Vowel Classification using Wavelet Decomposition during Speech Imagery

Basil M. Idrees
Department of Electronics Engineering
Z. H. College of Engineering & Technology
Aligarh Muslim University, Aligarh 202 002, UP, India
email: basilmidrees11@gmail.com

Omar Farooq
Department of Electronics Engineering
Z. H. College of Engineering & Technology
Aligarh Muslim University, Aligarh 202 002, UP, India
email: omarfarooq70@gmail.com

***Abstract*—Electroencephalography (EEG) has long been used for Brain computer interface (BCI). Recent researches have proved that EEG can be also used to classify data generated in speech imagery. This classification can further be utilized to develop speech prosthesis and synthetic telepathy systems. In this paper we wanted to check whether features extracted from beta, delta and theta rhythms of EEG can be used to classify the imagined English vowel sounds. A new approach is used to differentiate among the three classes of vowel sound /a/, /u/ and 'rest or no action' in pair-wise as well as 'combination of two sounds (tasks)' manner. Wavelet decomposition is performed to extract features in the 0-8 Hz and 16-32 Hz range. Energy sum and energy's waveform length of the approximate and detail coefficients are used as features. The algorithm is tested on 3 subjects and results showed that indeed the data from EEG rhythms can be used for classification. The pair-wise classification accuracy was found to be 65-82.5% which is a considerable improvement over the previous classification accuracies in the range of 56-82%, reported by DaSalla [4]. The 'combination of tasks' classification accuracy was found to be 81.25-98.75%.**

Keywords— Classification; Electroencephalogram (EEG); Energy Sum ; Energy's Waveform length; Imagined speech; Vowel; Wavelet Decomposition

I. INTRODUCTION

Thinking to oneself is a very common activity that every individual does. An equally common habit is to say words to oneself without actually saying it. These words or "imagined speech" are heard in one's own head. Capture and interpretation of imagined speech can open wide area of applications. A clear application for the people who have speaking disabilities such as advanced amyotrophic lateral sclerosis (ALS) [1], laryngectomy, paralysis, locked-in syndrome (LIS) etc. For those the interpretation of imagined speech can give "voice" to their thoughts. Other areas include a situation where visual and audio communications are undesirable such as war. In such situations the capture and decipher of this internal thought can give rise to some form of synthetic telepathy.

BCI based on EEG can provide a solution to these problems. BCIs aim to provide direct link between the neural activity of an individual and external devices.

EEG is a technique use to measure the electrical fields produced by brain activity. Since the work done by Dewan to transmit alphabets using EEG in 1967 there has been a growing interest in the area. Classification of Imagined speech using EEG is somewhat a newer venture.

D'Zmura [2] performed EEG based speech imagery experiments for classification but with subjects imagining syllable /ba/ or /ku/ without associated muscle movements. D'Zmura [2] classified the EEG signal by computing the Hilbert envelope of each electrode waveform and averaging signal envelopes over each electrode separately to form templates for each class. Matched filters were then used to classify the imagined syllable.

Jongin [3] conducted research regarding the phoneme representation in the brain and to find out whether EEG responses for each speech sound could be discriminated. Classification among the English vowels /a/, /i/ and /u/ was performed. Multivariate empirical mode decomposition and common spatial pattern was used to extract features. As proposed, the results confirmed that English vowel stimuli can be differentiated from the brain waves.

DaSalla [4] performed experiments where EEG data was recorded in subjects who imagined mouthing and vocalization of vowels "a" and "u". These vowels were specifically chosen because of different muscles involved in uttering them. Common Spatial Patterns method was used for feature extraction and subsequent classification was done using a nonlinear support vector machine (SVM). The classification was done between /a/ and control state, /u/ and control state and between /a/ and /u/ with classification accuracies ranging from 56-82%.

This paper uses data of DaSalla [4] to check whether data from EEG bands delta, theta and beta can be used to classify the imagined English vowel sounds /a/, /u/ and 'rest' to obtain a better classification accuracies.

II. DATA AQUISITION

The data is downloaded from a data base made publicly available by DaSalla [4] (www.brainliner.jp).

A. Subjects

EEG recordings of three healthy subjects viz S1, S2 and S3 out of whom 1 was a female and 2 were male, were obtained.

978-1-4673-9197-9/16/$31.00 ©2016 IEEE

Fluent English speaking participants were chosen. The experiment was conducted in accordance with the Declaration of Helsinki and from each participant prior consent was obtained.

