



Identification of vowels in consonant–vowel–consonant words from speech imagery based EEG signals

Sandhya Chengaiyan¹ · Anandha Sree Retnapandian¹ · Kavitha Anandan¹

Received: 3 January 2019 / Revised: 18 September 2019 / Accepted: 30 September 2019 / Published online: 4 October 2019
© Springer Nature B.V. 2019

Abstract

Retrieval of unintelligible speech is a basic need for speech impaired and is under research for several decades. But retrieval of random words from thoughts needs a substantial and consistent approach. This work focuses on the preliminary steps of retrieving vowels from Electroencephalography (EEG) signals acquired while speaking and imagining of speaking a consonant–vowel–consonant (CVC) word. The process, referred to as Speech imagery is imagining of speaking to oneself silently in the mind. Speech imagery is a form of mental imagery. Brain connectivity estimators such as EEG coherence, Partial Directed Coherence, Directed Transfer Function and Transfer Entropy have been used to estimate the concurrency and causal dependence (direction and strength) between different brain regions. From brain connectivity results it has been observed that the left frontal and left temporal electrodes were activated for speech and speech imagery processes. These brain connectivity estimators have been used for training Recurrent Neural Networks (RNN) and Deep Belief Networks (DBN) for identifying the vowel from the subject's thought. Though the accuracy level was found to be varying for each vowel while speaking and imagining of speaking the CVC word, the overall classification accuracy was found to be 72% while using RNN whereas a classification accuracy of 80% was observed while using DBN. DBN was found to outperform RNN in both the speech and speech imagery processes. Thus, the combination of brain connectivity estimators and deep learning techniques appear to be effective in identifying the vowel from EEG signals of subjects' thought.

Keywords Speech imagery · Electroencephalography (EEG) · Brain connectivity estimators · Recurrent Neural Networks (RNN) · Deep Belief Networks (DBN)

Introduction

Mental imagery is the ability of humans to remember, recollect, traverse, and make judgements. Also, mental imagery is effective in analyzing and treating many mental health disorders. It is a collective term referring to illustrations and the associated understanding of sensory

information without a direct external stimulus. Mental imagery representations are pure recall of memory induced due to original external stimuli. It is of various categories based on the senses involved and this work discusses the effect of speech imagery in terms of electrophysiological outcomes (Pearson et al. 2015). Speech imagery is a form of mental imagery which refers to imagining of speaking without any actual articulation. Speech imagery can be considered as a medium of thought, which is often referred to as “imagining of speaking”. Speech imagery has a high similarity to real voice communication (Perrone-Bertolotti et al. 2014).

Patients with neuronally affected speech disorders such as Autism find it difficult to communicate their thought. These impairments occur mainly due to the dysfunctions in the brain and severity depends on which area in the nervous system is affected. In a recent research conducted for children with speech and language impairments, it was

✉ Sandhya Chengaiyan
sandhyaaswathama@gmail.com

Anandha Sree Retnapandian
anandha.sree@gmail.com

Kavitha Anandan
KavithaA@ssn.edu.in

¹ Department of Biomedical Engineering, Centre for Healthcare Technologies, SSN College of Engineering, Chennai, Tamilnadu, India

observed that there were abnormalities in the left hemisphere of the brain in the regions related to speech (Mehta et al. 2015). Therefore, Speech imagery, a process involving imagined speech, can be used to decode the thoughts for such patients using non-invasive techniques. Electrophysiological methods such as Electroencephalography (EEG) (Suppes et al. 1999), Magnetoencephalography (MEG) (Ahissar et al. 2001) and Neuro-imaging techniques such as fMRI (Bookheimer 2002) are the commonly used techniques to study about the brain activations during speech production and speech comprehension.

Electroencephalography (EEG) is a highly effective and easily affordable process in analyzing the functional interactions between different regions of the brain. EEG measures the mental activity of the brain directly from the scalp while performing a given cognitive task. The neuronal oscillations related to speech perception and speech comprehension have been identified as theta and gamma frequency bands (Ghitza and Greenberg 2009; Ghitza 2013; Lin et al. 2015). Brain regions associated with speech and imagined speech are identified as homunculus, Broca's area and Wernicke's area (Hickok and Poeppel 2000; Poeppel and Hickok 2007). The neural correlations between these brain regions can be analyzed from Electroencephalography (EEG) signals using functional and effective brain connectivity parameters. EEG coherence is a functional connectivity parameter which is used to detect functional synchronization between two distinct brain regions for any given task (Weiss and Muller 2003). Righi et al. (2014) have analyzed neural connectivity in infants with ASD between 6 and 12 months of age. EEG signals were acquired from infants while listening to speech sounds consisting of a few consonant–vowel pairs. Coherence parameter was estimated for 6 months and 12 months old infants. It was observed from the results that the 12-months-of-age infants diagnosed with ASD showed a reduced functional connectivity in terms of EEG Coherence.

Multivariate autoregressive (MVAR) models represent the connections between multichannel EEG signals in the form of linear difference equations to predict the future values of one time signal based on the past values of another time series signal. By using multivariate AR representation of EEG signal, the direct or indirect causal influences can be analyzed (Shibata et al. 2004). Directed Transfer Function (DTF) (Kaminski and Blinowska 1991), Partial Directed Coherence (PDC) (Baccala and Sameshima 2001) are Granger causality parameters based on MVAR criteria, estimate the strength and directional influence in multivariate signals. In multivariate time series analysis, Transfer Entropy is an efficient information theoretic measure that quantifies the amount of information

transfer from one variable to the other and the connectivity strength between two time series (Schreiber 2000).

The exploration of recent research works reveals that, speech imagery brain signals could be helpful in interpreting the thoughts of an individual. Pei et al. (2011) have proven that it is possible to decode vowels and consonants from Electrocorticographic (ECoG) signals. Further, Wester has shown that it is feasible to recognize normal speech, imagined speech, whispering and silent speech from the recorded EEG signals. Results demonstrated that Broca's area and Wernicke's area work together to produce speech (Wester 2006). Classification of English vowels using imagined speech has been reported by DaSalla, where EEG signals have been acquired from subjects while imagining of speaking vowels/a/and/u/using Support Vector Machines (SVM) classifier (DaSalla et al. 2009).

DaSalla's work has been further enhanced by Idrees and Farooq in which statistical features were derived for different combinations of vowels and were classified using a linear classifier (Idrees and Farooq 2016). Rojas et al. has proved the identification of the vowels/a/and/e/in Spanish language using imagined speech (Rojas and Ramosm 2016). Min et al. has classified all the vowels/a/,e/,i/,o/,u/while imagining in a single trial of EEG recording. Statistical features computed, were fed to train the Extreme Learning Machine (ELM) for the classification of vowels (Min et al. 2016). Yoshimura et al. have proved it is possible to decode vowels from imagined articulation of Japanese vowels/a/and/i/using EEG cortical currents. Sparse logistic regression (SLR) method was used for classifying the three tasks (imagining the vowel/a/and/i/and no-imaginary task). The classification accuracy was found to be higher due EEG cortical currents (Yoshimura et al. 2016). Anandha Sree et al. have identified vowels from signals derived while imagining of speaking vowels, using wavelet decomposition of EEG signals and Deep Belief Networks (Sree and Kavitha 2017). Recent research on the analysis of speech imagery for similar and dissimilar sounding words and also on Consonant–Vowel syllable pairs has been carried out using brain connectivity estimators (Sandhya and Kavitha 2015; Sandhya et al. 2016).

Thus from the literature it has been reported that vowels can be classified from EEG signals of imagined speech using various machine learning techniques. However, to the best of authors' knowledge, the identification of vowels from imagined speech of CVC words using brain connectivity estimators and deep learning techniques have not been attempted so far. This work has been carried out on normal subjects to prove the possibility of interpreting a vowel from a mispronounced CVC word (Eg: an individual might have thought a CVC word 'CAT' but would have spoken it as 'COT').

In this work, EEG signals have been recorded from healthy subjects while speaking and imagining of speaking consonant–vowel–consonant (CVC) words for each vowel (/a/, /e/, /i/, /o/, /u/). The frontal and temporal lobe electrodes were analyzed exclusively, as these lobes are related to Broca's and Wernicke's area (Price et al. 2011). The brain connectivity parameters were analyzed exclusively for frontal and temporal electrodes in order to understand the neural correlations between the two interconnected speech processes, viz. speech and speech imagery. The vowels are classified using Recurrent Neural Network (RNN) and Deep Belief Networks (DBN) for two different protocols (Speech-S and Speech Imagery-SI) in order to identify a vowel from the subjects' thought.

Methodology

A total of six subjects ($N = 6$), three men and three women, with mean age group of 20 participated in this work. All subjects were healthy and native Indians with normal hearing. All subjects were chosen with no history of any neurological disorders. All subjects were right handed subjects and well familiar with English language. The experiment was conducted as per the guidelines of Institutional Ethics Committee of Department of Biomedical Engineering, SSN College of Engineering, India and validated by Speech pathologist from nearby hospital.

EEG acquisition

The electrical signals of the brain were recorded by placing Ag/AgCl electrodes on the scalp using the 10–20 electrode system. The 10–20 system describes the relationship between the location of an electrode and the corresponding region in the cerebral cortex of the brain (Klem et al. 1999). In most of the clinical applications, 19 recording electrodes are used. Figure 1 shows the international 10–20 electrode placement system (Courtesy Source: Rojas et al. 2018). Each electrode placement location has a letter to identify the lobes of the brain. The letters F denotes Frontal lobe (F3, F4, F7, F8), T—Temporal lobe (T3, T4, T5, T6), P—Parietal lobe (P3, P4) and O—Occipital lobe (O1, O2) and C (C3, C4) refers to Central region. Even numbered electrodes (2, 4, 6, 8) corresponds to placement of electrodes on the right hemisphere of the brain, whereas the odd numbers (1, 3, 5, 7) refers to those on the left hemisphere of the brain. The letter A (A1 and A2) are used for referencing of all EEG electrodes. Each electrode is connected to a differential amplifier. These amplifiers are used to amplify the voltage between the active electrode and the reference. The amplified signal is digitized using an

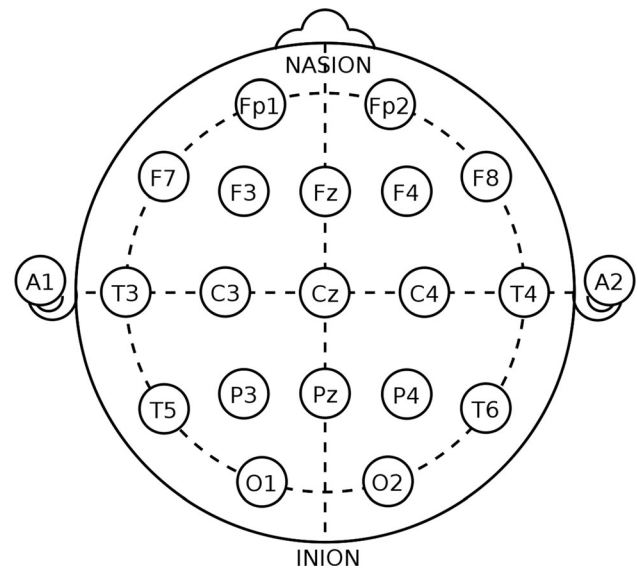


Fig. 1 Standard 10–20 electrode placement system

analog-to-digital converter. Analog-to-digital sampling generally occurs at 256–512 Hz in clinical scalp EEG.

