2017 4th International Conference on Signal Processing, Communications and Networking (*ICSCN* -2017), March 16 – 18, 2017, Chennai, INDIA

Vowel Classification from Imagined Speech Using Sub-band EEG frequencies and Deep Belief Networks

R. Anandha Sree, A. Kavitha
Centre for Healthcare Technologies,
Department of Biomedical Engineering,
SSN College of Engineering,
Chennai, Tamil Nadu, India.
Email: anandhasreer@gmail.com*, kavithaa@ssn.edu.in

Abstract: **This work has focused on the possibilities of classifying vowels 'a', 'e', 'i', 'o', 'u' from EEG signals, that has been derived while imagining the vowels, with minimum input features. The EEG signals have been acquired from 5 subjects while imagining and uttering the vowels during a well defined experimental protocol, have been processed and segmented using established signal processing routines. The signals have been segmented under various sub-band frequencies and subjected to Db4 Discrete Wavelet Transform. The various conventional and derived energy based features have been acquired from the sub-band frequency signals, trained and tested using Deep Belief Networks for classifying the imagined vowels. The experiments have been repeated on various electrode combinations. Results obtained from all sub-band frequency based features have shown a good classification accuracy. Further, classification protocol employing features that have been derived from each sub-band frequency has shown that the theta and gamma band frequency features have been more effective with a vowel classification accuracy ranging between 75-100%.**

Keywords - Electroencephalography (EEG); Discrete Wavelet Transform (DWT); Deep Belief Network (DBN)

I. INTRODUCTION

Speech has been the most common and vital means of communication among human beings. The partial or complete loss of audible speech has been due to many reasons like neural damage or neurodegenerative disorders. Research have been carried out on this wide problem, which mostly includes invasive procedures. For the past few years, researchers have shown interest on Brain Computer Interface (BCI) through non-invasive techniques like Electroencephalogram, Electromyogram[1]. EEG has been a non-invasive functional neuroimaging technique, which has measured the electric fields that have been generated by the activity in the brain, to obtain tangible information related to brain activity. The brain has reacted in a different way for each action performed and each action imagined. It has already been proved that imagined action can be identified only with the EEG signals. And more recently, studies have been performed on interpreting EEG signals recorded during imagined speech[2-4].

With the EEG recording of a person imagining that he has spoken something (speech imagery), there has been a possibility to identify what he has spoken. This work has put efforts to identify the vowel that has been imagined and to find out which frequency band (delta, theta, alpha, beta and gamma) has shown more activation during the speech imagery tasks. For the BCI systems, the use of optimized classification algorithm that categorizes a set of data into different classes has been essential. The accuracy of the classifier might vary depending on the type of application it has been used for. In medical applications the classification needs to be more accurate when compared to gaming applications.

The proposed work has aimed to classify the vowels 'a', 'e', 'i', 'o', 'u' from the imagined EEG signals. The EEG signals that have been acquired were pre-processed and segmented into the five frequency bands that have been analyzed separately. Then the Discrete Wavelet Transform has been used to extract features. The work has focused around the frontal and temporal regions of the brain, since they are related to the speech production and speech comprehension related tasks[5,6]. Hence various combinations of the frontal and temporal electrodes have been experimented over classification accuracy. Different combinations of the extracted features have been classified using the generative, probabilistic Deep Belief Network (DBN).

978-1-5090-4740-6/17/$31.00 ©2017 IEEE

II. METHODOLOGY

A. EEG Acquisition

The RMS EEG-32 super spec, with a resolution of 0.22V/ bit and 32 input channels, has been used for recording the EEG signals. All recordings have been conducted at a sampling rate of 128Hz. 19 Ag/AgCl paste/gel type electrodes have been placed on the scalp of the subjects based on the 10-20 electrode placement system[7]. The block diagram of the proposed system has been given in the figure 1.

