized between 8 and 10 s, to prevent the subjects from getting used to the length of the rest period. A single experimental session was comprised of the EEG acquisition for 25 trials of each word. The data were recorded over two separate sessions with varying word order contributing to a total of 50 trials of each word (total number of trials = 50 trials × 5 words). Each trial has five instances of subvocalized word. The EEG was recorded in a controlled environment. The EEG data were recorded in a

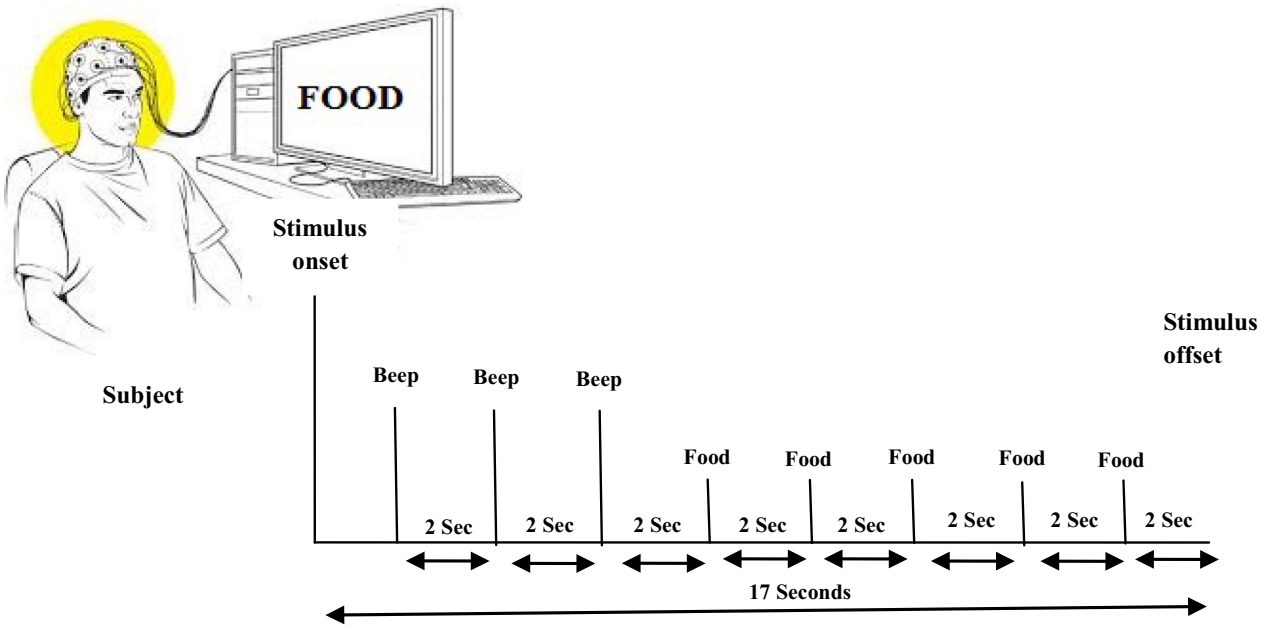

Fig. 2 Data acquisition paradigm includes the experiment design to capture brain signal behavior during subvocalized speech. The diagram shows capturing one trial of a particular word

continuous mode with Neuroscan Synamps 2 amplifiers at a sampling rate of 1000 Hz.

Preprocessing

The EEG data were analyzed off-line using Neuroscan's SCAN 4.5 software. The signals are filtered between 0.5 and 40 Hz using a band-pass filter and down-sampled to 250 Hz. The eyeblink artifact reduction was done in all of our experiments. The vertical and horizontal ocular artifacts were reduced using the independent component analysis-based blink artifact reduction algorithm implemented in SCAN 4.5. All blink activities were reduced from the continuous signal. Artifacts other than eye blinks were not removed. After the removal of artifacts, the signals are epoched and averaged.

For each trial, the signal was extracted with reference to the stimulus onset and offset markers on the continuous file. Each trial has five instances of a word (called epochs), each epoch with approximately 2-s duration. The epochs are extracted and averaged over each trial leaving only that activity that is consistently associated with the stimulus in a time-locked manner. All the spontaneous EEG signal that is random about the stimuli onset is averaged out, leaving only the event-related potentials. Finally, there are 50 epochs of averaged EEG signal of each word, forming a total of 250 epochs per subject (50 epochs/word $\times$ 5 words).

Feature Selection

Feature selection is a kind of dimensionality reduction that efficiently identifies and generates discriminatory features among different classes of data as a trampled feature vector. In the current work, EEG is measured from 64 channels with a 2-s epoch for each word, contributing to a huge amount of data. Hence, the need for dimension reduction is crucial in EEG data analysis. Due to volume conduction of the brain signals, the electric potential observed at the scalp becomes more widespread. Hence, the channels are highly correlated. Prominent signals are measured by scalp electrodes located above the active cerebral area involved in the mental processing. So, in multi-channel EEG data, groups of channels are interrelated. The reason for this multipronged data analysis is that more than one channel might be measuring the same EEG potential evoked by the mental processing. However, to avoid the redundancy of information, a group of channels can be replaced with a new single variable/channel. In our work, the representative feature SSM, using pairwise cross-correlation among the features, is used to reduce the size of the dataset with minimal loss of information. The desired outcome from the SSM, based on principal representative features (PRF), is to project the feature space onto a smaller subspace that represents the data with significant discrimination. This exercise facilitates analysis, as explained in the subsequent discussion.

Algorithm 1: Computing the most informative subspace in a space S of EEG signals for feature selection
Input: Observation matrix X containing m channels and N samples per channel.
Output: Feature vector containing n significant variances.