B. Experimental paradigm

Digastricus muscle controls vowel /a/ pronunciation which is marked by mouth opening. And the pronunciation of vowel /u/ is marked by lip rounding and controlled by orbicularis oris muscle.

The subjects were coached beforehand and rehearsed with real movements to ensure correct task execution. The subjects were then asked to imagine a particular vowel when a visual cue appeared. Data recording was performed in the subsequent manner: A fixation crosses remained on the screen for 2-3 seconds and a audible beep was sounded. After this a visual cue appeared and remained on screen for 2 seconds entailing that the subject must imagine speaking the displayed English vowel for 2s. On the screen, the following was used to depict different tasks:

a. 'mouth opening' for /a/
b. 'lip rounding' for /u/
c. 'fixation cross' for 'no action' [4].

The experimental set up is shown in Fig 1.

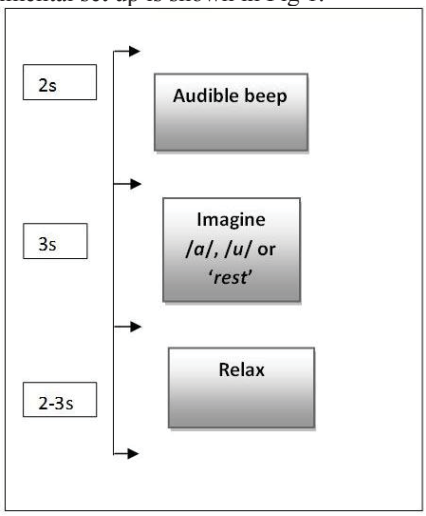

Fig. 1. The experimental setup as used by DaSalla [4]

The subjects were comfortably seated in a reclining armchair. On appearance of a visual cue, subjects were instructed to perform one of three tasks:

a. Vowel /a/- imagined mouth opening and imagined vocalization
b. Vowel /u/- imagined lip rounding and imagined vocalization
c. Control or rest-alert, no action [4]

C. Recording

Continuous EEG was recorded using 64+8 active Ag-AgCl electrodes with the help of a BioSemi ActiveTwo system. Accordingly, a BioSemi head cap was used to position electrodes on the scalp, in accordance with international 10-20 system. Originally sampled at 2048 Hz the data was down sampled at 256 Hz using software to reduce file size. Visual inspection of the signals was done during the recordings. The trials which showed movement artifacts were replaced by repeated trail in an extra session [4].

III. DATA PROCESSING

Out of the 64+8 available active electrodes, properly selecting the electrode which contains neuronal information about speech is important [5]. Speech musculature is located in the motor cortex. Consequently, only data from 4 electrodes positioned in the motor cortex region was chosen. MATLAB (The MathWorks. Inc) was used for data processing.

Out of 0-2 sec used to imagine the vowel, 0-500msec data was selected. This is justified as it takes approximately this much time to imagine vowels. Hence, 128 samples are present on each of the 4 channels. There are 50 trials for each task for each subject divided into 20 testing and 30 training set.

0-45Hz of EEG data essentially contains brain information. And EEG data is highly prone to be corrupted by various artifacts. So, the removal of these physiological and non- physiological artifacts is a must. These include low frequency baseline shift and electronic noise present in the signal. Other artifacts such as Electrocardiographic (ECG) artifacts (60-72 Hz), power line frequency (60Hz or 50 Hz), Electromyographic (EMG) artifacts, and lie above 45 Hz. [6].

Therefore, wavelet decomposition was performed to 3rd and 4th level which led to resolution of 0-16 Hz and 0-8 Hz respectively. Energies of Detail coefficients of 3rd level (16-32 Hz) and of approximate and detail coefficients of 4th level (0-8 Hz and 8-16 Hz resp.) were extracted. From those energies 11 features namely mean, variance, skewness, kurtosis, geometric mean, harmonic mean, inter-quartile range, energy sum, entropy, standard deviation and waveform length were calculated and tested for classification. Then this feature set is reduced taking only few features. The features selected are such that they extract maximum information from the EEG signal providing a decision boundary which is most significant for 2-class problem classification.

The data file available online consisted of the data from each subject for a 'combination of task' (for example, combination of /a/ with 'rest', /a/ with /u/ and so on). This was done to ease the pair-wise classification process as was done by DaSalla [4]. The classification can also be done among these combination of tasks (i.e. between 'combination of /a/ and /rest/' and 'combination of /u/ and /rest/ so on.)