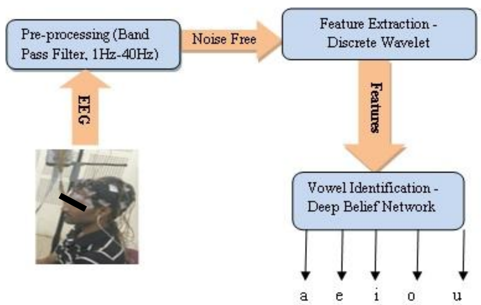

Figure 1: Block diagram of the proposed system

The experiment has been carried out in the Department of Bio-Medical Engineering of SSN College of Engineering, Chennai. Three female and two male subjects have been recruited in the process of data collection. The age of the subjects has been between 20 and 22 (mean age = 21 years). All the subjects have been right handed, with no visual, auditory or motor impairments and has had no history of any neurological disorders. The experiment has been conducted in a sound proof room. Visual stimuli has been given through a computer monitor placed in front of the subjects. The subjects have been seated comfortably in a chair, ensuring no motor activities at the time of EEG recording. A small quantity of gel has been used to improve electrical conductance. The impedance levels have been maintained below 10kΩ. The visual stimuli, utterance and imagery signals have been segmented using eye blink between each task. The task for each subject has spanned between 15 to 20 minutes that included two trails. Figure 2 has shown the experimental protocol of the EEG recording.

The overall recording course has consisted of four states:

1. A **rest** state for 5 seconds to physically and mentally complete one task and to get ready for the next task.
2. The **visual stimuli** state, in which the prompt text ('/a/', '/e/', '/i/', '/o/', '/u/') has been displayed on the monitor before the subject.
3. The **spoken** state, in which the subjects have uttered the prompt aloud.
4. The **imagery** state, in which the subjects have imagined speaking the vowel.

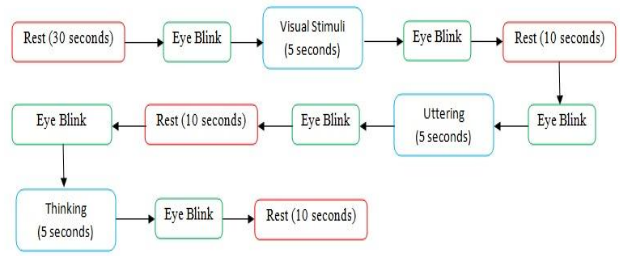

Figure 2: The Experimental Protocol

B. Pre-processing

The acquired EEG signals have been pre-processed using the interactive MATLAB toolbox EEGLAB[8]. The signals have been band-passed between 1Hz and 60Hz and mean normalized. And finally the signals have been segmented into different trials and each trial has been further segmented into the four states as described above with the help of eye blink signal/artifact. Then the data of channels F3, F7, F4, F8, T3, T4, T5, T6 have been chosen, based on the knowledge gained from the literatures[9]. The assumptions have made that the frontal and the temporal region of both the left and right hemisphere of the brain has shown more activation during the speech production and speech comprehension tasks.

C. Discrete Wavelet Transform Analysis

For discrete data, such as the data that has been used in the current work, the Discrete Wavelet Transform (DWT) has been better fit for analysis. A large selection of DWT mother wavelets have been available, but the Daubechies (Db) wavelet has been proved to be the most suitable family in similar applications [10]. So, in this work it has been decided to calculate the Daubechies orthogonal wavelets up to fifth level of decomposition.

The following energy coefficients[11] have been calculated from the obtained detailed coefficients, where the n^{th} sample of a wavelet decomposed detail at level i is $D_i(n)$:

- Root Mean Square (RMS)

$$RMS_i = \sqrt{\frac{1}{N}\sum_{n=1}^{N} D_i^2(n)} \tag{1}$$

- Mean Absolute Value (MAV)

$$MAV_i = \frac{1}{N}\sum_{n=1}^{N} |D_i(n)| \tag{2}$$

978-1-5090-4740-6/17/$31.00 ©2017 IEEE

- Integrated EEG (IEEG)

$$IEEG_i = \sum_{n=1}^{N}|D_i(n)| \quad (3)$$

- Simple Square Integral (SSI)

$$SSI_i = \sum_{n=1}^{N} |D_i(n)^2| \quad (4)$$

- Variance of EEG (VAR)

$$VAR_i = \frac{1}{N-1}\sum_{n=1}^{N} D_i^2(n) \quad (5)$$

- Average Amplitude Change (AAC)

$$AAC_i = \frac{1}{N}\sum_{n=1}^{N} |D_i(n+1) - D_i(n)| \quad (6)$$

The Daubechies wavelet has been used to analyze the channels(F3, F4, F7, F8, T3, T4, T5, T6) of each EEG record. The features RMS, MAV, IEEG, SSI, VAR and AAC have been calculated for the wavelet coefficients using (1) through (6). Figure 3 has shown the features extracted from a single channel during the imagery task. This process has been repeated for each band of all the events in our dataset. At the end of the calculations, a set of 19x5x6 (19 channels x 5 subjects x 6 features) features have been generated for each band - alpha, beta, gamma and theta (omitting the low frequency (delta) and the high frequency (noise)), from which the desired combination of channels of interest have been extracted. Then the features have been numerically represented in a format that has been suitable for use with neural network algorithms. Totally, 40 (5 vowels x 4 bands x 2 tasks) data sets have been generated. From each dataset 80% of the data have been used for training and the remaining 20% for testing.