Function ComputeVariance(X) // Calculates the variance for each trial of data
Calculate the empirical mean of each row of data // for each channel
Subtract the mean of the data from the original dataset
Compute the covariance matrix of the dataset
Compute the variance and *PRF* of the covariance matrix
Sort the variance and the associated *PRF* in decreasing order
Return n significant characteristic variances as the feature vector
end function

The subset method generates a new set of variables, called PRF. Each PRF is a linear combination of the original variables. All the PRFs are orthogonal to each other, so there is no redundant information. This relationship is ascertained by a simple pairwise cross-correlation coefficient computation. The PRFs as a whole form an orthogonal basis for the space of the data. The coefficients are calculated so that the first PRF defines the maximum variance. The second PRF is calculated to have the second highest variance and, importantly, is uncorrelated with the first PRF. Subsequent PRFs exhibit decreasing contribution of variance and are uncorrelated with all other PRFs. The full set of PRFs is as large as the original set of variables. However, it is common that the cumulative sum of the variances of the first few PRFs exceeds 80 % of the total variance of the original data as observed in our experimental procedure. Only the first few variances can be considered; the remaining is discarded, thus reducing the dimensionality of the data. The output generated by the SSM based on principal features is described in Algorithm (1).

The algorithm is explained in detail as follows. Let $X \in R^{m\times N}$ denote the original matrix, where m and N represent the number of channels and number of samples per channel, respectively. Let $Y \in R^{m\times N}$ denote the transformed matrix derived from a linear transformation P on X. The sample mean M, of each channel, given by $M = \frac{1}{N}\sum_{i=1}^{N} X_i$, is subtracted from every measurement of each channel. For m channels, the covariance matrix C is computed, which is an $m \times m$ square symmetrical matrix. The elements of C are defined as:

$$c_{ik} = c_{ki} = \frac{1}{N-1}\sum_{t=1}^{N}(X_{it} - M_i)(X_{kt} - M_k) \tag{1}$$

where X is the dataset with N samples and M_i denotes the mean of channel i. The entry C_{ik} in C for $i \neq k$ is called the covariance of X_i and X_k. C is positive definite [5] since it is of the form XX^{T}.

The SSM based on principal features finds an orthonormal $m \times m$ matrix P that transforms X into Y such that $X = PY$.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \dots & u_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{bmatrix} \tag{2}$$

Each row in P is a set of new basis vectors for expressing the columns of X. The new variables $y_1, y_2, \ldots, y_m$ are uncorrelated and are arranged in decreasing order. Each observation of x_i is transformed to y_i, by rotation and scaling, to align a basis with the axis of maximum variance, such that $x_i = Py_i$. The x_i is rendered into m new uncorrelated variables y_i. Obtaining the principal feature axes involves computing the eigenanalysis of the covariance matrix C. The eigenvalue λ_i is found by solving the characteristic equation, $|\mathbf{C} - \lambda\mathbf{I}| = 0$. The eigenvalue denotes the amount of variability captured along that dimension. The eigenvectors are the columns of matrix P such that

$$C = \mathrm{PDP}^{\mathrm{T}}, \quad \text{where} \quad D = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_m \end{bmatrix} \tag{3}$$

The vectors $u_1, u_2, \ldots, u_m$ are the unit vectors corresponding to the columns of the orthogonal matrix P. The unit vectors $u_1, u_2, \ldots, u_m$ are called the PRF vectors. They are derived in decreasing order of importance. The first PRF u_1 determines the new variable y_1 as shown in Eq. (4). Thus, y_1 is a linear combination of the original variables x_1,

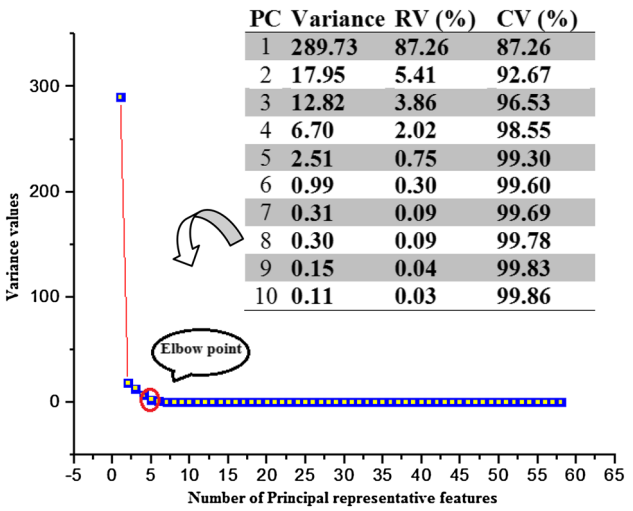

Fig. 3 Variance, relative variance (RV), and cumulative variance (CV) for word1 (first ten values), and the corresponding scree plot is shown

Table 1 Coefficients of the first four PRFs

Channels	Coefficients of PRF-1	Coefficients of PRF-2	Coefficients of PRF-3	Coefficients of PRF-4
FP1	0.0022	0.0616	0.0317	0.0801
FPZ	**0.6193**	**−0.3484**	**−0.4277**	**0.1128**
FP2	0.0016	0.0775	0.0188	0.0778
AF3	0.0018	0.0563	0.0327	0.0806
AF4	0.0016	0.0603	0.0200	0.0737
F7	0.0008	0.0324	0.0163	0.0588
F5	0.0028	0.0494	0.0314	0.0547
F3	0.0038	0.0509	0.0569	0.0717
F1	0.0026	0.0563	0.0405	0.0818
FZ	**0.5202**	**0.5275**	**0.4171**	**0.4084**
F2	**0.5587**	**−0.1185**	−0.0903	**−0.2703**
CP6	**0.1829**	0.0911	**0.4793**	**−0.7529**
PO4	0.0151	**−0.7200**	**0.6110**	**0.2819**
FC2	0.0013	0.0434	0.0254	0.0710

The channels from the frontal region and the channels with prominent variance selected from the PRF matrix are shown

$x_2, \ldots, x_m$ where $a_1, a_2, \ldots, a_m$ are the entries in PRF vector u_1. Similarly, u_2 determines the variable y_2 and so on.

$$y_1 = a_1x_1 + a_2x_2 + \cdots + a_mx_m \tag{4}$$

Hence, the SSM based on principal features generates a subset of features endowed with large representative variances, thus embodying impressive structure, while the features with lower variances represent noise and are omitted from the feature space.