In this work, impedance was kept below 10 k Ω and the recording was taken continuously. The software used to acquire the data was RMS EEG-32 Super Spec. The recorded signals were amplified in head box and connected to the adaptor box. The Adaptor box consist the circuitry for signal conditioning and further connected to the computer via USB port. The RMS acquisition software was used to select the required montages. All montages were unipolar, with respect to the reference electrode. Signals from Fronto-polar (Fp1, Fp2), Frontal (F3, F7, F4, F8), Temporal (T3, T4, T5, T6), Parietal (P3, P4) and Occipital lobes (O1, O2) were acquired during speech and speech imagery processes with a sampling rate of 256 Hz. Eye Blink/Artifacts were manually discarded by selection.

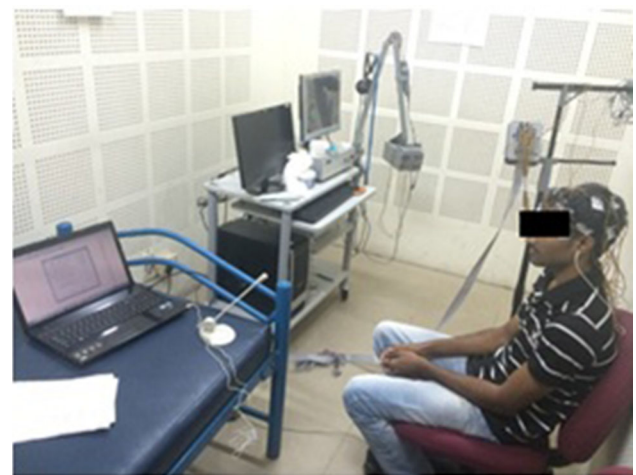


Fig. 2 EEG data acquisition of a representative subject

EMG filter was switched on during recording procedure. Figure 2 shows the EEG data acquisition of a representative subject.

Experimental protocol was designed for vowels: *a, e, i, o, u* with each vowel containing 10 Consonant–vowel–consonant (CVC) words (Enlisted in Table 1). These CVC words were given as visual stimuli to the subjects for 5 s. Protocol for the signal acquisition is shown in Fig. 3. EEG signal for each CVC word were acquired in single trial i.e. each subject were asked to perform the task (speaking and imagining of speaking a CVC word) only once. An initial and final rest of 30 s were provided to the subjects in order to make them feel comfortable during the recording process of each CVC word. After the initial rest period the subjects were instructed to visualize the CVC word for 5 s. The subjects were asked to speak the word and imagine of speaking the word silently in the mind with a period of 5 s. The subjects were asked to blink before and after performing each task of speaking and imagining of speaking the words. This resulted in easier segmentation of the task based signals for further processing.

In order to have a flawless acquisition of the EEG signal, the entire protocol and time span of each task (speaking-5 s and imagining of speaking a CVC word-5 s) were explained to the subjects before the recording session. After that, a trial recording of twice or thrice for each CVC word was performed in order to ensure that the subject feels comfortable with the experimental protocol and registers the time span of each task in the mind before the actual recording. This training session ensured that the subjects performed the tasks (speaking and imagining of speaking) with duration of 5 s exactly during the actual EEG acquisition. Though that the time span cannot be fixed for the tasks, the protocol is designed to involve such time spans to complete the task within the mentioned 5 s and take a rest of 10 s to switch to the next process. During the EEG acquisition, subjects were comfortably seated in a

chair with fully air-conditioned isolated room. If the acquired signal was found to have artifacts, then after an ample amount of rest period, the procedure was performed again.

Though imagining of speaking something takes lesser time than speaking the same itself, a fixed time frame of 5 s has been allocated for every task, to allow the brain to settle for the particular task. Also, since eyeblink has been mentioned as the start and end signs of the task, the actual time of the task have been segmented between the processes though they differed in time. The rest periods were maintained consistently to allow the brain to switch between processes so as to make the data belong to the mentioned task. As eye blink has been used as a discriminating factor, the task based signals were segmented accordingly for further processing.

EEG pre-processing

The acquired EEG signals were pre-processed and segmented for further processing. An IIR high pass filter with a frequency of 0.1 Hz and filter order of 5 has been implemented to remove the slow drift and low frequency noise (Widmann et al. 2015). 50 Hz powerline noise was removed using notch filter respectively. This work focuses on the brain connectivity analysis between different combinations of electrodes in each band. Further processing has been implemented using a IIR bandpass filter of order 5 with a passband frequency of 0.5 Hz and stopband frequency of 50 Hz. The filtered signals were further segmented into various EEG sub-bands using butterworth band-pass filter with their corresponding passband and stopband frequencies. The prominent sub-bands are Delta (0.5–3 Hz), Theta (3–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz) and Gamma bands (> 30 Hz) (Teplan 2002). Figure 4 shows the flow of the work carried out in this process.

Feature extraction

Functional connectivity estimators

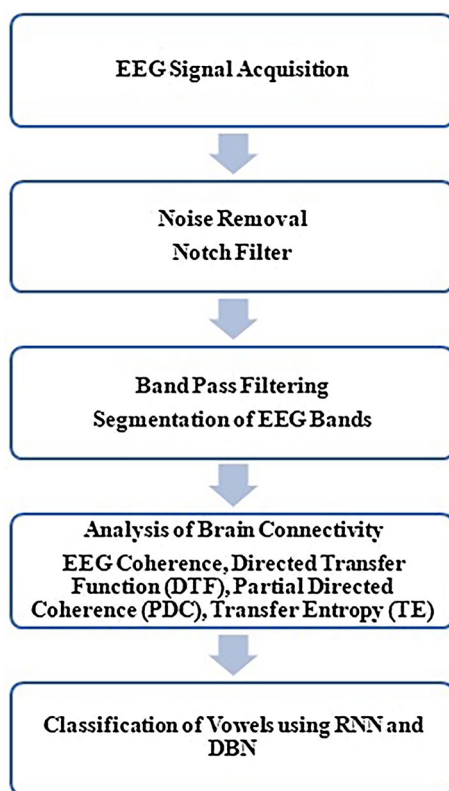
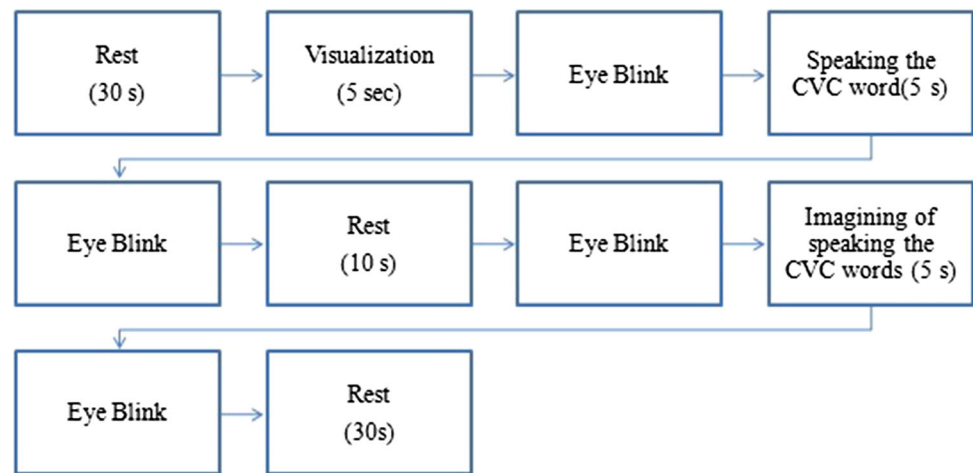
EEG coherence Coherence is a mathematical parameter that quantitatively measures the linear dependency between two distant brain regions as expressed by their EEG activity (Thatcher et al. 2004). It is represented by the mathematical formula:

$$Coh(x) = \frac{|S_{AB}|^2}{S_{AA}(x)S_{BB}(x)} \quad (1)$$

where $S_{AB(x)}$ is the cross power spectral density of the convolution of two EEG signals where as the $S_{AA(x)}$ and

Table 1 CVC words for each vowel

Vowels	A	E	I	O	U
Words	Can	Bed	Did	Box	Bun
	Car	Den	Fit	Cop	Bus
	Cat	Hen	Kit	Dog	Cup
	Bad	Her	Lip	Fog	Gum
	Dad	Jet	Pig	Jog	Hug
	Gas	Led	Pin	Lot	Hut
	Lab	Let	Rip	Not	Jug
	Man	Net	Sim	Pot	Pup
	Rat	Red	Sit	Rod	Sum
	Tap	Vex	Zip	Sob	Sun

Fig. 3 Protocol used for EEG acquisition**Fig. 4** Workflow

$S_{BB(x)}$ are the auto power spectral densities of the individual EEG Signals. EEG coherence values vary from 0 to 1.

Effective brain connectivity estimators

Granger causality measures based on multivariate autoregressive model (MVAR) For two simultaneously measured signals $x(t)$ and $y(t)$, if time series signal $x(t)$ causes time series signal $y(t)$, the continuous wave patterns in $x(t)$ are

approximately repeated in $y(t)$ after some time lag. Thus, past values of $x(t)$ can be used for the prediction of the future values of $y(t)$ (Granger 1969). Autoregressive Model (AR) of time series assumes that x_t , the (sample of data) of the process at a time t depends on its p previous values weighted by coefficients a plus a random white noise component ε . The a_j parameters are the model coefficients and p is the model order. Equation (2) describes linear regression of the time series on its own previous values, which explains the name autoregressive.

$$x_t = \sum_{j=1}^p a_j x_{t-j} + \varepsilon_t \quad (2)$$

Multivariate Autoregressive models (MVAR) extend the univariate AR model to multiple time series so that the vector of current values of all variables is modelled as a linear sum of previous activities. Consider X time series generated from X variables within a system such as a functional network in the brain and where p is the order of the model. MVAR model predicts the next value in a X -dimensional time series, X_t as a linear combination of the p previous vector values. In a multivariate (k channels) case $X(t)$, the process value at a time t is a vector of size k , each $A(j)$ is a k -by- k matrix of model coefficients (weights), and the noise component E is a vector of size k .

$$X(t) = (X_1(t), X_2(t), \dots, X_k(t))^T$$

$$X(t) = \sum_{j=1}^p A(j)X(t-j) + E(t) \quad (3)$$

Equation (3) can be transformed to the frequency domain (by applying the Z transform and substituting $(z = e^{-2\pi i f \Delta t})$ which yields

$$E(f) = A(f)X(f) \quad (4)$$

where $A(f) = -\sum_{j=0}^p A(j)e^{-i2\pi f\Delta t}$ and Δt is the data-sampling interval.

With $A(0) = -I$ (the identity matrix), Eq. (4) can be rewritten (with the sign of $A(j)$ changed) in another form:

$$X(f) = A^{-1}(f)E(f) = H(f)E(f) \quad (5)$$

The matrix H is called the transfer matrix of the system. It contains the information flow among data channels constituting the system. $H(f)$ is asymmetric matrix, so it allows for finding causal dependencies between different combinations of electrodes.