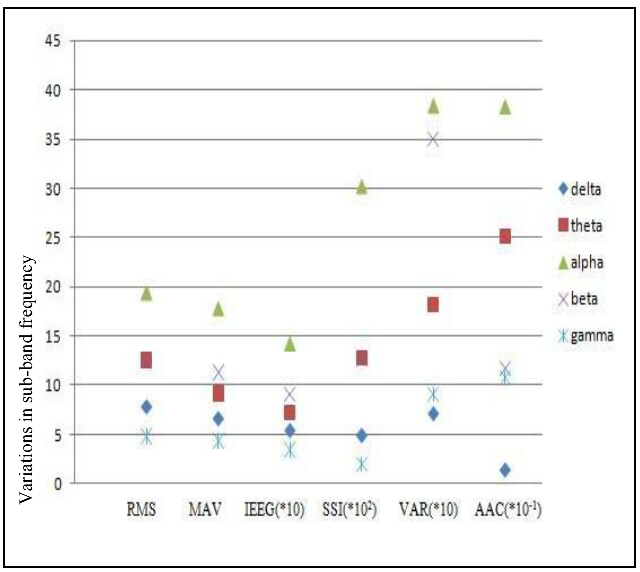

Figure 3: The features extracted using Db5 from a single channel (F3) during speech imagery ('a')

D. Deep Belief Network

DBN, probabilistic generative models have been composed of multiple layers of stochastic, latent variables. The variables or observable units that have not been externally available are called hidden units. The first two layers have undirected, symmetric connections between them and form an associative memory. The lower layers have received top down, directed connections from the top layers. The states of the units in the lowermost (visible) layer have represented the output data vector. A DBN has been built as a stack of its constituents, called Restricted Boltzmann Machines(RBMs). In this experiment, a DBN with 7 hidden layers has been built. DBN itself has consisted of *three* hidden layers with 1000 units (RBMs) per layer. The DBN has trained itself with a supervised learning rate of 0.002.

A typical RBM, a special type of Markov random field, has one layer of hidden units and one layer of visible or observable units. RBMs can be represented as bipartite graphs, where all I visible units have been connected to J hidden units, with no visible unit - visible unit connections, similarly there have not been any hidden unit - hidden unit connections.

Gaussian-Bernoulli RBMs can be used to convert real-valued stochastic variables to binary stochastic variables, which can then be further processed using the Bernoulli-Bernoulli RBMs. Taking the gradient of the log likelihood log p(v;θ), the RBM weights have been derived using the update rule as given in (7).

$$\Delta w_{ij} = E_{data}\left(v_i h_j\right) - E_{model}\left(v_i h_j\right) \quad (7)$$

where $E_{data}\left(v_i h_j\right)$ has been the expectation observed in the training set and $E_{model}\left(v_i h_j\right)$, the same expectation under the distribution defined by the model[12].

III. RESULTS AND DISCUSSION

From the generated data sets of each band, 80% of the samples have been used for training and the remaining 20% of samples for testing. This has been repeated for all the five vowels. The classification has been first carried out for the spoken EEG signal and the DBN has been tested for its performance. The classification accuracy of the DBN has varied between 69-95%. In the same way the experiments have been repeated for the imagined EEG signals. The overall classification accuracy for imagined EEG signals has been greater than 87.5% as shown in figure 4.

978-1-5090-4740-6/17/$31.00 ©2017 IEEE

Also, the vowels were more accurately classified in the theta and gamma bands. While analyzing the band-wise activity, theta band shows higher activation during the learning, memory, speech comprehension, imagery related activities and the gamma band shows higher activation during speech production related tasks. Accuracy of the vowel 'a' classified in each band for various electrode combinations is given in table 1.