Figure 3 shows the variance of the covariance matrix, computed using the SSM based on principal features. The cumulative variance (CV) illustrates that the first four variances explain 99 % of the total variance. The remaining components contribute less than 1 % each. Therefore, the first four components are chosen to form the feature vector, and the remaining variances are discarded. A scree plot helps to select the specific number of variance. The number of variances to choose depends on the "elbow" point. After the elbow point, the remaining variance values are relatively small and are all about the same size, and hence, can be discarded.

The first two PRFs are typically responsible for the bulk of the variance. They display most of the variance in the data and give the direction of the maximum spread of the data. The first PRF gives the direction of the maximum spread of the data. The second gives the direction of the maximum spread, perpendicular to the first direction. The loading plot in Fig. 4a reveals the relationships between variables/channels in the space of the first two PRFs. An intense loading for PRF-1 is observed in channels FPZ, F2, CP6, and FZ. Similarly, an intense loading for PRF-2 is found in electrode channels PO4 and FC2. A three-dimensional loading plot of PRF-1, PRF-2, and PRF-3 is shown in Fig. 4b.

In Table 1, a significant difference in the values is observed for FPZ, FZ, F2, and CP6 of PRF-1. Also, note that the majority of the variance in the dataset is along the aforementioned channels. So, the information from these channels alone is just sufficient to infer the result.

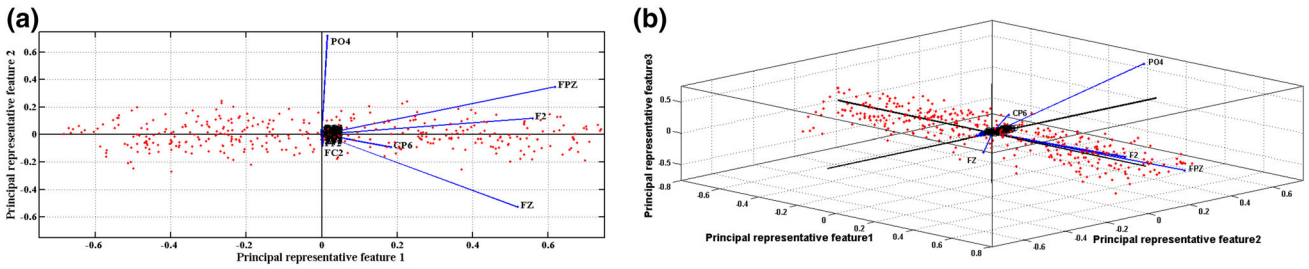

Fig. 4 **a** Two-dimensional PRFs plot and **b** three-dimensional PRFs plot reveal the relationship between variables in different subspaces

Therefore, the information from the remaining channels can be discarded, thus reducing the computational burden on the system. The PRF-1 expressed as a linear combination of the original variables is shown in Eq. (5).

$$\begin{aligned}\mathrm{PRF1} = {} & 0.0022 * \mathrm{FP1} + 0.6192 * \mathrm{FPZ} + 0.0016 * \mathrm{FP2} \\ & + 0.0018 * \mathrm{AF3} + 0.0016 * \mathrm{AF4} + 0.0008 * \mathrm{F7} \\ & + 0.0028 * \mathrm{F5} + \cdots \end{aligned} \tag{5}$$

slab parallel to the hyperplane that has no interior data points. Given the features of the data, the support vector machine is first trained to compute a model to distinguish the data from two classes. The trained model is then used to classify the new incoming data. The details are given in Algorithm (2).

Algorithm 2: Classify the EEG signals (one-against-all) based on the feature vector
Input: feature vector from algorithm 1 for 5 different classes (train, test)
Output: classified label

Function BuildMultiSVM(train, test)
training set (x_i, y_i, i=1,2,..,5)
for K Є { 1: I } **do** // number of classes
compute SVM solution w and b for all x_i with input labels y_i, such that
$\Phi(w) = \frac{1}{2}||w||^2$, is minimized
subject to the constraint $y_i(w^T x_i + b) \geq 1$
binarize the group such that the current class is +1 and all other classes are -1
compute output $f_i = \langle w, x_i \rangle + b$ for all x_i in +1 class
set $y_i = \mathrm{sgn}(f_i)$ for every K Є I, $y_i = +1$
end for
for J Є {1:size(test)} // classify test cases
for k Є {1:I} // number of Classes
if(models(k)==test(j))
break;
end if
end for
return k;
end for

end function

The current study shows significant differences in the variance across different words (given in Table 2). These variances correspond to the EEG change due to a particular component of mental rehearsal of different words. These features from the training set and the testing set were fed to the classifier, in the form of feature vectors, for classification of the test data.

