Vowel Sound Imagery Decoding by a Capsule Network for the Design of an Automatic Mind-Driven Type-Writer

Sayantani Ghosh[1], Mousumi Laha[1], Amit Konar[1], Pratyusha Rakshit[1], Atulya K. Nagar[2]
Electronics & Telecommunication Engineering Department[1], Jadavpur University, India[1]
Department of Math and Computer Science, Liverpool Hope University, Liverpool, UK[2]
sayantani.sonrisa25@gmail.com[1], lahamou@gmail.com[1], konaramit@yahoo.co.in[1], pratyushar1@gmail.com[1], nagara@hope.ac.uk[2]

***Abstract*— This paper intends to develop a novel methodology for modeling a mind-controlled type-writer system to cater the needs of individuals suffering from various communication related disorders. This objective is fulfilled by first capturing EEG signals from ten subjects involved in mental utterance of seven vowel sounds. The eLORETA analysis of the acquired signals confirms the involvement of occipital, parietal and pre-frontal lobes for this cognitive activity. The procured signals undergo the process of feature extraction after eradication of artifacts and are transferred to a novel capsule network module for categorization of seven class labels. Performance analysis undertaken confirms the superlative behavior of the proposed classifier with respect to other standard ones. Moreover, statistical evaluation also assures the superior performance of the proposed classifier model. A coding scheme has also been proposed to signify the consonants by the coalescence of two vowel sounds segregated by a space. Thus, the proposed methodology can be effectively utilized as a mind controlled type writing system to serve the needs of disabled individuals. Additionally, this technique can also be used in certain military scenarios that demand non-verbal communication as a secure option and in various BCI based gaming applications.**

***Keywords*— mind controlled type writer, vowel sound imagery, EEG, eLORETA, capsule network.**

I. Introduction

The most common form of communication amongst human beings involves the impartation of information in the form of speech, writing and gestures. However, the usage of these communication modalities becomes quite inaccessible to individuals suffering from speech production and/or muscle related impairments such as Amyotrophic Lateral Sclerosis (ALS), Progressive Bulbar Palsy (PBP), Locked-in-Syndrome (LIS) etc. Thus, there is a grievous need for implementation of any mechanism that will aid in providing a "voice to the thoughts" of such disabled people. The present study aspires to attain the above stated objective by designing a Mind Controlled Type Writer (MCTW). The main characteristic of such a system includes the ability to abstract vowel sounds from the mental imagery of subjects and transform them into written/printed form of words or sentences.

Existing Literature on MCTW primarily focuses on the classification of mentally uttered vowel sounds using EEG based brain signal analysis [1-5]. However, these research studies provide only a preliminary instance of MCTW and fail to throw any light on the complete design and development of such a system. The present work is an endeavor to nullify this void.

The first step of our research involves the investigation of the active brain lobes from the acquired EEG signals of subjects who mentally utter seven vowel sounds pertaining to Bengali language semantics (A, AA, AeA, EE, O, Ae, UU) using eLORETA software. Next, the signals are pre-processed to eradicate the effects of various artifacts. Then, the pre-processed signals are fed to a feature extraction module to abstract the important features of vowel sound imageries. The extracted features are transferred to a classifier that can accurately categorize the vowel sounds into seven distinct classes. In order to involve the consonants in printed form, a code book has been developed where each consonant is depicted by the collocation of two vowel sounds separated by a space. For example, the letter B is represented as A_A, the letter C is denoted as A_EE and suchlike. So, the combination of two vowel sounds will be able to represent ${}^{7}C_{2} = 21$ consonants. Thus, this encoding scheme will be of immense help to the diseased subjects for communicating words or sentences constituting of both vowels and consonants.