The causal dependence and information flow between different combinations of electrodes while speaking and imagining of speaking has been estimated as mentioned in Sandhya et al. (2015).

Directed Transfer Function (DTF) The normalized version of DTF (Kaminski et al. 2001) is defined as:

$$\gamma_{ij}^2(f) = \frac{|H_{ij}(f)|^2}{\sum_{m=1}^k |H_{im}(f)|^2} \quad (6)$$

where H_{ij} is an element of a transfer matrix of MVAR model. The normalization has been performed in such a way that γ_{ij} represents a ratio of the inflow of information to channel i from channel j to the inflows of information to channel i from all the channels. The normalized DTF value varies from 0 to 1. It indicates both the direct flows and also indirect flows among the electrodes.

Partial Directed Coherence (PDC) PDC provides the frequency domain parameter. It is based on modelling time series by multivariate autoregressive (MVAR) processes (Baccala and Sameshina 1999). PDC between two signals i.e. from signal j to signal i is given by:

$$P_{ij} = \frac{A_{ij}(f)}{\sqrt{a_j^*(f)a_j(f)}} \quad (7)$$

where $A_{ij}(f)$ is the Fourier transform of the MVAR coefficient matrix. $a_j(f)$ is the j th column of $A(f)$ matrix and the $a_j^*(f)$ denotes the complex conjugate operation. PDC is used to analyze the direct informational flow between different electrode combinations. PDC represents the ratio between the outflows from channel j to channel i to all the outflows to channel j .

Entropy measures

Information theory measures are used to quantify the information of a discrete random variable X . In general

Shannon entropy quantifies the reduction in the uncertainty of variable X when it is measured (Shannon and Weaver 1949) which given by:

$$H(X) = -\sum_x p(x) \log_2 p(x) \quad (8)$$

where $p(x)$ is the probability of a given symbol.

Transfer entropy Transfer entropy is a non-linear extension of Granger causality (Schreiber 2000). This property can be useful in analyzing non-linear signals where minimal a priori knowledge is available. Transfer entropy estimates directional relationships between two time series variables. Transfer entropy of two time series from x to y :

$$T_{X \rightarrow Y} = \sum_{y_{t+1}, y_t^n, x_t^m} p(y_{t+1} | y_t^n, x_t^m) \log \left(\frac{p(y_{t+1} | y_t^n, x_t^m)}{p(y_{t+1} | y_t^n)} \right) \quad (9)$$

The above equation measures the directed information flow from X to Y . Transfer entropy was computed for inter-hemispheric and intra-hemispheric frontal and temporal electrodes.

Data analysis

The features extracted from brain connectivity estimators and entropy measures were fed to various deep learning techniques for classifying a vowel. The deep learning techniques employed in this work are Recurrent Neural Networks (RNN) and Deep Belief Networks (DBN).

Recurrent Neural Network

Recurrent neural networks (RNN) execute the same task for every element of a given sequence, with the output being dependent on the previous input. The entire output sequence generated over time is considered the result of the computation and they have a memory which captures information about what has been calculated so far. For example, to predict the next word in a sentence it's better to know which words came before it. Many varieties of RNN have been proposed, such as Elman networks (Elman 1990), Jordan networks (Jordan 1990), time delay neural networks (Lang et al. 1990) and echo state networks (Jaeger 2001). The network structure of a typical RNN is depicted in Fig. 5 (Mohammadi et al. 2015). Each hidden layer holds a number of neurons performing a linear matrix operation on its input. For each time-step t , x_t is input to the hidden layer to produce a prediction output \hat{y} and output features h_t . W^{hx} is the weight matrix for the input x_t . h_{t-1} represents the output of the previous time step $t-1$.

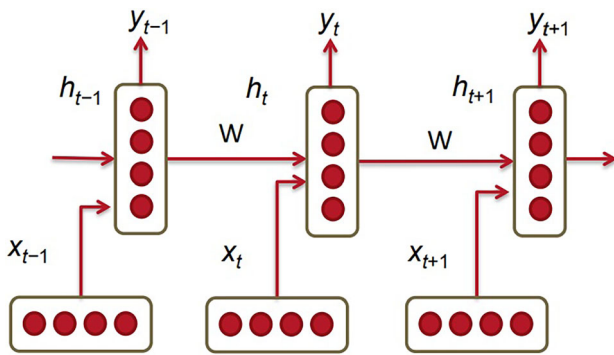


Fig. 5 RNN Architecture

$$h_t = Wf(h_{t-1}) + W^{hx}x_t \quad (10)$$

$$\hat{y} = W^{(S)}f(h_t) \quad (11)$$

Deep Belief Network

Deep Belief Network (DBN) is composed of multiple layers of typical variant of Boltzmann machine called Restricted Boltzmann Machine (RBM). The variables that are not externally available can have binary values and are often called hidden units. The top and the lower layers are connected through top down directed connections between them and form an associative memory. The states of the units in the lowermost (visible) layer represent the output data vector (Hinton 2007). In this work, a DBN with 7 hidden layers was built. DBN itself consists of three hidden layers with 1000 units (RBMs) per layer. The DBN trained itself with an unsupervised learning rate of 0.01.

RBM is a two layered Markov random field that has one layer of hidden units and one layer of visible units. RBMs can be represented as bipartite graphs as shown in Fig. 6, where all I visible units are connected to J hidden units, and there are no single unit in visible layer that are connected to any other unit in the same visible unit, similarly there are no hidden unit–hidden unit connections.

A practical implementation of the RBM training is given by Hinton (Hinton 2012). For a Gaussian (visible)–Bernoulli (hidden) RBM, the energy function is given by,

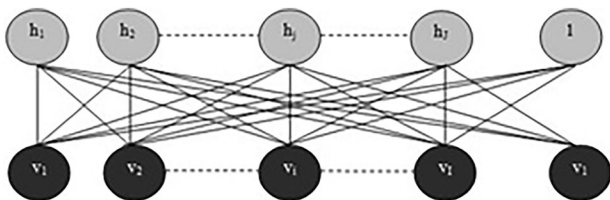


Fig. 6 RBM layer

$$E(v, h; \theta) = - \sum_{i=1}^I \sum_{j=1}^J w_{ij} v_i h_j + \frac{1}{2} \sum_{i=1}^I (v_i - b_i)^2 - \sum_{j=1}^J a_j h_j \quad (12)$$

where $E(v, h; \theta)$ is the energy function over the visible units v and hidden units h , w_{ij} is the symmetric interaction term between visible unit and hidden unit.

RBM weights are given by

$$\Delta w_{ij} = E_{data}(v_i h_j) - E_{model}(v_i h_j) \quad (13)$$

where $E_{data}(v_i h_j)$ is the expectation in the training set and $E_{model}(v_i h_j)$ is the same expectation defined by the model.

Similarly, Deep Belief Network (DBN) has been trained for classifying the vowels for two different protocols [Protocol I- Speech (S), Protocol II-Speech imagery (SI)]. The Deep belief network structure consists of a single input layer with 6 input units each representing a feature corresponding to a particular vowel, 4 hidden layers with number of units 500, 500, 1000, 2000 in h_1, h_2, h_3, h_4 respectively and a single output layer with 2 output units representing each class.

Classification of vowels from speech and speech imagery protocol using RNN and DBN

The functional and effective brain connectivity features were extracted from all the EEG frequency bands (delta, theta, alpha, beta and gamma) for the chosen inter and intra hemispheric electrode combinations. From the extracted features individual datasets for each frequency band were obtained for both the processes (Speech-S, Speech imagery-SI). For speech process- speech task (articulation of a word) alone has been considered for training and testing. Similarly for speech imagery process, imagined speech task has been considered. The datasets were obtained with total of four electrode combinations (F3, F7, T3, T5) with their corresponding brain connectivity features (EEG coherence, DTF, PDC and Transfer Entropy). Recurrent Neural Networks (RNN) and Deep Belief Networks (DBN) were used for classifying the vowels. Recurrent neural networks were trained in such a way that it identifies the vowel for example as either 'a' or not 'a' and Deep Belief Networks has been trained to classify the corresponding vowel based on the input training dataset. In order to decide the amount of training and testing data leave one out cross validation technique has been implemented. Based on the nature of dataset involved in this work and to perform an unbiased validation this technique (Martin et al. 2016) was chosen. As an outcome of the validation technique, from a total six subjects, three subjects' data were used for training and three subject's data were used for testing. The classifier performance was measured in terms of the

accuracy. The pipeline for classification procedure for vowels using RNN is shown in Fig. 7.

The size of the dataset chosen for one EEG band and for one vowel in speech and speech imagery processes has been represented in Table 2. Similarly for other bands and other vowels the same dimension has been retained to train the classifiers. These datasets were trained and tested for each process using RNN and DBN for the classification of vowels.

The training and testing datasets for each band were obtained based on the following algorithm using CVC words. The algorithm was designed in such a way that from a total of six subjects' data, three subjects' data were used as training set and three subjects' data were used as testing set for speech and speech imagery protocols. These datasets were trained and tested for each protocol by using RNN (ML1) and DBN (ML2). This was done for 10 CVC words per each vowel.

The training and testing datasets are obtained based on the following algorithm.

Results and discussions

EEG coherence analysis

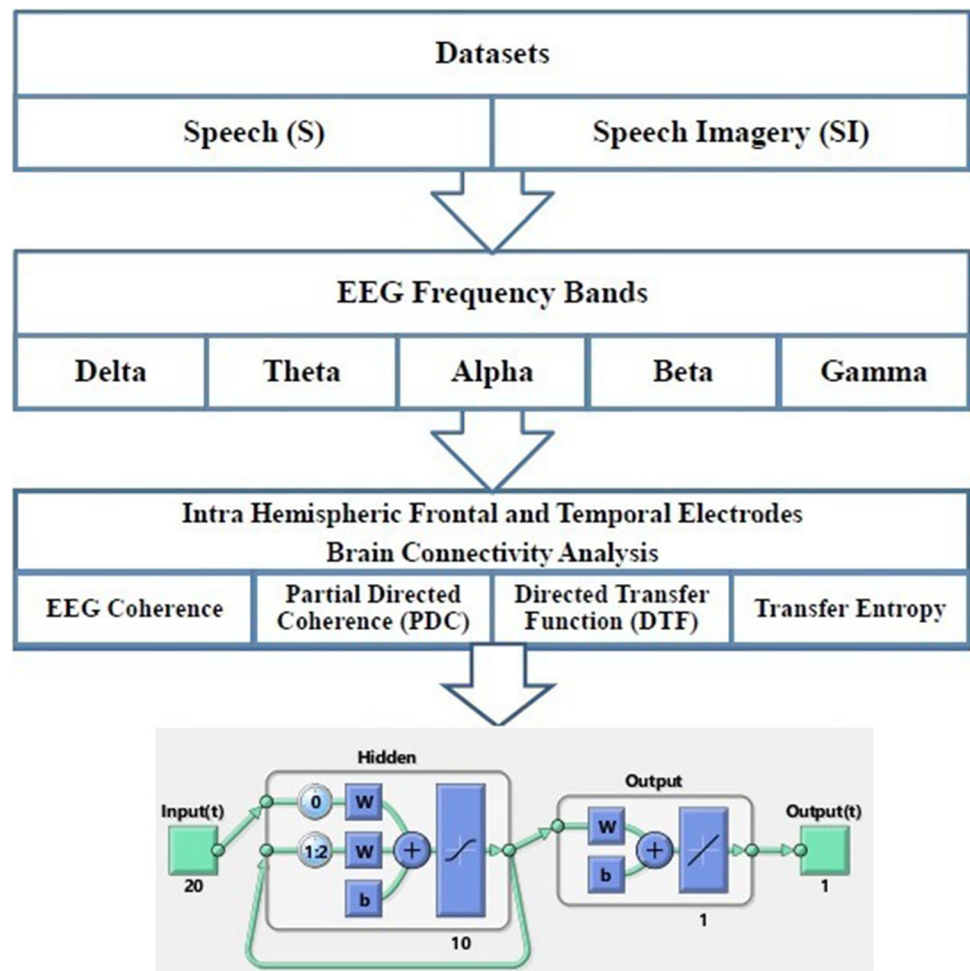
The inter and intra hemispheric cross-coherences have been estimated from the frontal, temporal and occipital electrodes. Cross-coherence values have been compared for two words that contain the same vowel and one consonant in common (e.g. 'Car' and 'Can'). Also, it was seen that Consonant–Vowel–Consonant (CVC) words of the same vowel having one consonant in common had similar coherence values. Table 3 represent the intra-hemispheric cross-coherence obtained while speaking the representative CVC word 'Can' respectively. Table 4 represent the intra-hemispheric cross-coherence obtained while imagining of speaking the representative CVC word 'Can' respectively. Specific to frequency bands, it has been observed from intra-hemispheric coherence analysis that Gamma band were dominant band while speaking a CVC word and theta band were dominant imagining of speaking the CVC word.