For the pair-wise classification (i.e. /a/ form /u/ etc) the feature selected was 'energy sum' of approximation coefficients of 4th level wavelet decomposition of all the 4 channels. This corresponds to frequency range 0-8 Hz (delta & theta rhythm). So this 'energy sum' of Channel 1, 2, 3 and 4 formed the feature number 1, 2, 3 and 4 respectively. 'Energy sum' is defined by equation (1).

$$\text{Energy Sum} = \sum_{i=1}^{N} |C_i|^2 \quad (1)$$

where C_i is the approximation coefficient of 4th level wavelet decomposition of EEG signal and N number of coefficients. Fig 2 shows the graph of energy sum of channel 1 of Subject S1 for class /a/ and /re/. Fig 3 shows the scatter plot of 'energy sum' with channel 1 and channel 2 chosen as axes of Subject S1 for class /a/ and 'rest'. From these figures it is clear that these values are separable for both classes.

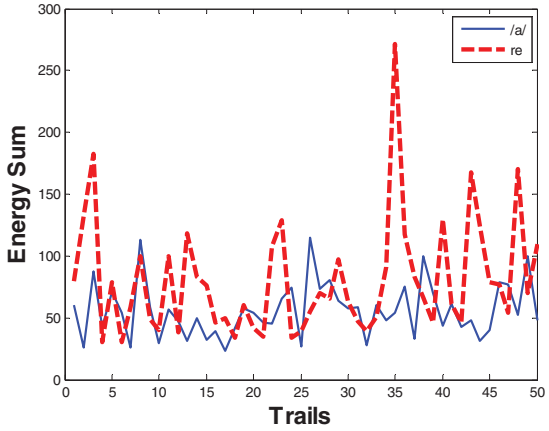

Fig. 2. 'Energy sum' of Channel 1 in S1. /a/ shown by regular line and /re/ by dashed line

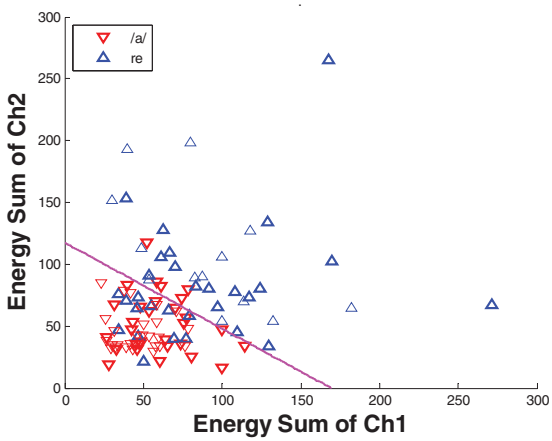

Fig. 3. Scatter plot of 'energy sum' along Channel 1 and Channel 2 in S1.

For the classification of the 'combination of tasks' the features selected were 'energy sum' and 'energy's waveform length' of detail coefficients of 3rd level wavelet decomposition. This corresponds to frequency band 16-32 Hz (Beta rhythm). As before, the 'energy sum' of Channel 1, 2, 3 and 4 is selected as feature number 1, 2, 3 and 4 respectively. Similarly the 'energy's waveform length' of Channel 1, 2, 3 and 4 is taken as the feature number 5, 6, 7 and 8 respectively. 'Energy's Waveform length' is defined by equation (2)

$$\text{Energy's Waveform Length} = \sum_{k=2}^{N} ||C_k|^2 - |C_{k-1}|^2| \qquad (2)$$

where C_k is the detail coefficient of 3rd level wavelet decomposition of EEG signal and N number of coefficients. Fig 4 shows the graph of energy's waveform length of channel 1 of Subject S3 for class /aiui/ and /uire/. Fig 5 shows the scatter plot of S3 along channel 1 and channel 2 for these classes. From these figures it is clear that these values are separable for both classes.

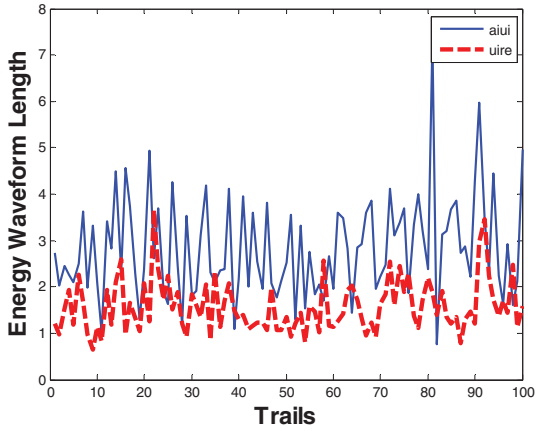

Fig. 4. 'Energy Waveform length' of Channel 1 in S3. /aiui/ shown by regular line and /uire/ by dashed line