Feature Classification

An SVM is a supervised learning model that classifies the data by finding the best hyperplane [6] that separates all data points of one class from those of the other class. The best hyperplane for an SVM is the one with the maximum margin between the two classes. Margin is the width of the

For a specified set of training data (x_i, y_i), where $i = 1, \ldots, N$, and $x_i \in R^d$ and $y_i \in \{+1, -1\}$ (representing two different classes of the subvocalized word), train a classifier $f(x)$ such that:

$$f(x_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases} \tag{6}$$

The linear classifier is of the form $f(x_i) = w^T x_i + b$ (dot product) where w is the normal to the hyperplane, known as weight vector, and **b** is the bias. For a linear classifier, the **w** is learned from the training data and is needed for classifying the new incoming data. The support vectors are the data points x_i on the boundary, for which $y_i f(x_i) = 1$. The optimal hyperplane can be represented as $|w^T x_i + b| = 1$. The distance between a support vector x_i and the

Table 2 Range (mean ± standard deviation) of the first four features (variances) across 50 trials of EEG signals for the five subvocalized words

Features	Words				
	Word1	Word2	Word3	Word4	Word5
Feature 1	232.07 ± 0.20	424.73 ± 0.28	143.12 ± 0.20	807.84 ± 0.09	322.01 ± 0.34
Feature 2	15.38 ± 0.23	20.00 ± 0.16	17.12 ± 0.29	28.44 ± 0.11	14.56 ± 0.29
Feature 3	8.05 ± 0.34	10.66 ± 0.24	8.89 ± 0.17	9.96 ± 0.25	9.02 ± 0.24
Feature 4	5.33 ± 0.18	5.70 ± 0.30	5.14 ± 0.17	5.72 ± 0.36	4.89 ± 0.14

A major difference in the variance across each word facilitated classifying the word appropriately

hyperplane can be written as shown in Eq. (7). For a canonical hyperplane, the numerator is equal to one. Therefore, the distance from the hyperplane to the support vectors is $\frac{1}{||W||}$

$$\text{Distance} = \frac{|w^{\mathrm{T}}x_i + b|}{||w||} = \frac{1}{||w||} \tag{7}$$

The margin M is twice the distance from the hyperplane to the support vectors. Therefore, $M = 2/||w||$. To find the best separating hyperplane, estimate **w** and **b** that maximize the margin $2/||w||$, such that for $y_i = +1$, $w^{\mathrm{T}}x_i + b \geq 1$ and for $y_i = -1$, $w^{\mathrm{T}}x_i + b \leq -1$ or equivalently, minimize $\frac{1}{2}||w||^2$ subject to the constraint $y_i(w^{\mathrm{T}}x_i + b) \geq 1$. Learning an SVM can be formulated as a convex quadratic optimization problem, subject to linear inequality constraints for a unique solution. The objective function [7] of this problem is formulated as:

$$\begin{aligned} &\min_{w \in R^d} \; J(w) = \frac{1}{2}||w||^2 \\ &\text{s.t} \quad \left\{ y_i\left(w^{\mathrm{T}}x_i + b\right) \geq 1, \quad i = 1, 2, \ldots, N \right. \end{aligned} \tag{8}$$

We can express the inequality constraint as $C_i(w) = y_i(w^{\mathrm{T}}x_i + b) - 1$. The Lagrangian function is used as the method to find the solution for constrained optimization problems with one or more equalities. However, when the function has inequality constraints, we need to extend the method to Karush–Kuhn–Tucker (KKT) conditions. The KKT defines the necessary conditions for a local minimum of constrained optimization. The necessary conditions define the properties of the gradients of the objective and constraint functions. According to the KKT dual complementarity condition—$\alpha_i C_i(x) = 0$, the objective function of Eq. (8) can be expressed by a Lagrangian function as shown in Eq. (9).

$$\begin{aligned} &\min \quad L(w, b, \alpha_i) = \frac{1}{2}||w||^2 - \sum_{i=1}^{d} \alpha_{\mathrm{i}} \left[y_i\left(w^{\mathrm{T}}x_i + b\right) - 1 \right] \\ &\text{s.t} \quad \alpha_i \geq 0, \quad i = 1, 2, \ldots, N \end{aligned} \tag{9}$$

The scalar quantity α_i is the Lagrange multiplier for the corresponding data point x_i. The optimal condition for the Lagrange function is at some point w when no first-order feasible descent direction exists (saddle point). At this point w, there exists a scalar α_i such that

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{d} \alpha_i y_i x_i \tag{10}$$

and

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{d} \alpha_i y_i = 0 \tag{11}$$

If we exploit the definition of **w** from Eq. (10) and substitute it in the Lagrangian Eq. (9), then simplify, we get

$$L(w, b, \alpha_i) = \sum_{i=1}^{d} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{d} \alpha_i \alpha_j y_i y_j x_i^{\mathrm{T}} x_j - b\sum_{i=1}^{d} \alpha_i y_i \tag{12}$$

However, from Eq. (11) the last term in Eq. (12) must be zero. Positioning the constraints $\alpha_i \geq 0$ and the constraint given in Eq. (11), we obtain the dual optimization problem shown in Eq. (13).

$$\begin{aligned} &\max \; W(\alpha) = \sum_{i=1}^{d} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{d} \alpha_i \alpha_j y_i y_j x_i^{\mathrm{T}} x_j \\ &\text{s.t} \quad \sum_{i=1}^{d} \alpha_i y_i = 0 \quad \text{and} \quad \alpha_i \geq 0 \quad \forall i \end{aligned} \tag{13}$$

The optimal value of α, substituted in Eq. (10), gives the optimal value of w in terms of α. There exists a Lagrange multiplier α_i for every training data point x_i. Suppose we have fit our model's parameters to a training set, and now wish to make a prediction at a new input point, x. We would then calculate the linear discriminate function $g(x) = w^{\mathrm{T}}x + b$ and predict $y = 1$ if and only if this quantity is greater than zero. Using Eq. (10), the discrimination function can be written as:

$$\begin{aligned} w^{\mathrm{T}}x + b &= \left(\sum_{i=1}^{d} \alpha_i y_i x_i \right)^{\mathrm{T}} x + b \\ w^{\mathrm{T}}x + b &= \sum_{i=1}^{d} \alpha_i y_i \langle x_i, x \rangle + b \end{aligned} \tag{14}$$

The prediction of the class labels from the Eq. (14) depends on the inner product between the input point x and the support vectors x_i of the training set. In the solution, the points that have $\alpha_i > 0$ are called the support vectors.