Training and Testing Algorithm:

1. Input: Dataset (N), Target (T), Computational models (RNN-ML1, DBN-ML2)
 2. Divide the Dataset (N) into two disjoint subsets T_r (Training) and T_t (Testing)
 3. Train models M L1 and ML2 on T_r for two mutually-exclusive protocols
 4. Assign protocols, PI for Speech, PII for Speech Imagery

for PI in size $N[m*n*l]$
 $m=1$ (no of subjects)
 $n=1$ (no of CVC words)
 $l=1$ (no of features)
 while $m \leq 3, n \leq 10, l \leq 20$ do
 Training1 ($Tr1[30 * 20]$)
 Train RNN (ML1) for input dataset $Tr1$
 for PII in size $N[m*n*l]$
 while $m \leq 3, n \leq 10, l \leq 20$ do
 Training2 ($Tr2[30 * 20]$)
 Train RNN (ML1) for input dataset $Tr2$
 5. Repeat Step 1:4 for DBN [(ML2)]
 6. Evaluate the performance of the models on Testing dataset T_e

for PI in size $N[m*n*l]$
 while $m \leq 3, n \leq 10, l \leq 20$
 Testing1 ($Te1[30 * 20]$)
 Test RNN (ML1) for test dataset $Te1$
 for PII in size $N[m*n*l]$
 while $m \leq 3, n \leq 10, l \leq 20$
 Testing2 ($Te2[30 * 20]$)
 Test RNN (ML1) for test dataset $Te2$
 7. Repeat Step 6 for DBN [(ML2)]
 8. Terminate once the vowels are classified according to their respective classes
-

Fig. 7 Classification procedure using RNN**Table 2** Dataset for each protocol

Protocol	Training set	Testing set
Speech-S (PI)	30 * 20 (600)	30 * 20 (600)
Speech imagery-SI (PII)	30 * 20 (600)	30 * 20 (600)

Figures 8 and 9 show the inter hemispherical coherence obtained while speaking and imagining of speaking the CVC word ‘Can’ with respect to vowel ‘a’. From Figs. 8 and 9 it has been observed that theta band is active in the inter-hemispheric coherence analysis for both the tasks.

Effective brain connectivity analysis while of speaking and imagining of speaking the CVC word

For a better understanding on the correlations of brain during speech and speech imagery processes, effective brain connectivity estimators such as Partial Directed

Table 3 Intra-hemispheric coherence variations obtained while speaking the representative CVC word

Electrode pairs	Intra-hemispheric coherence				
	Delta	Theta	Alpha	Beta	Gamma
F3–F7	0.393	0.574	0.500	0.721	0.951
F3–T3	0.296	0.581	0.546	0.615	0.718
F3–T5	0.491	0.461	0.620	0.444	0.639
F7–T3	0.597	0.387	0.582	0.336	0.597
F7–T5	0.494	0.463	0.429	0.664	0.634
T3–T5	0.398	0.488	0.428	0.537	0.555

The value highlighted in Bold indicates that the coherence value of the electrode combination F3–F7 is high in gamma band

Coherence (PDC) and Directed Transfer Function (DTF) were computed for each frequency band (delta, theta, alpha, beta and gamma).

Table 4 Intra-hemispheric coherence variations obtained while imagining of speaking the representative CVC word

Electrode pairs	Intra-hemispheric coherence				
	Delta	Theta	Alpha	Beta	Gamma
F3–F7	0.596	0.570	0.547	0.429	0.528
F3–T3	0.498	0.475	0.646	0.595	0.602
F3–T5	0.356	0.583	0.747	0.497	0.489
F7–T3	0.296	0.697	0.680	0.465	0.760
F7–T5	0.388	0.874	0.826	0.361	0.563
T3–T5	0.468	0.987	0.527	0.413	0.688

The value highlighted in Bold indicates that the coherence value of the electrode combination of T3–T5 is high in theta band

Effective brain connectivity while speaking the CVC word

PDC has been estimated for frontal (F7, F3, F4, F8), temporal electrodes (T7, T8, P7, P8) and occipital electrodes while speaking the CVC word ‘Can’ in single trial. A 10×10 matrix were obtained for the frontal, temporal

and occipital electrode combinations. These matrixes were obtained for all five sub-band EEG frequencies (delta, Theta, alpha, beta and gamma). In all cases it has been found that PDC values of gamma band were dominant while speaking the CVC word. Figure 10 shows the directional connectivity estimated using PDC during speaking of the CVC word ‘Can’ in gamma band with respect to full view of the brain (Lateral, medial, dorsal and ventral). The electrodes are represented in terms of Brodmann’ areas. Frontal electrodes are represented as F7–47L, F8–47R, F3–8L, F4–8R (47 L&R-Inferior Prefrontal Gyrus, 8 L&R-Frontal Eye Field with respect to left and right hemisphere of the brain). Similarly temporal electrodes are represented as T3–22L, T4–22R, T5–37L, T6–37R (22 L&R- Wernicke’s area of left and right hemisphere, 37 L&R-Fusiform Gyrus of left and right hemisphere of the brain) and occipital electrodes are represented as 18L and 18R. The scale ranges from blue to maroon color. The maroon color indicates the maximum flow of information from source to destination electrode whereas the blue color indicates the lesser connectivity or

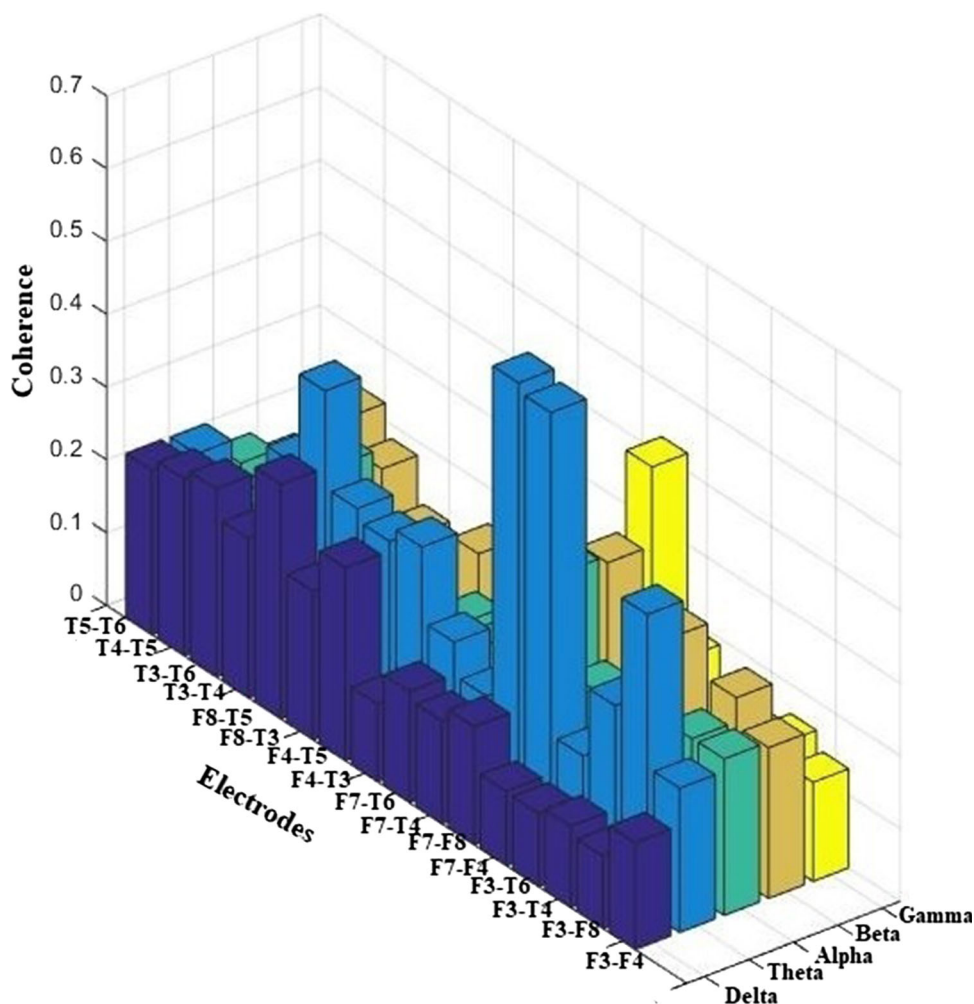
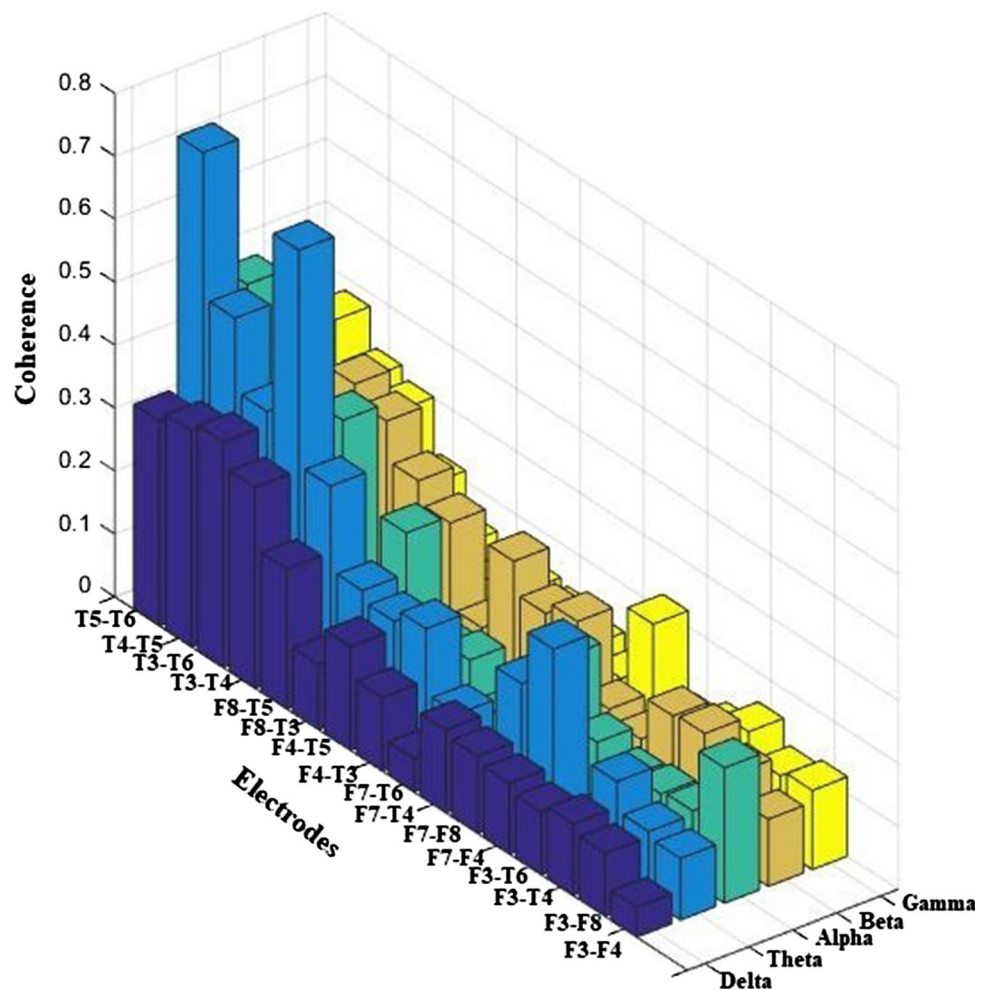
Fig. 8 Inter-hemispheric sub-band frequency coherence variations obtained from the chosen electrode pairs while speaking a representative CVC word