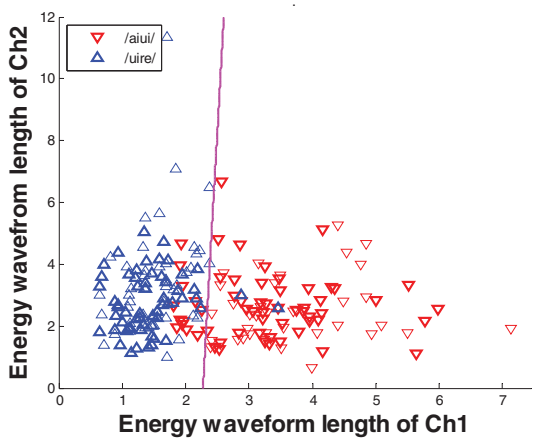

Fig. 5. Scatter plot of 'Energy waveform length' along Channel 1 and Channel 2 in S3.

IV. Classifiaction

The extracted features' matrix was divided into 2 groups. 60% of the data was used for training and 40% of the data was used for testing. This was done for all the 3 subjects S1, S2 and S3. Linear classifier was used for classification. The results were very encouraging and are discussed in the subsequent section.

V. RESULTS

The classifier output performance was measured in terms of 5 parameters defined below:

Accuracy: It is defined as the ratio of number of events correctly classified to the number of value classified.

$$\text{Accuracy} = \frac{\text{No. of correctly classified events}}{\text{No. of classifed events}} \quad (3)$$

Sensitivity: It is defined as the ratio of number of positive correctly classified value to the true positive value.

$$\text{Sensitivity} = \frac{\text{No. of positive correcly classified value}}{\text{No. of true positive value}} \quad (4)$$

Specificity: It is defined as the ratio of number of negative correctly classified value to the true negative value.

$$\text{Specificity} = \frac{\text{No. of negative correcly classified value}}{\text{No. of true negative value}} \quad (5)$$

Positive Predictive Value (PPV): It is defined as the ratio of number of positive correctly classified value to the positive classified value.

$$\text{PPV} = \frac{\text{No. of positive correctly classifed value}}{\text{No. of positive classifed value}} \quad (6)$$

Negative Predictive Value (NPV): It is defined as the ratio of number of negatively correctly classified value to the negative classified value.

$$\text{NPV} = \frac{\text{No. of negative correctly classifed value}}{\text{No. of negative classifed value}} \quad (7)$$

Table I shows the performance metric of linear classifier for pair-wise classification. Similarly, Table II shows the performance metric of linear classifier for the 'combination of task'.

TABLE I. PERFORMANCE (IN PERCENTAGE) OF THE PAIR-WISE CLASSIFICATION

		Accuracy	Sensitivity	Specificity	PPV	NPV
S1	'a'/ 'rest'	82.5	90	75	88.23	78.26
	'u'/ 'rest'	77.5	95	60	92.3	71.37
	'a'/ 'u'	75	90	60	85.71	69.23
S2	'a'/ 'rest'	72.5	65	80	69.56	76.47
	'u'/ 'rest'	75	75	75	75	75
	'a'/ 'u'	77.5	80	75	78.94	76.19
S3	'a'/ 'rest'	72.5	90	55	84.6	66.66
	'u'/ 'rest'	80	80	80	80	80
	'a'/ 'u'	65	70	60	66.66	63.63

TABLE II. PERFORMANCE (IN PERCENTAGE) OF 'COMBINAION OF TASK' CLASSIFICATION

		Accuracy	Sensitivity	Specificity	PPV	NPV
S1	'aire'/ 'uire'	91.25	87.5	95	94.59	88.37
	'uire'/ 'aiui'	92.5	95	90	90.47	94.73
	'aire'/ 'aiui'	81.25	85	77.5	79.06	83.78
S2	'aire'/ 'uire'	93.75	97.5	90	90.69	97.29
	'uire'/ 'aiui'	98.75	97.5	100	100	97.56
	'aire'/ 'aiui'	95	97.5	92.5	92.85	97.36
S3	'aire'/ 'uire'	92.5	97.5	87.5	88.63	97.22
	'uire'/ 'aiui'	98.75	100	97.5	97.56	100
	'aire'/ 'aiui'	98.75	100	97.5	97.56	100

i. 'aire' is the combination of epochs of /a/ and /rest/
ii. 'uire' is the combination of epochs of /u/ and /rest/
iii. 'aiui' is the combination of epochs of /a/ and /u/

VI. DISCUSSION

Classification accuracy of 81.25%-98.75% in the case of 'combination of task' classification and 65-82.5% in case of pair-wise classification was obtained. This is a sizeable improvement over the pair-wise classification results obtained

by DaSalla which had accuracy of 56-82%. Table III and Fig 4 further illustrates this.

