Indian Journal of Pure & Applied Physics
Vol. 59, March 2021, pp. 180-183

Transforming Imagined Thoughts into Speech Using a Covariance-Based Subset Selection Method

Prabhakar Agarwal & Sandeep Kumar*
Department of Electronics & Communication Engineering, National Institute of Technology, Delhi-110 040, India

Received 27 January 2021; accepted 17 February 2021

With the advancement of engineering solutions in the medical domain, the patient's life can become comfortable. This work recognizes the silent speech of three words. The decoding of silent speech can be useful for patients who are in a locked-in syndrome state. Moreover, it is also applicable to entertainment, cognitive biometrics, and brain-computer interfacing. Brain waves of these imagined words in the delta, theta, alpha, beta, gamma, and high gamma frequency bands are analysed. Covariance based connectivity features are extracted in each frequency band. The principal features which represent more than 95% of the variance are selected as a subset of the covariance connectivity matrix. This sub-set is tested on five classifiers. The maximum accuracy achieved is 76.4% in the theta band. Also, theta and high gamma band contain maximum information about imagined speech with average accuracies of 68.32% and 62.09% respectively.

Keywords: Brain-computer interfacing, Covariance, Electroencephalogram

1 Introduction

In 1929, Hans Berger[1] introduced a new device to the world called Electroencephalogram (EEG) to measure the electric potentials of the human brain. The brain currents vary, depending upon the nature of the conditions like epilepsy, movement, mental calculation, sleep, *etc.* and hence this paves the way to Brain-computer interfacing (BCI). BCI can be done with techniques like Electrocorticography (ECoG)[2], EEG, Electromyography (EMG)[3], and functional magnetic resonance imaging (fMRI)[4]. ECoG is invasive, fMRI is costly and in EMG, magnets are placed on the face of the subjects and hence making it uncomfortable. Conversely, EEG is non-invasive, cheap, portable, and easy to handle. So we focus on the EEG technique in this work. BCI has been used in many applications like movement imagery, entertainment, biometrics, Internet of things (IoT)[5-6], *etc*., under the umbrella of artificial intelligence. Yet another beneficial application of BCI is to recognize imagined speech/thoughts. This can act as a speech prosthesis for the patients having partial/complete paralysis but otherwise sensible with their cognitive features. The work of imagined speech using EEG is an extremely challenging task and is confined to decoding vowels, syllables, and short words primarily because of its low signal to noise ratio (SNR).

Early works of vowel imagery were done by DaSalla *et al.*[7] in 2009. They employed common spatial patterns (CSP)[7] to recognize tasks /a/, /u/ and rest interval. The authors classified with support vector machine (SVM) and achieved an accuracy between 68% to 78%. On the same imagined vowels, Prabhakar *et al.*[8] used statistical features with the Random Forest (RF) classifier and got a maximum accuracy of 89%. K. Brigham and Vijaya Kumar[9] extracted Autoregressive (AR) coefficients from imagined syllables /ba/ and /ku/. They utilized the 3-nearest neighbour classification algorithm and got an average accuracy of 61%. Balaji *et al.*[10], classified a mixture of Hindi (haan/na) and English (yes/no) words and got an average classification accuracy of 75.38%. Nguyen *et al.*[11], decoded imagined vowels, short and long words. They used covariance matrix based features[12-13] and attained a maximum accuracy of 70%. In 2018, Qureshi *et al.*[14] recognized 5 imagined words 'go', 'back', 'left', 'right', and 'stop'. They used covariance-based features and an Extreme learning machine as a classifier. The maximum classification accuracy was 40.3%. In the current work, three imagined words 'sos', 'stop', and 'medicine' are recognized. These words are medium to long size in length. Covariance-based features are extracted and five classifiers are used to decode these words. The rest of the manuscript is drafted as follows. Materials and methods used in the

*Corresponding author: (E-mail: sandeep@nitdelhi.ac.in)

experiment are described in Section 2. Section 3 describes the results and discussion followed by the conclusion in Section 4.

2 Material and Methods

2.1 Experimental setup and procedure

The subjects were seated on a chair comfortably and a fixation cross was shown for 5 seconds on the computer screen so that they can focus on the experiment. Then a word was displayed on the screen for 2 seconds. The subject imagines it without producing any vocal sound or movement while keeping their mouth closed. This is followed by a 3-second rest interval. Each word is repeated for 3 trials per subject. A total of thirteen subjects took part in the experiment, hence the total number of trials for each word was 39. Three medium and long-sized English words namely 'sos, 'stop' & 'medicine' were used in this study. The subjects were right-handed, mentally sound, and submitted written consent for the experiment. The experiment consists of recording brain potentials of imagined speech and rest intervals from the scalp by 32 channel, water-based electrode EEG device MOBITA® at the National Institute of Technology, Silchar. The recordings performed were according to the 10-20 international system[15] with a sampling frequency (fs) of 250 Hz. The analysis was done in MATLAB software offline.