The classifier required for the aforesaid problem must be able to categorize the seven vowel sound imageries. Although any conventional supervised learning based classifier could have been utilized for the current application, we opt for a capsule network based classifier for its precision in classifying features and low computational complexity [6-7]. Our proposed classifier is based on the original model introduced by S. Sabour [6] with two major modifications. First, we investigate the use of a new activation function, Exponential Linear Sigmoid SquasHing (ELiSH) [8] instead of the classical Rectified Liner Unit (ReLU), as a strategy to enhance the accuracy of the classifier module. The prime motivation behind the choice of this function is that it is quite efficient in tackling the vanishing gradient issue and also allows better information propagation during training phase. The next contribution towards designing the classifier module involves the utilization of Sigsoftmax [9] function in the routing algorithm rather than the primitive Softmax. This function has the capability to enhance the likelihood of many higher level probabilities amongst features in neighboring capsule layers during the competition of a primary capsule with each of the secondary capsules. Moreover, it can efficiently combat the Softmax bottleneck [9] issue as well as possess high representation capacity without compromising the computational cost.

Experimental results obtained infer that the performance of the proposed classifier surpasses other primitive classifier techniques. Additionally, statistical analysis undertaken also assures the superlative results provided by the proposed classifier over other standard ones. Besides, the proposed methodology has also shown supercilious performance when tested as a MCTW for both vowels and consonants by utilizing the classification of mental vowel sound imageries. Thus, this methodology can act as an efficacious tool to convey the messages of disabled individuals. Moreover, the proposed technique can be utilized in military applications where non-verbal communication can provide a better and secure possibility under grave scenarios like war. Additionally, this system can play a vital role in numerous gaming applications that will aid in exhilarating the player's experience.

The remaining segments of the paper have been compiled as follows. The abridgment of the entire system module has been described in Section II. Section III illustrates the architecture of the proposed classifier. Section IV depicts the experimental protocols undertaken. Section V discusses the performance analysis and statistical validation of the proposed model. The inference has been portrayed in Section VI.

II. SYSTEM OVERVIEW

This section introduces a summarized description of all principles and methodologies adopted to classify vowel sound imageries of subjects from their brain response. Fig. 1 provides a detailed description of the overall system. The EEG signals are captured from the scalp of the subjects using Ag/Agcl$_2$ electrodes when they are involved in imagination of vowel sounds. The visual stimuli begin with a fixation cross of 3 seconds duration that aids a subject to concentrate his/her attention towards the instructed task. The next window consists of the presentation of vowel sound to be imagined which has duration of 2 seconds. This window segment is succeeded by a blank window of 5 seconds duration where the subject mentally utters the vowel sound presented by the previous frame. The procured EEG signals are then fed to the exact Low Resolution Topographic Analysis (eLORETA) software, to identify the active brain regions responsible for this work. Next, to eradicate the effects of numerous artifacts from the raw EEG data, the signals are filtered using an Elliptical Band pass filter of order 10. The filtered signals are then transferred to an ICA (Independent Component Analysis) [21] module to eradicate the artifacts concomitant in the pass band of the filter. The artifact free signals are transferred to a spectrogram analyzer module to extract time-frequency domain features using STFT (Short Time Fourier Transform) algorithm. The acquired feature space is reduced by selecting features that have frequency components within alpha and theta bands. These two bands have been chosen as they portray strong alliance with vowel sound imagination. The reduced spectrogram features are transferred to a novel capsule network to classify seven vowel sounds imageries.

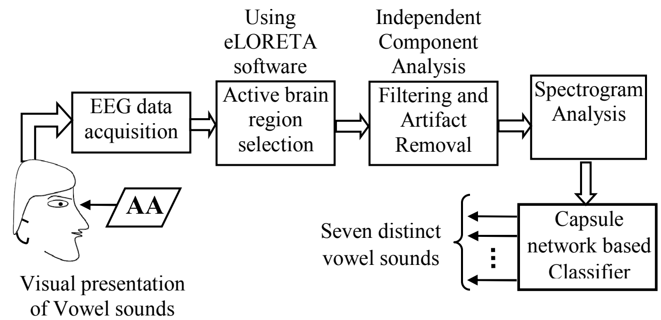

Fig. 1 Block diagram of the complete system

III. CLASSIFIER DESIGN

The architecture of the proposed classifier is illustrated in Fig. 2. The function of each sequential layer of the classifier model has been described below.