In general, if the problem does not have a simple hyperplane as a separating criterion, we need nonlinear separators. A nonlinear classifier can be created by applying the kernel trick. A kernel function maps the data points onto a higher-dimensional space, hoping to improve the separateness of data. The kernel function is expressed as a dot product in an infinite dimensional feature space. Therefore, the dot product between the input point x and the support vectors x_i in Eq. (14) can be computed by a kernel function. Using kernels, the discriminate function $g(x)$ with support vectors x_i can be written as:

$$g(x) = w^{\mathrm{T}}x + b = \sum_{i=1}^{d} \alpha_i y_i k(x_i, x) + b \tag{15}$$

Using the kernel function, the algorithm can be carried into a higher-dimensional space without explicitly mapping the input points into this space. This is highly desirable as sometimes our higher-dimensional feature space could even have infinite dimension and, thus, be infeasible to compute. With the kernel functions, it is possible to operate in a theoretical feature space of infinite dimension. Some standard kernel functions include the polynomial function, the radial basis function, and the Gaussian functions.

In the present work, a one-against-all multiclass SVM, with the default linear kernel, was constructed to discriminate the five subvocalized words competently. The feature classification, using the SVM classifier, is described in Algorithm (2). Linear kernel SVM was used as the data are found to be linearly separable. The linearity of the data was verified using the perceptron learning algorithm. The one-against-all model constructs N ($N = 5$ in the present work) binary SVM classifiers, each of which separates one class from the rest. The jth SVM is trained with the features of the jth class and labeled as a positive class; all of the others are labeled as a negative class. The N classes can be linearly separated such that the jth hyperplane puts the jth class on its positive side and the rest of the classes on its negative side. However, the drawback of this method is that when the results from the multiple classifiers are combined into the final decision, the outputs of the decision functions are directly compared, without considering the competence of the classifiers [8]. Another drawback of the SVM is that there is no definite method to select the best suitable kernel for the problem at hand.

Results and Discussion

A number of experiments were conducted to evaluate the performance of the designed BCI model in classifying the EEG signals of subvocalized words. The SSM, based on principal features, was applied to the preprocessed EEG signals of five subvocalized words. The dataset had 50 trials of each word, measured for 2 s, from a 64-channel EEG headset. A total of 250 trials, measured from five subvocalized words, were used for evaluating the possibility of recognizing the subvocalized word from the EEG signals. Due to the vast dimension of the dataset, the SSM, based on principal features, was used to project the data to reduce the dimension while preserving maximum useful information. An optimal number of coefficients, contributing to 99 % of the variance, were selected as features for each trial of the EEG signal. The classification performance is evaluated by a multiclass SVM (one-against-all) using a fivefold classification procedure. The features selected, using the SSM based on principal features, were used to build the classifier. To develop a generalized, robust classifier that performs well when new samples are input, we choose a fivefold cross-validation data re-sampling technique for training and testing the classifier. In this procedure, the data are split into five equal-sized subsamples. Four subsamples of the data are used for training the classifier, and one subsample is used for testing. This procedure is repeated five times using a randomly picked different subsample for testing in each case. Based on the results obtained, the precision, recall, F-measure, and accuracy are calculated. The average performance over fivefold is taken as the actual estimate of the classifier's performance.

The classifier performance is determined by computing the precision, recall, F-measure, and classification accuracy drawn from the confusion matrix. The confusion matrix illustrates the true positive (TP), false negative (FN), false positive (FP), and true negative (TN) of the classified data. The metrics are calculated using the following formulae:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{16}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{17}$$

$$F\text{-measure} = 2((\text{precision} * \text{recall}) / (\text{precision} + \text{recall})) \tag{18}$$

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{19}$$

Table 3 shows the precision, recall, F-measure, and accuracy of the model in classifying the data. The recall represents the ability of the test to retrieve the correct information. In the present work, a recall of 0.6 was achieved which means 60 % of the activity was detected (TP), but 40 % of the activity was undetected (FN). Precision identifies the percentage of the selected information that is correct. A precision of 0.5 detected 50 % of the activity correctly (TP), but the remaining 50 % of the activity was mistaken as belonging to the same class (FP). Higher recall indicates that most of the relevant information was extracted, and higher precision means that

Table 3 Precision, recall, F-measure, and accuracy assessment for different words with three subjects

Task	Recall	Precision	F-measure	Classification accuracy
Word1	0.60	0.50	0.55	0.80
Word2	0.40	0.40	0.40	0.76
Word3	0.20	0.50	0.28	0.80
Word4	0.60	1.00	0.75	0.92
Word5	0.40	0.20	0.27	0.60

substantially more relevant than irrelevant information was retrieved. The precision and recall are inversely related. Often it is possible to increase one at the cost of reducing the other. The feature selection and classifier model used in data analysis affect the level of recall and precision. The balanced F-measure is a combined measure that assesses the precision-recall trade-off. It is the average of the two parameters and varies between a best value of 1 and a worst value of 0. In the current work, the F-measure ranged between 0.27 and 0.75. The classification accuracy varied between 60 and 92 %, which appreciably is good compared to the results given in Table 4. The results indicate that there is a significant potential for the use of subvocalized speech in EEG-based direct speech communication.