Fig. 9 Inter-hemispheric sub-band frequency coherence variations obtained from the chosen electrode pairs while imagining of speaking a representative CVC word



less amount information flow from one region to other region. From Fig. 10 it was observed that the left hemisphere frontal electrodes were activated during the speech process. The information flow (i.e. direction and strength of the connectivity) from left hemisphere frontal electrode combination (F3 with F7) was found to be higher while speaking the CVC word ‘Can’. Thus it was inferred that the directional connectivity from left hemisphere to right hemisphere was found to be high than that of the right hemisphere to left hemisphere of the brain.

The connectivity matrix of PDC obtained while speaking of a CVC word is shown in Fig. 11. From Fig. 11, it has been observed that the connectivity between F7 and F3 was found to be the highest value compared to other electrode combinations. The inter hemispheric electrode combinations showed a lower connectivity during speaking of a CVC word. The color coded scale ranges from blue to yellow color wherein blue indicates lower connectivity and yellow indicates higher connectivity between the electrodes. Lower connectivity was observed between temporal electrode combinations. The connectivity between frontal

electrodes was high compared to that of temporal combinations which proves that Broca’s area was active during speech process.

Figure 12 shows the Directed Transfer Function (DTF) in axial view of the brain while speaking the CVC word. In all cases it has been found that DTF values of gamma band were dominant while speaking the CVC word. The color coded scale for this parameter ranges from blue to maroon colour wherein blue colour indicates lower connectivity and maroon colour indicates stronger connectivity. From Fig. 12 it has been observed that there exists a stronger connectivity in left hemisphere while speaking the CVC word whereas the connectivity in right hemisphere is lesser compared to left hemisphere of the brain. Moreover, the information from left to right hemisphere of the brain is more than right to left hemisphere of the brain. Predominantly electrode combination (F3 with F7) labelled as 8L and 47L respectively, of left hemisphere displayed a stronger connectivity compared to other electrode combinations.

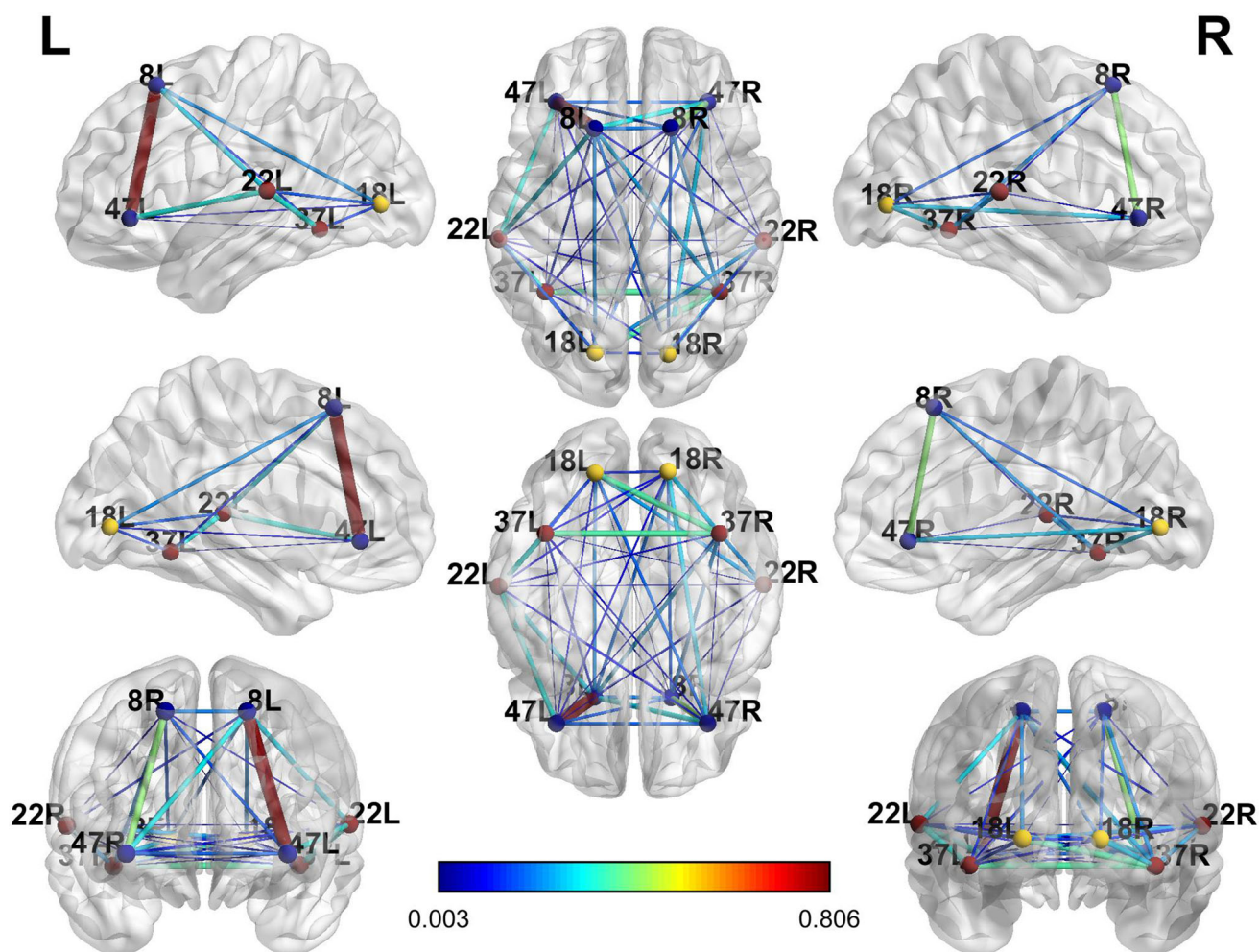


Fig. 10 Estimation of directional connectivity using PDC for intra and inter hemispheric electrodes from EEG signals obtained while speaking a representative CVC word

Effective brain connectivity while imagining of speaking the CVC word

PDC has been estimated for frontal (F7, F3, F4, F8), temporal electrodes (T7, T8, P7, P8) and occipital electrodes while imagining of speaking the CVC word in single trial. A 10×10 matrix was obtained for the frontal, temporal and occipital electrode combinations as discussed for speech process. These matrices were obtained for all five sub-band EEG frequencies (delta, theta, alpha, beta and gamma). In all cases it has been found that PDC values of theta band were dominant while imagining of speaking the CVC word. Figure 13 shows the directional connectivity estimated using PDC during imagining of speaking the CVC word ‘Can’ in theta band with respect to lateral and medial, coronal, axial view of the brain. The electrodes are represented in terms of Brodmann’ areas. From Fig. 13 it was observed that the left hemisphere temporal electrodes were activated during the speech process. The information

flow (i.e. direction and strength of the connectivity) from left hemisphere temporal electrode combination (T3 with T5) was found to be higher while imagining of speaking a CVC word ‘Can’. Thus it was inferred that the directional connectivity from left hemisphere to right hemisphere was found to be high than that of the right hemisphere to left hemisphere of the brain.

The connectivity matrix of PDC obtained while imagining of speaking of the CVC word is shown in Fig. 14. From Fig. 14, it has been observed that the connectivity between T3 and T5 was found to be the highest when compared to other electrode combinations. The color coded scale ranges from blue to yellow color wherein blue indicates lower connectivity and yellow indicates higher connectivity between the electrodes. Lower connectivity was observed between frontal electrode combinations. The connectivity between temporal electrodes was high compared to that of frontal combinations which proves that Wernicke’s area was active during speech imagery process.