Two statistical features extracted from wavelet coefficients of EEG data namely energy's waveform length and energy sum are used. The use of these statistical features obtained by time-frequency analysis elaborates that data from delta, theta and beta rhythms can be used to classify the vowel sounds. Only two features are used for classification which makes the algorithm simpler, faster. Linear classifier is used to differentiate between the classes which give added advantage of this algorithm. It gives high performance which again makes the algorithm simpler and efficient.

TABLE III. COMPARISON OF CLASSIFICATION ACCURACIES (IN PERCENTAGE) OF THE CURRENT WORK WITH THE WORK DONE BY DASALLA (UNDERLINED VALUES SHOW IMPROVEMENT)

		S1	S2	S3
'a'/'rest'	DaSalla [4]	79	71	67
	Current Work	82.5	72.5	72.5
'u'/'rest'	DaSalla [4]	82	72	80
	Current work	77.5	75	80
'a'/'u'	DaSalla [4]	72	60	56
	Current Work	75	77.5	65

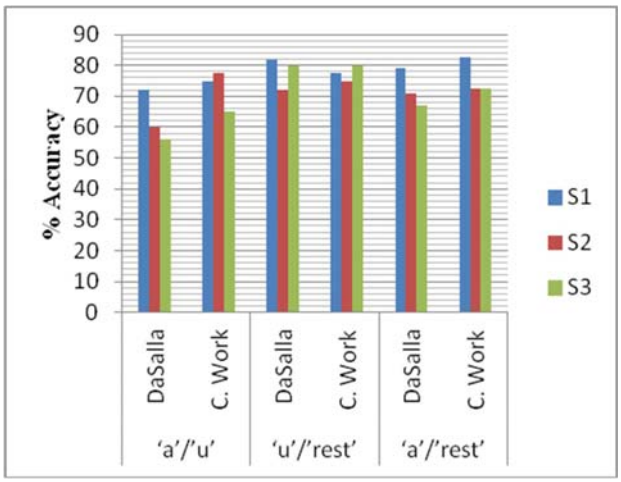

Fig. 6. Comparison of classifcation accuracies of Current work with DaSalla [4].

VII. CONCLUSION

Data obtained from DaSalla have been pre-processed and classified using statistical features obtained from various EEG frequency bands. Wavelet decomposition was applied to resolve the EEG signal from English vowel sounds and the wavelet coefficients were generated. 'energy waveform length' and 'energy sum' of these wavelet coefficients were found to be good features for two class problem classification. The classifier performance is evaluated on the basis of 5 parameters. The proposed approach works well on all the 3 subjects. The use of two features and a linear classifier makes the algorithm fast, efficient while maintaining computational simplicity. Good classification accuracy is obtained using this proposed algorithm.

Also, the algorithm has brought better results than the work reported in literature by DaSalla [4] in the pair-wise classification of all the three subjects. The accuracy obtained is 65-82.5 % which is a significant improvement over 56-82% as obtained by DaSalla [4] and that too using simpler and faster method. An accuracy of 82.25-98.75% was also obtained in case of classification of 'combination of tasks'.

In conclusion this paper proposes a new approach whose performance shows an improvement over the earlier reported in literature by DaSalla [4]. Consequently, the method can be further developed to be implemented in advanced BCIs for the synthetic telepathy systems and in developing prostheses for speech impaired patients.

REFERENCES

[1] K. Barrett, H. Brooks, S. Boitano and S. Barman – *Ganong's Review of Medical Physiology*: 23rd edition: Tata McGraw Hill, 2009.

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

[3] K. Jongin, S. K. Lee, and B. Lee, "EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition", Journal of neural engineering, 11.3: 036010, 2014.

[4] C. S. DaSalla, H. Kambara, M. Sato and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns", Neural Networks, Vol 22(9), pp. 1334-1339, 2009.

[5] X. Pie, D. Barbour, E.C. Leuthardt and G. Schalk, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans", Journal of neural engineering, 8.4: 046028, 2011.

[6] S. Iqbal, P.P M. Shanir, Y U Khan, O Farooq , "Time Domain Analysis of EEG to Classify Imagined Speech" , Proceedings of the Second International Conference on Computer and Communication Technologies, Advances in Intelligent Systems and Computing, Volume 380, pp 793-800 ,2015.