The scalp maps in Fig. 5a show the brain electrical activity during subvocal verbalization of the words. Note that the neural activations are significantly prominent in the frontal lobes during subvocalized speech. Since these regions are directly responsible for speech production, the results appear promising. Figure 5b shows the plot of the scalp maps after using the SSM. The discrete sources of EEG signals are decomposed to distinguish the co-occurrence of brain electrical activity in the spatial domain, and then the signal components are mapped to a lower-dimensional space, retaining the most discriminate features. The discrimination of the brain signals corresponding to five subvocalized words is shown in Fig. 5. The plot of classification accuracy against the increasing number of features used in the feature vector to discriminate the subvocalized speech of word1 is displayed in Fig. 6. It is observed that the classification accuracy increases as the number of features increases and remains constant after the fourth value. So in the current work, only four discriminating features are used to form a feature vector. The

Table 4 Comparison of the results obtained for speech communication using ECoG and EEG signals by different researchers

Neuroimaging method	Modality	Description	Authors	Recognition rate
Invasive (implanted electrodes)	Intracortical microelectrode	Speech BCI using ECoG. Intended vowel productions by the user. The decoded signals from the attempted speech are used to drive an artificial speech synthesizer	Brumberg et al. [18–20] and Guenther et al. [21, 22]	Maximum rate of 80–90 % accuracy with a mean accuracy of 70 %
	ECoG signals	Control a one-dimensional computer cursor with ECoG features of different overt and imagined phoneme articulations	Leuthardt et al. [23–25]	74–100 % in a one-dimensional binary task and accuracies between 68 % and 91 % from the higher gamma frequency
	ECoG signals	To control a visual keyboard through BCI. Predict the intended target letters	Krusienski and Shih [26]	Greater than 70 % accuracy with 12 bits per minute bit rate
	ECoG signals	Decode elements of speech production using ECoG	Mugler et al. [28]	Classified phonemes with up to 36 % accuracy when classifying all phonemes and up to 63 % accuracy for a single phoneme
Noninvasive (EEG)	Speech imagery	Recognition of words silently spoken for auditory and visual comprehension	Suppes et al. [13]	Recognition rates varied between 34 and 97 %
Direct methods				
	Speech imagery	Imaginary speech of the English vowels /a/ and /u/, and a no-action state as control	Da Salla et al. [14]	Classification accuracies ranged from 68 to 78 %
	Speech imagery	Subjects imagining two syllables, /ba/ and /ku/, without speaking or performing any overt actions	D'Zmura et al. [15] and Brigham et al. [16]	87 % in the beta band
	Speech imagery	Recognizing unspoken/imagined speech of five words	Porbadnigk et al. [27]	Average recognition rate of 45.50 %

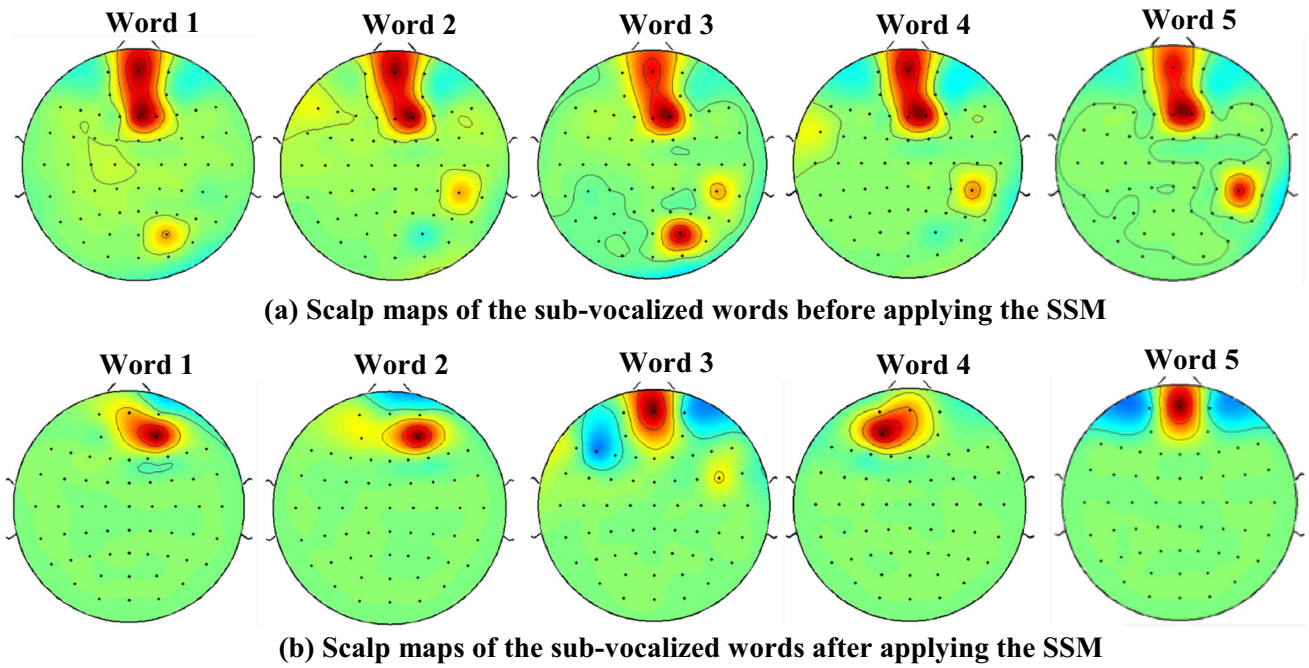

Fig. 5 Scalp topographical presentation of the brain activity, recorded from 64-electrode EEG, during subvocal verbalization of the words by subject-1. **a** The brain activity during subvocal verbalization of five different words. **b** The signal components mapped to a lower-dimensional space using the SSM

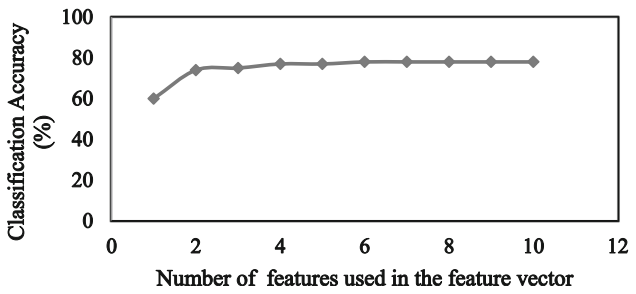

Fig. 6 *Graph* shows the variation in classification accuracy versus number of features used in the feature vector to discriminate the subvocalized speech of word1