Fig. 11 PDC connectivity matrix while speaking a representative CVC word

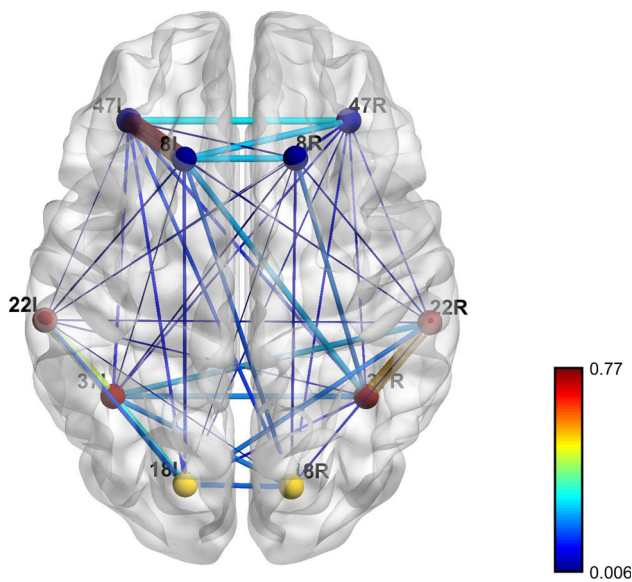
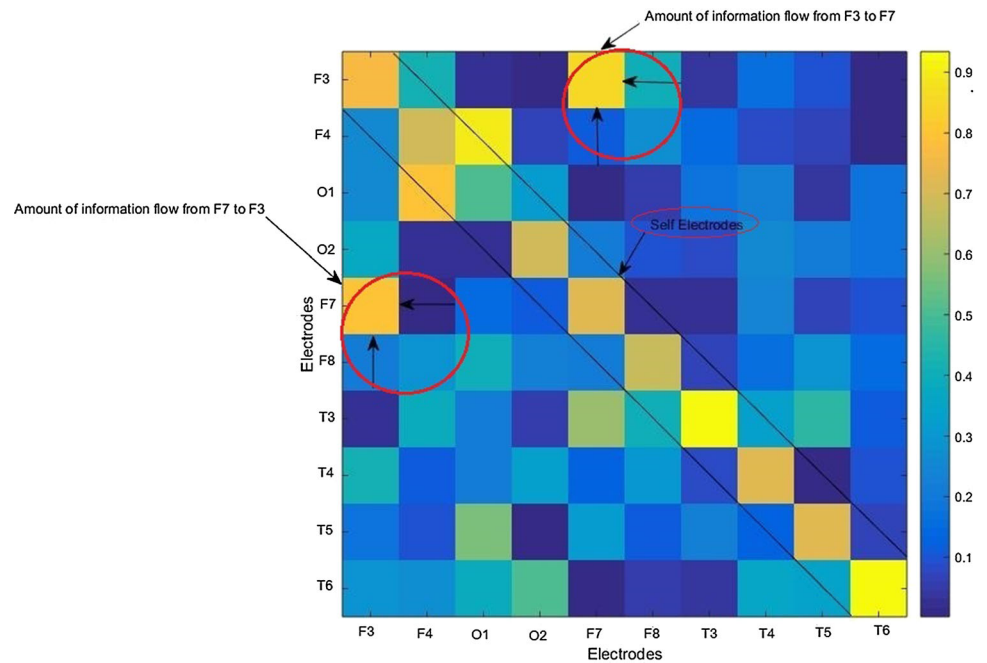


Fig. 12 DTF obtained from EEG signal while speaking a representative CVC word

Figure 15 shows the DTF in axial view of the brain while imagining of speaking the CVC word ‘Can’. The color coded scale ranges from blue to maroon colour wherein, blue colour indicates lower connectivity and maroon colour indicates stronger connectivity. From Fig. 15 it has been observed that there exists a stronger connectivity in left hemisphere, while imagining of speaking the CVC word whereas the connectivity in right hemisphere is lesser compared to left hemisphere of the brain. The information from left to right hemisphere of the brain is more than right to left hemisphere of the brain. The

electrode combination (T3 with T5) which is labelled as 22L and 37L showed a stronger connectivity compared to other electrode combinations.

Transfer entropy analysis

Table 5 shows the transfer entropy while speaking and imagining of speaking the representative CVC word ‘Can’. Outflow of information represents the signal direction and strength from one particular electrode to other electrode whereas inflow of information represents the information into a particular electrode from other different electrodes. From Table 5 it has been observed that the amount of information shared by frontal electrodes and temporal electrodes are high compared to other electrodes. The inflow and outflow of information for F3, F7, T3, T5 are high compared to other electrodes. The inferences that were observed from Table 5 are represented in Figs. 16 and 17 for speech and speech imagery processes in lateral, medial and dorsal view of the brain. From transfer entropy results it is evident that left frontal electrodes (F3 with F7) were active during speech process and left temporal electrodes (T3 with T5) were active during speech imagery process.

Classification of vowels using Recurrent Neural Networks (RNN) and Deep Belief Networks (DBN)

The datasets for each EEG band was obtained based on the brain connectivity features (Coherence, PDC, DTF, Transfer Entropy). From brain connectivity analysis it has

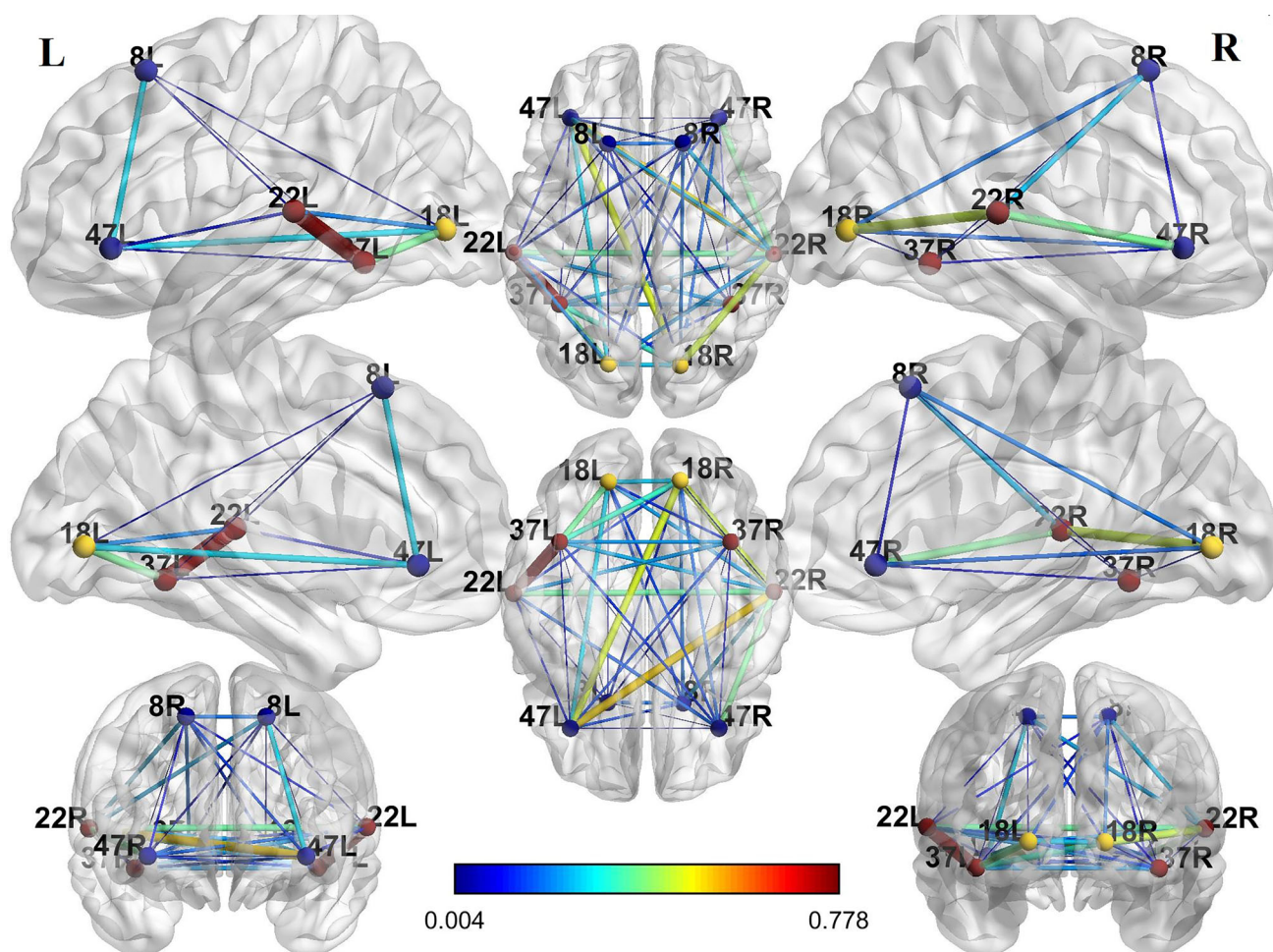
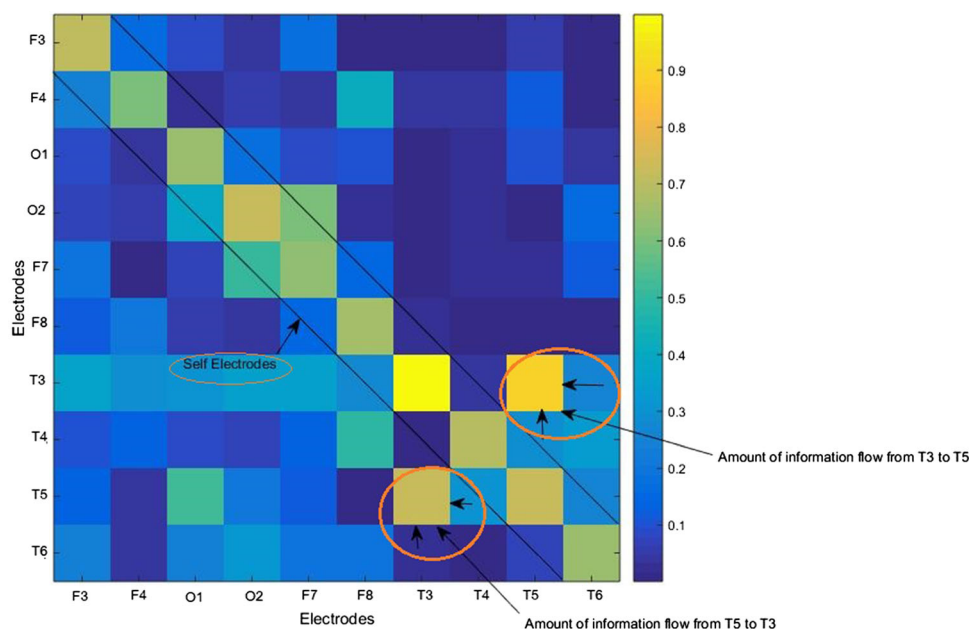


Fig. 13 Estimation of directional connectivity using PDC for intra and inter hemispheric electrodes from EEG signals obtained while imagining of speaking a representative CVC word

Fig. 14 PDC connectivity matrix while imagining of speaking a representative CVC word



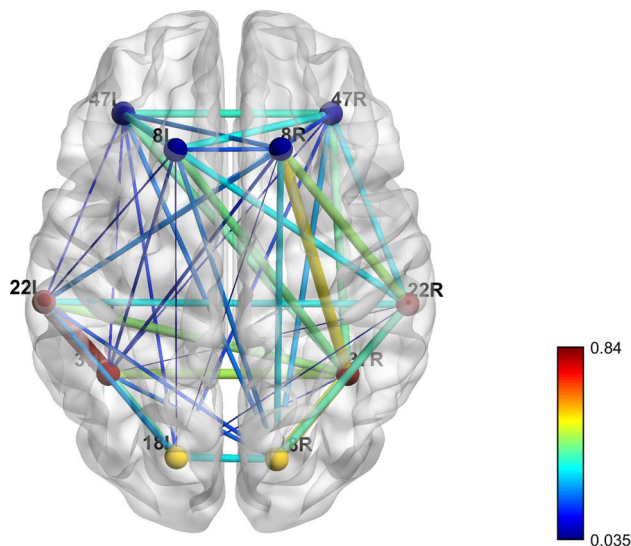


Fig. 15 DTF obtained from EEG signal while imagining of speaking the representative CVC word

been observed that gamma band was found to be dominant band while speaking whereas theta band was dominant band while imagining of speaking. It has been observed that the classification accuracy of each vowel in gamma band was high for speech protocol and classification accuracy of each vowel in theta band was high for speech imagery protocol compared to other EEG bands. Figure 18 shows the classification accuracy for each vowel using RNN and DBN for protocol-I speech (S) in gamma band and for protocol-II speech imagery in theta band.