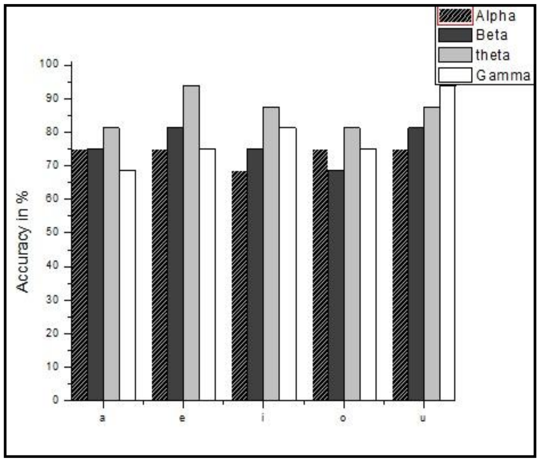

Figure 4: Classification accuracy of imagined vowels using DBN

TABLE 1: Accuracy of vowel 'a' identified by DBN for different electrode combinations in sub-bands alpha, beta, theta and gamma

Electrode Combinations	Frequency Bands			
	Theta	*Alpha*	*Beta*	*Gamma*
Frontal (F3, F7, F4, F8)	87.5%	50%	62.5%	75%
Temporal(T3, T4, T5, T6)	75%	50%	75%	87.5%
Left hemisphere (F3,F7,T3,T5)	100%	67%	75%	93%
Right hemisphere (F4,F8,T4,T6)	92%	75%	67%	87.5%
Frontal and Temporal	93.5%	87.5%	81.25%	93.5%

From table 1, the classification protocol employing features that have been derived from each sub-band frequency has shown that the theta and gamma band frequency features have been accurate when compared with alpha and beta sub-bands.

IV. CONCLUSION

With this work, the research in the domain of speech imagery through EEG signals has been expanded. The implemented principal features have been the conventional and modified energy based features. The features have been extracted using Daubechies wavelets and have been classified using Deep Belief Networks. This work has completely focused on band-wise analysis of the EEG signals. The model has shown a good performance record of above 87% in the theta and gamma bands. The work will be further extended to identify vowels from Consonant - Vowel (CV) and Consonant - Vowel - Consonant (CVC) syllables.

V. REFERENCES

[1] J. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767-791, 2002.

[2] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan. "Toward EEG sensing of imagined speech," *Human Computer Interactions and New Trends*, pp. 40-48, 2009.

[3] K. Brigham, B. V. K. Vijaya Kumar, "Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy," 4th *International Conference on Bioinformatics and Biomedical Engineering*, 2010.

[4] F. Guenther, J. S. Brumberg, E. J. Wright, A. N. Castanon, J. A. Tourville, M. Panko, et al., "A wireless brain-machine interface for real-time speech synthesis," *PLoS ONE* 4(12): e8218, vol. 4, no. 12, Dec. 2009.

[5] C. Sandhya, R. Anandha Sree, A. Kavitha, "Analysis of Speech Imagery using Consonant-Vowel Syllable Speech Pairs and Brain Connectivity Estimators", *Proceedings of 2nd International Conference on Biomedical Systems, Signals and Images*, IIT Chennai, February 2015.

[6] T. Jennifer Crinion, A. Matthew Lambon-Ralph, A. Elizabeth Warburton, J. David Howard and S. Richard Wise, "Temporal lobe regions engaged during normal speech comprehension", *Brain*, Vol. 126, no. 5, pp. 1193-1201, 2003.

[7] Manzoor Khazi, Atul Kumar, M J. Vidya, "Analysis of EEG Using 10:20 Electrode System", *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 1, issue 2, pp. 185-191, December 2012.

[8] A. Delorme, & S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics", *Journal of Neuroscience Methods, vol. 134, no.* 1, pp. 9-21, 2004.

[9] A. Mohamed, D. Yu, and L. Deng, "Investigation of full-sequence training of deep belief networks for speech recognition," in *Proc. Inter. speech*, Sept. 2010.

[10] S. Michahial, K. R. Ranjith, K. P. Hemath & K. A. Puneeth, "Hand rotate EEG signal feature extraction by second order Daubechies wavelet transform (DWT)." *Third International Conference on Computing Communication & Networking Technologies* (ICCCNT), Coimbatore (pp. 1-6), 2012.

[11] H. Mohammad Alomari, A. Emad Awada, Aya Samaha & Khaled Alkamha, "Wavelet-Based Feature Extraction for the Analysis of EEG Signals Associated with Imagined Fists and Feet Movements", *Computer and Information Science*, Vol. 7, No. 2, pp. 17-27, 2014.

[12] G. Hinton, "A practical guide to training restricted Boltzmann machines", Univ. Toronto, Tech. Rep. 2010-003, Aug.2010.

978-1-5090-4740-6/17/$31.00 ©2017 IEEE