2.2 Pre-processing

The channels associated with speech and language areas of the brain *i.e.* Broca's (F7, F3, FC5) and Wernicke's (CP5, P3, T3, T5, C3) region[14] were selected. Then from each sample of the data, the mean of the whole EEG signal is subtracted to nullify the effect of the common signal. This is known as the common average referencing[16] and is calculated by Eq. 1.

$$x_i^c = x_i - \sum_{j=1}^{N} x_j / N \quad \text{... (1)}$$

Here x_i^c is the common average referenced signal, x_i is the signal of each channel and N is the total number of channels. Now, to reduce the power-line interference, a butter-worth notch filter of order 2 is applied at 50 Hz and its harmonics (100 Hz). Discrete wavelet transform[17] (DWT) is applied to decompose and reconstruct the signal into various EEG frequency bands. We have used 'dmeyer' wavelet up to five levels of decomposition (i = 1,2,3,4,5). Each frequency band is responsible for different activities of the brain[18] as mentioned below.

a) Delta (0.5 to 4 Hz): This frequency range has the highest amplitude and occurs during deep sleep.
b) Theta (4 to 8 Hz): This band occurs due to emotional stress, meditation, and inspiration.
c) Alpha (8 to 13 Hz): These frequencies arise when eyes are closed, during mental activity and stress.
d) Beta (13 to 30 Hz): This frequency range is generated when the mind is doing mental activity and some focussed task.
e) Gamma (>30 Hz): This frequency band is concerned with cognitive & motor functions. This band can be further classified as high gamma, the frequency of which lies approximately above 80 Hz.

Now, in terms of approximation (Ai) and detailed (Di) coefficients of DWT, the frequency bands correspond to delta (A5), theta (D5), alpha (D4), beta (D3), gamma (D2), and high gamma (D1). The various frequency bands are shown in Fig. 1. The amplitude of all the waves shown is in microvolts. The duration of

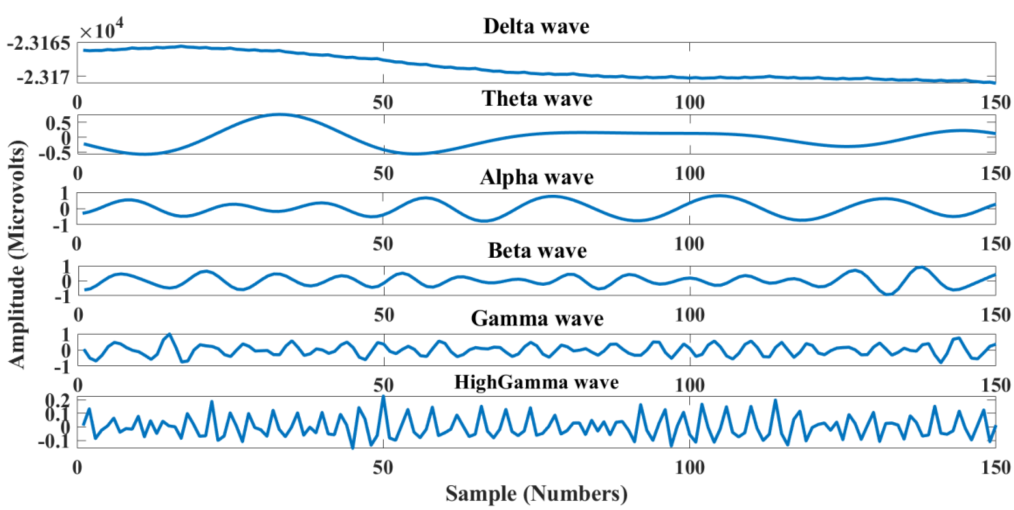

Fig. 1 — EEG frequency bands of a sample subject obtained after DWT

each imagined word is 2 seconds. For analysis, the first 1.5 seconds of all the words which contain 375 samples each, are taken and the rest is discarded.