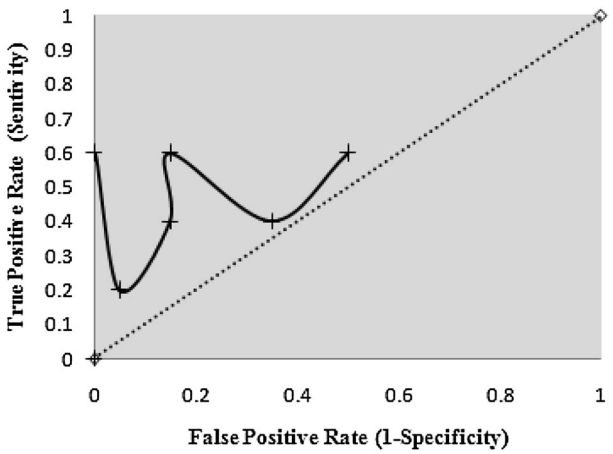

Fig. 7 ROC curve drawn for the classification of the subvocalized words using the proposed SSM algorithm

receiver operating characteristic curve (ROC) is drawn to show the effective discrimination of the proposed SSM algorithm through the multiclass SVM classifier (Fig. 7). The ROC curve serves as a measure of performance of the algorithm by plotting the true positive rate verses the false positive rate in a unit square. An ROC curve of Fig. 7 reflects that the performance of the proposed BCI model is better than chance level.

Related Work

Research on synthetic telepathy is being carried out by the US Army, with the intention to allow its soldiers to communicate just by thinking [9]. The aim is to build a thought-helmet, a device that can read and broadcast the unspoken speech of soldiers. The goal is to enable soldiers to communicate silently. Silent speech communication is one of the most exciting future technologies. Silent speech communication [10] allows people to communicate with each other by using a whispering sound or even soundless, speech. This technology is used by NASA astronauts who need to communicate despite surrounding noise. Currently, electromyography (EMG) signals captured by small, button-sized sensors affixed below the jawbone on either side of the throat [11] are used to collect the signals. A new

class of biometrics, based on the cognitive aspects of human behavior, called cognitive biometrics, presents a novel approach to user authentication. The brain state of individuals, used for the authentication mechanism, increases the robustness and enables cross-validation when used in combination with traditional biometric methods. The cognitive biometric cannot be hacked, stolen, or transferred from one person to another; they are unique for each individual. The BCI for speech communication is used as an alternative augmentative communication (AAC) device for severely disabled people who can communicate only through computer interfaces that can interpret neurological signals. For example, people suffering from amyotrophic lateral sclerosis (ALS) and locked-in syndrome (LIS) are the targeted beneficiaries.

In the past decade, several BCI techniques have been developed to restore communication in patients with varied and severe paralysis. The indirect communication devices generally used in these types of communication, such as the speller device or a virtual keyboard, suffer from slow selection rate of just one word per minute [12]. This sometimes limits the user's fluency and comprehension. Moreover, these indirect methods fail to improve the patients' behavioral abnormalities. Besides, these methods do not improve the subjects' psychological condition and constrain related speech communication. To address the above-mentioned problems and make BCI speech production more natural and fluent, direct methods are being developed. The direct approach involves capturing the neural activity of the intended speech through an EEG. The signals are then processed to predict the speech and synthesize speech production in real-time. Suppes et al. [13] used electrical and magnetic brain waves for recognition of words and sentences that were supposed to have been silently spoken. DaSalla et al. [14] developed a BCI using EEG for "imaginary speech" of the English vowels /a/ and /u/, and a no-action state as a control. The potential use of EEG as a means of silent communication has been explored by D'Zmura et al. [15] and Brigham et al. [16]. Subjects imagining two syllables, /ba/ and /ku/, without speaking or performing any overt actions, are assessed for feasibility and considered for subject identification [4, 17]. In Table 4, the authors are showing the evidence that though considerable amount of research is being carried out on silent speech, most of the work is using invasive method, which has lot of shortcomings. And the noninvasive teams are able to decode only the phonemes and syllables. Nobody has reported about decoding a complete meaningful word using EEG signals during subvocalized speech production, so we claim that the present work is novel.

Though speech communication has extensive scope in various domains of application, the challenges in processing the EEG signals in real-time are significant. At the very outset, it must be acknowledged that the EEG signals are extremely complex and prone to internal and external interference.

Conclusion

The motivation for this study was to build a practical BCI framework for speech communication using current technology. The priority is to enable communication with a simple BCI setup providing high performance and speed. The study was conducted with an eye on the vast number of applications for BCI speech communication. Potential applications include synthetic telepathy, speech communication in LIS patients, silent communication, and cognitive biometrics. EEG was chosen for this experiment since it is low cost, portable, and has high temporal resolution compared to other brain-imaging modalities. Also, EEG can detect covert processing in the brain, even without the external stimulus; our input was mainly from covert activities.

An essential contribution of the present work is the usage of subvocalized speech for the development of an EEG-based BCI for speech communication. Subvocal verbalization is associated with activation in the frontal and temporal cortex, with bi-hemispheric lateralization. This activation alleges the frontal and temporal lobes to be involved in the articulation of speech output.

The EEG signals were acquired from three healthy subjects, while they subvocally articulated one of the five words. The EEG patterns for those five essential words were then selected from each subject. The data acquired were for five complete words that were felt to relate to patients' needs as opposed to phonemes and syllables. The signals were processed using an SSM algorithm devised to reduce the magnitude of sensor data processed for feature selection.

A multiclass SVM classifier (one-against-all) was used to classify the data. The developed model was verified/evaluated using standard metrics. Several performance measures were used to investigate the feasibility and limitations of the developed BCI model. In the present system, a satisfactory accuracy in the range 80–92 % was achieved. The results show that the presented model of BCI for speech communication, using subvocalized speech, is viable, but needs improvement in classification accuracy.