Classification accuracy using RNN

The overall classification accuracy for vowels ‘o’ and ‘e’ was found to be higher in both the protocols while using RNN. For vowel ‘e’, ‘o’ and ‘u’ the classification accuracy was found to be higher when compared to other vowels when RNN was trained with Protocol I. Similarly, for vowels ‘e’ and ‘o’ the classification accuracy was higher in Protocol II.

Classification accuracy using DBN

On the other hand, vowel ‘i’ has the maximum accuracy compared to other vowels while DBN was trained and tested with Protocol I. Classification accuracy was found to be varying from 79 to 81% for vowels ‘e’, ‘i’, ‘o’ when the network was trained with Protocol II. In general, for both the protocols, the overall accuracy specific to gamma band was found to be higher.

The overall classification accuracy for DBN was found to be high when compared to RNN when trained and tested with two different protocols (Speech-S and Speech imagery- SI) separately. For protocol I (Speech-S), the overall classification accuracy for RNN was observed to be 75% whereas, for DBN it was 82%. Similarly, for protocol II (Speech imagery-SI) the overall classification accuracy for RNN was 70% whereas, for DBN it was found to be 80%. From the above results, it is evident that DBN outperforms RNN for both the protocols proving that a vowel can be identified from any CVC word that is being articulated and imagined by an individual.

Table 5 Transfer Entropy for a representative CVC word while speaking and imagining of speaking

Electrode pairs	Transfer entropy			
	Speaking the CVC word		Imagining of Speaking the CVC word	
	In-flow of information	Out-flow of information	In-flow of information	Out-flow of information
F3–F4	0.661	0.469	0.527	0.433
F3–F7	0.801	0.880	0.541	0.509
F3–F8	0.431	0.575	0.491	0.445
F4–F7	0.457	0.481	0.337	0.323
F4–F8	0.370	0.444	0.473	0.442
F7–F8	0.594	0.664	0.365	0.409
T3–T4	0.488	0.518	0.657	0.565
T3–T5	0.369	0.369	0.724	0.894
T3–T6	0.402	0.409	0.581	0.534
T4–T5	0.485	0.392	0.531	0.586
T5–T6	0.256	0.299	0.554	0.617

The value highlighted in Bold indicates the transfer entropy value of the F3–F7 electrode combination is high while speaking in both inflow and outflow of information. Similarly, transfer entropy value of T3–T5 electrode combination indicates high value while imagining of speaking in both inflow and outflow of information

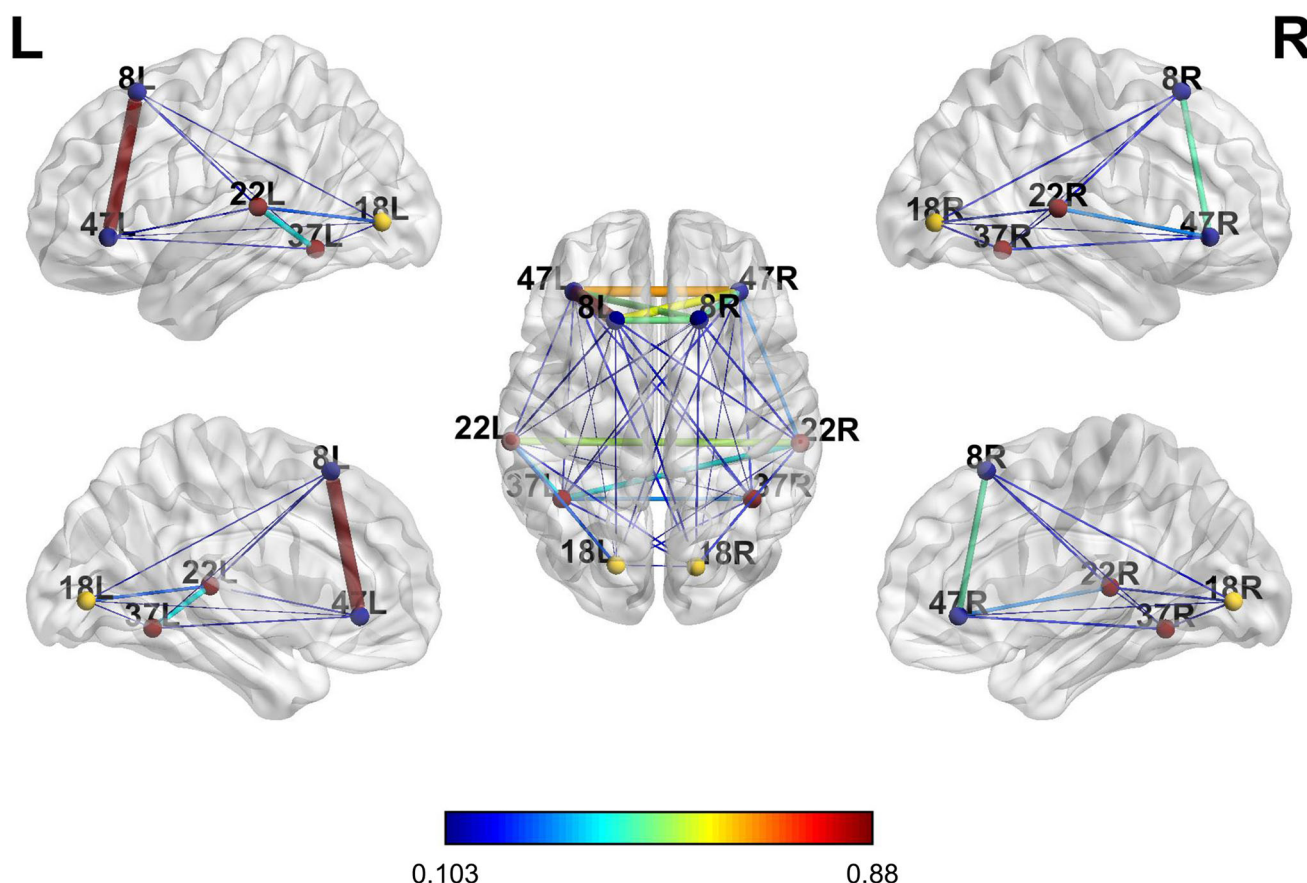


Fig. 16 Transfer entropy analysis while speaking a representative CVC word

Conclusion

It is a known fact that Speech is a powerful communication medium to voice one's thought, verbally and in many cases, patients with neural and speech impairments intend to speak something but end up in speaking a different one. In such situations, identifying the intended speech through thoughts seems very resourceful. Working towards such a process, this work addresses the implications of correlations that occur in the brain connectivity derived from EEG signals obtained during speech and speech imagery processes. Brain connectivity estimators derived from the EEG recordings have been found to reveal significant features from the spoken and imagined words. The hypothesis of identifying relevant features from EEG waves acquired while speaking as well as imagining of speaking has been exploited in this work in identifying the appropriate vowel involved in the task.

Through brain connectivity analysis, correlation between speech and speech imagery processes has been derived. Results confirm that the signals acquired from Broca's and Wernicke's area according to 10–20 electrode placement system (F3, F7, T3, T5) get more activated while speaking and imagining of speaking the word. The

results of increased functional connectivity in the frontal and temporal regions confirm the participation of Broca's area meant for speech production and Wernicke's area meant for speech recognition, in such speech imagery processes. The extracted features were then fed to Recurrent Neural Networks (RNN) and Deep Belief Networks (DBN) for the classification of vowels. The vowels were thus attempted to be identified through the features of the signals acquired during the different speech and speech imagery tasks. Classification accuracies of both the deep neural networks, DBN and RNN were obtained for each protocol. Classification results support the possibility of identifying the vowel from any CVC word that was imagined to be spoken.

Thus, this work aims at identifying words correctly, by identifying the vowel in any CVC word, spoken or thought by an individual. Even if the spoken word is misinterpreted, there appears to be a possibility of identifying the correct word by recognising the actual vowel that was thought to have been involved in the word using the combination of signal processing methods, brain connectivity estimators and machine learning techniques. Therefore, identifying the intended vowel from a set of CVC words by using the combination of above mentioned

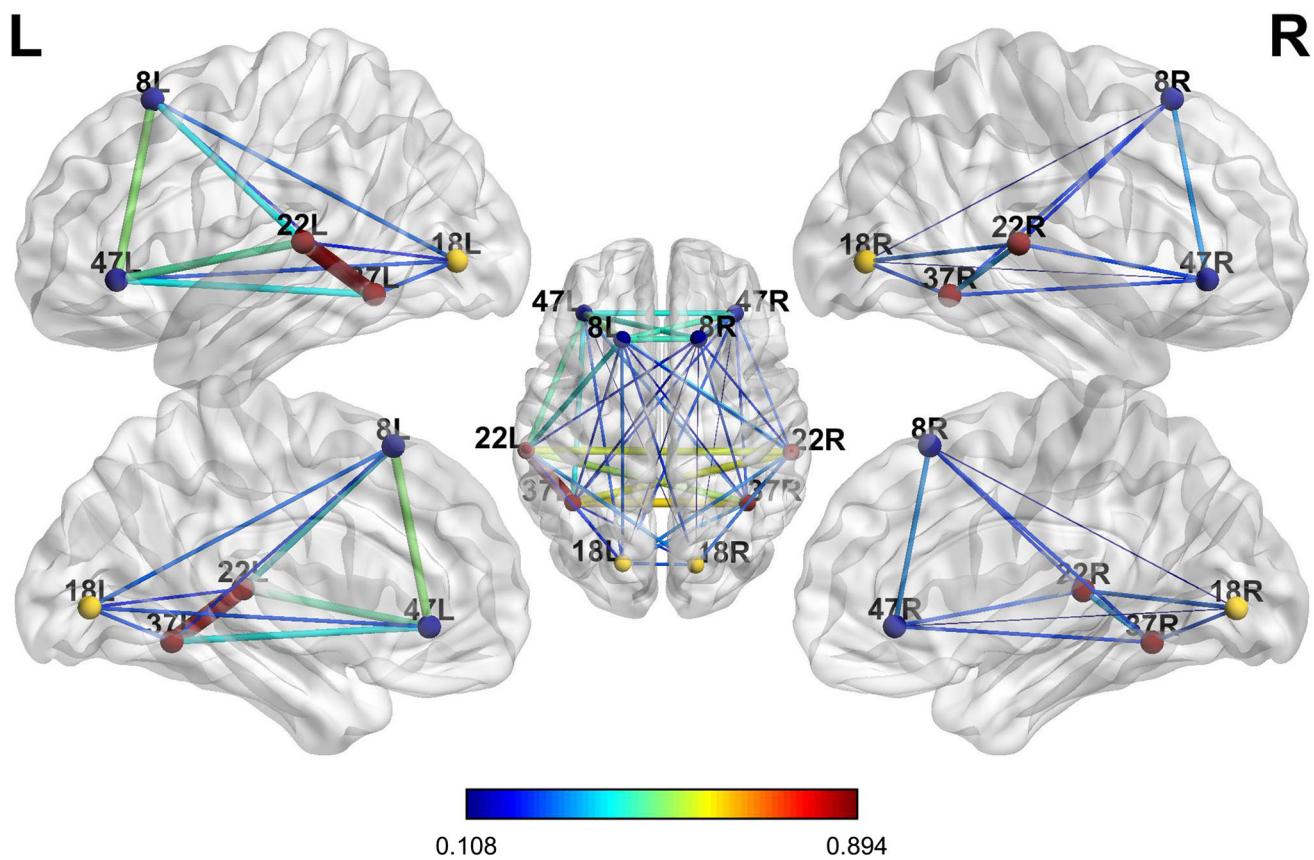


Fig. 17 Transfer entropy analysis while imagining of speaking a representative CVC word