2.3 Feature Extraction and Classification

Covariance as features for silent speech classification has been recently used by many researchers[14,19]. For each imagined word of dimension 375*8, we calculate the covariance matrix of dimension 8*8 according to Eq. 2.

$$cov\ (x) = \frac{1}{N-1}\sum_{i=1}^{N}|x_i - \mu_i|^2 \qquad \dots (2)$$

In the above equation, x is any random signal, N is the number of samples and μ is the mean of x. The Eigen values of the resultant covariance matrix are calculated and sorted in descending order. Eigen values of a covariance matrix are the variances in the independent coordinate frame. Only the top few Eigen values are sufficient to represent the signal as they contain more than 95% of the variance. This is a subset of the co-variance connectivity matrix. We have chosen the top 3 Eigen values in this work. This also reduces the dimension and additional computational cost. These features are extracted in the delta, theta, alpha, beta, gamma, and high gamma bands separately. Five classifiers are used in this work namely Decision Tree (DT), Naïve Bayes (NB), SVM with Gaussian kernel, K nearest neighbour (KNN), and RF. 83% and 17% of the data is used in training and testing respectively with 5 times cross-validation.

Table 1 — Accuracy of the words sos/stop

	Decision Tree	Naïve Bayes	SVM	KNN	Random Forest
Delta (0.5-4)	58.10	61.10	55.00	48.88	51.94
Theta (4-8)	76.40	58.30	57.20	76.38	73.34
Alpha (8-13)	45.80	65.20	51.90	45.84	36.66
Beta (13-30)	55.00	56.00	58.10	58.06	57.03
Gamma (30- 70)	36.70	45.80	45.80	42.78	45.83
High Gamma (70-124)	64.20	62.10	58.10	58.00	57.00

Table 2 — Accuracy of the words sos/medicine

	Decision Tree	Naïve Bayes	SVM	KNN	Random Forest
Delta (0.5-4)	61.10	55.00	58.10	58.06	53.30
Theta (4-8)	65.30	53.40	56.40	51.94	53.20
Alpha (8-13)	48.90	48.90	51.90	55.00	58.06
Beta (13-30)	54.00	51.90	61.10	61.12	61.12
Gamma (30-70)	48.90	55.00	58.10	51.94	51.42
High Gamma (70-124)	64.20	59.10	55.80	64.17	67.22

Table 3 — Accuracy of the words stop/medicine

	Decision Tree	Naïve Bayes	SVM	KNN	Random Forest
Delta (0.5-4)	51.90	58.10	64.20	61.12	61.12
Theta (4-8)	39.70	53.90	51.90	52.70	42.78
Alpha (8-13)	64.20	45.80	55.00	61.30	58.06
Beta (13-30)	61.10	51.90	52.40	66.90	62.00
Gamma (30- 70)	45.80	64.20	55.00	48.80	48.00
High Gamma (70-124)	64.10	42.80	42.80	64.16	67.22

3 Results and Discussion

Classification accuracies in delta, theta, alpha, beta, gamma, and high gamma frequency bands were calculated by five classifiers. Table 1 shows the pairwise accuracies of the words ‘sos’ and ‘stop’. The two highest average accuracies are present in theta and high gamma band with average values of 68.32% and 59.88% respectively. DT and KNN classifiers gave the best results with maximum values of 76.4% and 76.38% respectively. The results of the words ‘sos’ and ‘medicine’ are shown in Table 2. The high gamma band contains the maximum average accuracy of 62.09%. DT and RF classifiers gave good results with maximum values of 65.3% and 67.22% respectively. In Table 3, the accuracies of ‘stop’ and ‘medicine’ imagined words are given. Delta band shows the highest average accuracy of 59.28%. RF classifier gave the maximum accuracy of 67.22% in the high gamma band followed KNN giving 66.90% accuracy in the beta band. RF classifier is an ensemble of many decision trees in which each DT predicts by selecting some random samples and the final classification result is based upon averaging the result of each tree. This could be the reason for the high accuracy of the RF classifier. Accuracies of different pairs of words can differ depending upon the manner of articulation and length of words.