The significant challenge in analyzing the EEG signals is the low signal-to-noise ratio; they are prone to internal and external noise. A more refined data analysis and comparatively large number of data are required to extract useful information from EEG. In addition, as the number of words to be classified increases, we need to build

intelligent algorithms that learn the most discriminatory features. In real-time application, classification of a vast number of words needs to be developed to make the system scalable. Furthermore, an accurate mechanism to capture the subject's focus is required to impose discrimination among the different words. Advances in sensor technology, acquisition protocols, and intelligent algorithms will be needed by BCIs to meet the desired performance. The data were acquired in a restricted atmosphere, but promises also to work well in outside situations.

Future Work

As a future work, an improved method of feature selection and classification, using advanced machine learning techniques, could be explored. Furthermore, classifier ensembles could be used to capture the significant variability in EEG data and augment the accuracy and receptiveness of the system. The next step would be to work on an expanded list of recognized subvocalized words.

The work reported so far is just an embryo in the development of a BCI system for speech communication. Future steps would be to build a commercial headset with a minimum number of electrodes to enable speech communication through subvocal verbalization. We are confident that the technologies and methodologies presented in this study provide a foundation for future development that will enable the speechless to generate assisted speech in a geometrically augmenting mode.

References

1. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113(6):767–791
2. Mohanchandra K, Saha S (2014) Optimal channel selection for robust EEG single-trial analysis. AASRI Procedia 9:64–71
3. Mohanchandra K, Saha S, Lingaraju GM (2015) EEG based brain computer interface for speech communication: principles and applications. In: Hassanien AE, Azar AT (ed) Intelligent systems reference library, brain–computer interfaces: current trends and applications, vol 74. Springer, Berlin. doi:10.1007/978-3-319-10978-7
4. Brigham K, Kumar BV (2010b) Subject identification from electroencephalogram (EEG) signals during imagined speech. In: The fourth international IEEE conference in biometrics: theory applications and systems (BTAS), 27–29 September, Washington, pp 1–8
5. Johnson CR (1970) Positive definite matrices. Am Math Mon 77(3):259–264. doi:10.2307/2317709
6. Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2(2):121–167
7. Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13(2): 415–425
8. Liu Y, Zheng YF (2005) One-against-all multi-class SVM classification using reliability measures. In: Proceedings 2005 IEEE international joint conference on neural networks, 2005. IJCNN'05, vol 2. IEEE, pp 849–854
9. Discover magazine: the army's bold plan to turn soldiers into telepaths. http://discovermagazine.com/2011/apr/15-armys-bold-plan-turn-soldiers-into-telepaths#.UZe6-9isOSo. Accessed 22 May 2015
10. Denby B, Schultz T, Honda K, Hueber T, Gilbert JM, Brumberg JS (2010) Silent speech interfaces. Speech Commun 52(4): 270–287
11. NASA. NASA develops system to computerize silent 'subvocal speech' (March 17 2004). http://www.nasa.gov/home/hqnews/2004/mar/HQ_04093_subvocal_speech.html. Accessed 22 May 2015
12. Brumberg JS, Guenther FH (2010) Development of speech prostheses: current status and recent advances. Expert Rev Med Devices 7(5):667–679
13. Suppes P, Lu ZL, Han B (1997) Brain wave recognition of words. Proc Natl Acad Sci USA 94(26):14965–14969
14. 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
15. D'Zmura M, Deng S, Lappas T, Thorpe S, Srinivasan R (2009) Toward EEG sensing of imagined speech. In: Jacko JA (ed) Human–computer interaction new trends, Part I, HCII 2009, LNCS 5610. Springer, Berlin, pp 40–48
16. Brigham K, Kumar BV (2010a) Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In: The 4th international IEEE conference on bioinformatics and biomedical engineering (iCBBE), 18–20 June, 2010, Chengdu, China, pp 1–4
17. Mohanchandra K, Lingaraju GM, Kambli P, Krishnamurthy V (2013) Using brain waves as new biometric feature for authenticating a computer user in real-time. Int J Biom Bioinform 7(1):49
18. Brumberg JS, Kennedy PR, Guenther FH (2009) Artificial speech synthesizer control by brain–computer interface. In: Proceedings of the 10th annual conference of the international speech communication association (INTERSPEECH 2009).International Speech Communication Association, Brighton, 6–10 September 2009, pp 636–639
19. Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH (2010) Brain–computer interfaces for speech communication. Speech Commun 52(4):367–379
20. Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR (2011) Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. Front Neurosci 5:65
21. Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M et al (2009) A wireless brain–machine interface for real-time speech synthesis. PLoS ONE 4(12):e8218
22. Guenther FH, Brumberg JS (2011) Brain–machine interfaces for real-time speech synthesis. In: The 2011 annual international conference of the IEEE on engineering in medicine and biology society, EMBC, 30 Aug–03 Sept 2011, Boston, MA, USA, pp 5360–5363
23. Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW (2004) A brain–computer interface using electrocorticographic signals in humans. J Neural Eng 1(2):63
24. Leuthardt EC, Miller KJ, Schalk G, Rao RP, Ojemann JG (2006) Electrocorticography-based brain computer interface-the Seattle experience. IEEE Trans Neural Syst Rehabil Eng 14(2):194–198

25. Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z et al (2011) Using the electrocorticographic speech network to control a brain–computer interface in humans. J Neural Eng 8(3):036004
26. Krusienski DJ, Shih JJ (2011) Control of a visual keyboard using an electrocorticographic brain–computer interface. Neurorehabil Neural Repair 25(4):323–331
27. Porbadnigk A, Wester M, Calliess J-P, Schultz T (2009) EEG-based speech recognition impact of temporal effects. Biosignals 2009, Porto, Portugal, Jan 2009, pp 376–381
28. Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, Shih JJ, Krusienski DJ, Slutzky MW (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng 11(3):035015