Classification of vowels using RNN and DBN

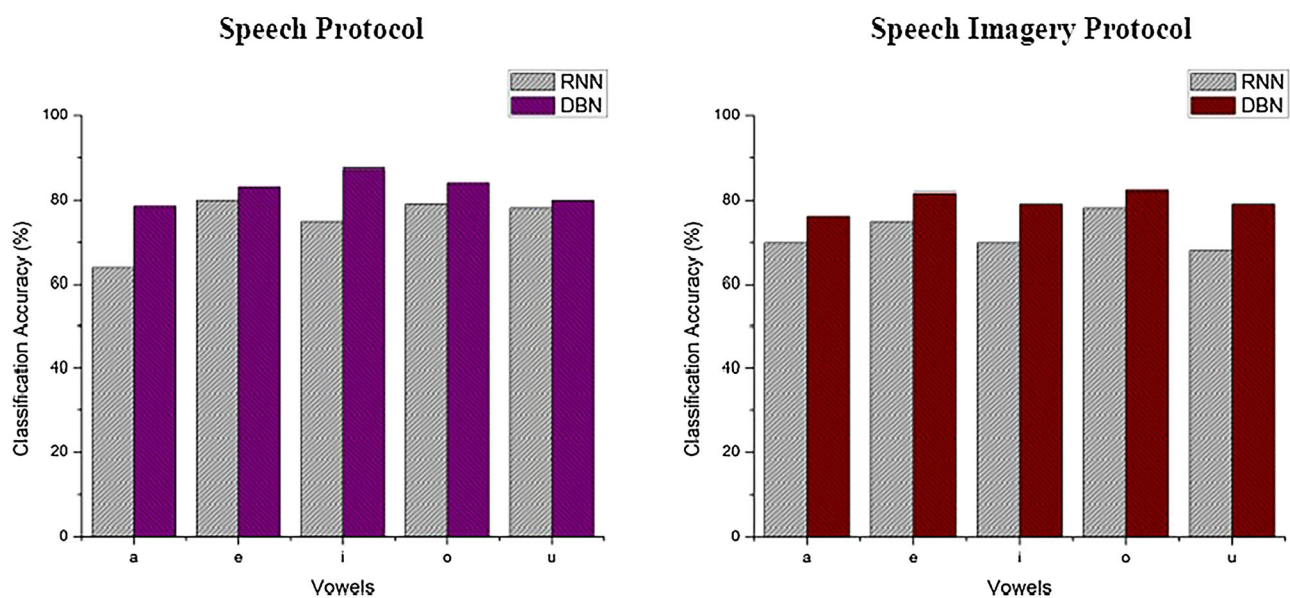


Fig. 18 Classification accuracy for each vowel using RNN and DBN

techniques, a control scheme for speech prostheses can be developed for the speech impaired people. This can clinically assist in bringing change in patients with speech disorders and their impaired communication to the external world. This work is preliminary in the era of speaking from one's thoughts as it involves EEG signals acquired, processed and machine learning involved to identify the simple vowels alone in the words. But it appears to be a promising source in believing the concept of reading the mind of speech impaired, through continuous attempts and unremitting experimental studies in the longer run.

Compliance with ethical standards

Conflict of interest Sandhya Chengaiyan, Anandha Sree Retnapan-dian and Kavitha Anandan declare that they have no conflict of interests.

Informed consent Informed Consent was obtained from all the individual participants included in the study.

Ethical approval Ethical approval was given by the Institutional Ethical Committee of Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India. This work was performed in the Department of Biomedical Engineering as per the guidelines of the Institutional Ethical Committee of SSN College of Engineering, India for human participants.

References

- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci USA* 98:13367–13372
- Baccala L, Sameshima K (2001) Partial directed coherence: a new concept in neural structure determination. *Biol Cybern* 84(6):463–474
- Baccala L, Sameshima K (1999) Using partial directed coherence to describe neuronal ensemble Interactions. *J Neurosci Methods* 94(1):93–103
- Bookheimer S (2002) Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25:151–188
- 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
- Elman JL (1990) Finding structure in time. *Cogn Sci* 14:179–211
- Ghitza O (2013) The theta syllable: a unit of speech information defined by cortical function. *Front Psychol* 4(138):1–5
- Ghitza O, Greenberg S (2009) On the possible role of brain rhythm in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66:113–126
- Granger C (1969) Investigating causal relations by econometric models and cross-spectral methods. *J Econom Soc* 37(3):424–438
- Hickok G, Poeppel D (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131–138
- Hinton GE (2007) Learning multiple layers of representation. *Trends Cogn Sci* 11(10):428–434
- Hinton GE (2012) A practical guide to training restricted Boltzmann machines. In: Montavon G, Orr GB, Muller KR (eds) *Neural networks: tricks of the trade*, vol 7700. Springer, Berlin, Heidelberg, pp 599–619
- Idrees BM, Farooq O (2016) EEG based vowel classification during speech imagery. In: *IEEE 3rd international conference on computing for sustainable global development (INDIACom)*, pp 1130–1134
- Jaeger H (2001) The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. *Bonn Ger Ger Natl Res Center Inf Technol GMD Tech Rep* 148(34):13
- Jordan MI (1990) Attractor dynamics and parallelism in a connectionist sequential machine. *IEEE Press, Piscataway*, pp 112–127
- Kaminski M, Blinowska KJ (1991) A new method of the description of the information flow. *Biol Cybern* 65:203–210
- Kaminski M, Ding M, Truccolo W, Bressler S (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern* 85(2):145–157
- Klem GH, Lüders HO, Jasper HH, Elger C (1999) The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol* 52(3):3–6
- Lang KJ, Waibel AH, Hinton GE (1990) A time-delay neural network architecture for isolated word recognition. *Neural Netw* 3(1):23–43
- Lin Y, Liu B, Liu Z, Gao X (2015) EEG gamma-band activity during audiovisual speech comprehension in different noise environments. *Cogn Neurodyn* 9:389–398
- Martin S, Brunner P, Iturrate I, Millán JDR, Schalk G, Knight RT, Pasley BN (2016) Word pair classification during imagined speech using direct brain recordings. *Sci Rep* 6:25803
- Mehta B, Chawla VK, Parakh M, Parakh P, Bhandari B, Gurjar AS (2015) EEG abnormalities in children with speech and language impairment. *J Clin Diagn Res* 9(7):CC04–CC07
- Min B, Kim J, Park HJ, Lee B (2016) Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram. *Biomed Res Int* 2016:1–11
- Mohammadi M, Mundra R, Socher R (2015) Deep learning for NLP. In: *Lecture notes: part IV²*, Stanford University, Spring
- Pearson J, Naselaris T, Holmes EA, Kosslyn SM (2015) Mental imagery: functional mechanisms and clinical applications. *Trends Cogn Sci* 19(10):590–602
- Pei X, Barbour DL, Leuthardt EC, Schalk G (2011) Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng* 8(4):046028
- Perrone-Bertolotti M, Rapin L, Lachaux JP, Baciau M, Loevenbruck H (2014) What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance and its relation to self-monitoring. *Behav Brain Res* 261:220–239
- Poeppel D, Hickok G (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402
- Price CJ, Crinion JT, Mac Sweeney M (2011) A generative model of speech production in Broca's and Wernicke's areas. *Front Psychol* 2:1–9
- Righi G, Tierney AL, Tager-Flusberg H, Nelson CA (2014) Functional connectivity in the first year of life in infants at risk for autism spectrum disorder: an EEG study. *PLoS ONE* 9(8):e105176
- Rojas DA, Ramos OL (2016) Recognition of Spanish vowels through imagined speech by using spectral analysis and SVM. *J Inf Hiding Multimed Signal Process* 7(4):889–897
- Rojas GM, Alvarez C, Montoya CE, de la Iglesia-Vayá M, Cisternas JE, Gálvez M (2018) Study of resting-state functional

- connectivity networks using EEG electrodes position as seed. *Front Neurosci* 12:235
- Sandhya C, Kavitha A (2015) Analysis of speech imagery using functional and effective EEG based brain connectivity parameters. *Intl J Cogn Inform Nat Intell* 9(4):33–48
- Sandhya C, Srinidhi G, Vaishali R, Visali M, Kavitha A (2015B) Analysis of speech imagery using brain connectivity estimators. In: *Proceedings of the IEEE 14th international conference on cognitive informatics and cognitive computing*, Tsinghua University, Beijing, China, pp 352–359
- Sandhya C, Anandha Sree R, Kavitha A (2016) Analysis of speech imagery using consonant–vowel speech syllable pairs and brain connectivity estimators. In: *Second international conference on biomedical signals, systems, images*, IIT Madras, India
- Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85(2):461–464
- Shannon CE, Weaver W (1949) *The mathematical theory of communication*. University of Illinois Press, Urbana
- Shibata T, Suhara Y, Oga T (2004) Application of multivariate autoregressive modelling for analyzing the interaction between EEG and EMG in humans. *Int Congr Ser* 1270:249–253
- Sree RA., Kavitha A (2017) Vowel classification from imagined speech using sub-band EEG frequencies and deep belief networks. In: *2017 fourth international conference on signal processing, communication and networking (ICSCN)*, pp 1–4
- Suppes P, Han B, Epelboim J, Lu ZL (1999) Invariance between subjects of brain wave representations of language. *Proc Natl Acad Sci USA* 96:12953–12958
- Teplan M (2002) Fundamentals of EEG measurements. *Meas Sci Rev* 2(2):1–11
- Thatcher RW, Biver CJ, North D (2004) EEG coherence and phase delays: comparisons between single reference, average reference and current source density. *NeuroImaging Lab, VA Medical Center, Bay Pines, FL*. <http://www.appliedneuroscience.com/Comparisons-Commonref-Avelaplacian.pdf>, 64
- Weiss S, Muller HM (2003) The contribution of EEG coherence to the investigation of language. *Brain Lang* 85:325–343
- Wester M (2006) *Unspoken speech: speech recognition based on electroencephalography*. Master's thesis, Institute for Theoretical Computer Science, Karlsruhe Institute of Technology, Karlsruhe, Germany
- Widmann A, Schroger E, Maess B (2015) Digital filter design for electrophysiological data: a practical approach. *J Neurosci Methods* 250:34–46
- Yoshimura N, Nishimoto A, Belkacem AN, Shin D, Kambara H, Hanakawa T, Koike Y (2016) Decoding of covert vowel articulation using electroencephalography cortical currents. *Front Neurosci* 10(175):1–15

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.