In Fig. 2 the average accuracies of all the classifiers in different frequency bands are shown. The accuracies

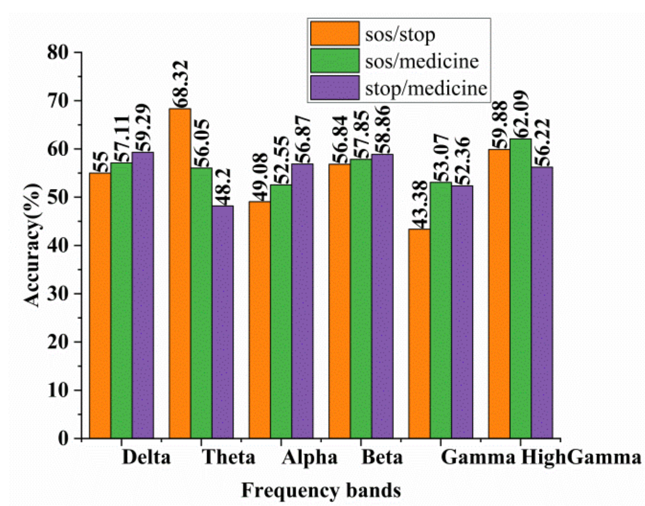

Fig. 2 — Average classification accuracy of the classifiers

are computed for the three pairwise combinations of words separately. We can see from the results that the highest average accuracy is in the theta band followed by a high gamma band. Also, the word pair 'sos' and ‘stop’ are classified with the highest accuracy.

4 Conclusion

We can draw the following conclusions from the results. First, the accuracy of pair-wise medium and long word classification is above chance (50%) for most of the classifiers. With non-invasive devices such as EEG whose SNR is poor, these results are encouraging. Also, we have performed subject independent BCI as the training and testing data are taken from different subjects. Second, silent speech content can be observed mainly in theta and high gamma bands with a maximum accuracy of 76.4%. So, this work complements the previous findings for vowels, syllables, words and provides additional insight with a comparison among 5 classifiers. Third, RF, DT, and KNN classifiers gave good results. We have performed the classification from a very limited data set with linguistic variability. This means the subjects who participated in the experiment were from different geographical locations and were well versed with their vernaculars though the experiment was performed in the English language. Thus, this is a practical case for patients with locked-in syndrome. To conclude, a subject independent BCI including arterial variability for recognition of imagined speech has been proposed using a non-invasive modality.

References

1 La Vaque T J, *J Neurotherapy*, 3 (1999) 1.
2 Martin S, Brunner P, Iturrate I, Millán J del R, *Sci Rep*, 6 (2016) 25803.
3 Wand M & Schultz T, Speaker-Adaptive Speech Recognition Based on Surface Electromyography in Biomedical Engineering Systems and Technologies: Edited by Fred A, Filipe J & Gamboa H (Springer, Berlin), (2010) 271.
4 Formisano E, De Martino F, Bonte M & Goebel R, *Science*, 322 (2008) 970.
5 Pineda J A, Allison B Z &VankovA, *IEEE Trans Rehabilit Eng*, 8 (2000) 219.
6 Kaur B, Singh D & Roy P P, *Multimed Tools Appl*, 76 (2017) 25581.
7 DaSalla C S, Kambara H, Sato M & Koike Y, *Neural Networks*, 22 (2009)1334.
8 Agarwal P, Kale R K, Kumar M & Kumar S, Silent speech classification based upon various feature extraction methods' in 7th *Int Conf Signal Process Integr Networks*, IEEE, Noida, India, (2020) 16.
9 Brigham K & Kumar B V K V, Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy, 4th *Int Conf Bioinform Biomed Eng*, (IEEE, Chengdu, China), (2010) 1.
10 Balaji A, Haldar A & Patil K, EEG-based classification of bilingual unspoken speech using ANN' in 39th *Ann Int Conf IEEE Eng Med Biol Soc*, (IEEE, Seogwipo), (2017) 1022.
11 Nguyen C H, Karavas G K & Artemiadis P, *J Neural Eng*, 15 (2017) 016002.
12 Tu L W, An Introduction to Manifolds (Springer, New York), 2010.
13 Nguyen C H & Artemiadis P, *Neurocomputing*, 275 (2018) 1871.
14 Qureshi M N I, Min B & Park H J, *IEEE Trans Biomed Eng*, 65 (2018) 2168.
15 Klem G H, Lüders H O, Jasper H H & Elger C, *Electroencephal Clinical Neurophys Supplement*, 52 (1958) 3.
16 Torres-García A A, Reyes-García C A, Villaseñor-Pineda L & García-Aguilar G, *Expert Syst Appl*, 59 (2016) 1.
17 Daubechies I, Ten Lectures on Wavelets (Society for Industrial and Applied Mathematics, USA), 1992.
18 Siuly S, Li Y & Zhang Y, EEG Signal Analysis and Classification: Techniques and Applications, (Springer International Publishing), 1stEdn, 2017.
19 Mohanchandra K & Saha S, *Augmented Human Res*, 1 (2016) 3.