A. 2D Convolutional layer with ELiSH activation

This is the first layer of the neural network that performs convolution operation on $X_g \in \Re^{M\times M}$ extracted EEG feature map where $M \times M$ denoted the size of each feature map. For a stride value of A and $B \in \Re^{K\times K}$ number of convolution filters, where $K \times K$ denotes the kernel size, this computation is expressed as

$$y_h = f_a((x_g * w_{gh}) + b_h) \tag{1}$$

where, $y_h \in \Re^{P\times P\times B}$ indicates the h^{th} output feature map each having $P \times P$ dimension and $P = \frac{M-K+1}{A}$. w_{gh}, b_h and $f_a(\cdot)$ represent the weights of the convolution filter, bias term and activation function respectively.

We utilize a new activation function Exponential Linear Sigmoid Squashing (ELiSH) [8] to introduce the non-linearity operation in this layer which is defined as follows.

$$f_a(x) = \begin{cases} \left(\dfrac{x}{1+e^{-x}}\right), x \ge 0 \\ \left(\dfrac{e^x - 1}{1+e^{-x}}\right), x < 0 \end{cases} \tag{2}$$

This function is a combination of ELU [10] and sigmoid [11] in the negative part while the positive part involves the Swish function [12]. The sigmoid part of this function helps in better propagation of information during training instance while the linear part aids in tackling the vanishing gradient issue. Moreover, this function has showed marvelous results when applied on ImageNet dataset [8]. These promising characteristics have motivated us to utilize this activation function for the current application.

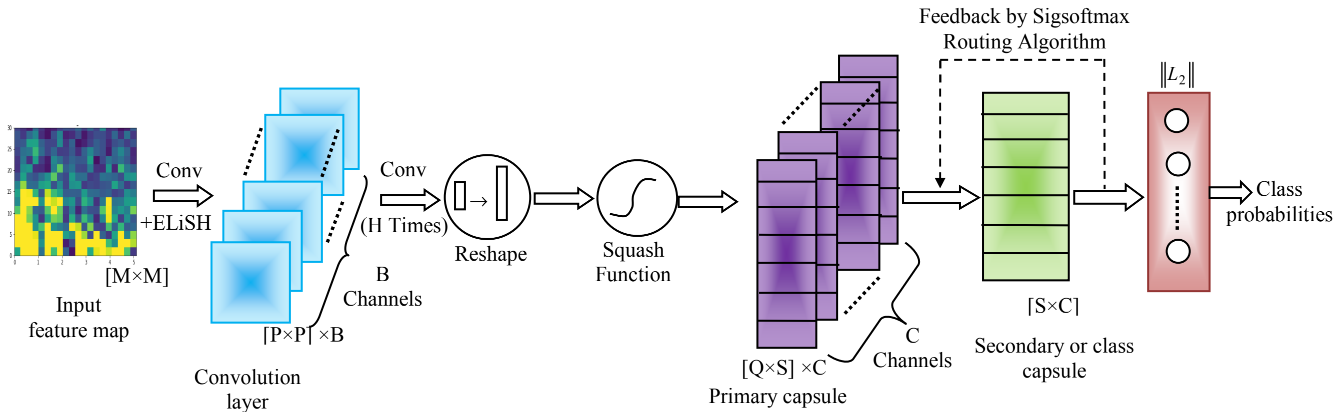

Fig. 2 Architecture of proposed Capsule Network

B. *Primary capsules*

In this layer a multi-convolution operation is performed H times on the output feature map to learn higher level features. Then the output of this multi-convolution operation is reshaped and squashed to form a set of feature maps referred to as Primary capsules. For C number of classes these series of operation is expressed by equation (3)

$$u_i = f_s(y_h * w_{hi} + b_i) \tag{3}$$

Where, $u_i \in \Re^{Q \times S \times C}$ denotes the i^{th} output feature map or the primary capsule unit within which each capsule has a dimension of $Q \times S$ and $f_s(\cdot)$ indicates the non-linear squashing function.

Since, each unit of the reshaped output of the multi-convolution operation must depict the probability of an entity's presence, its length is limited within [0, 1] using a non-linear squashing function denoted by equation (4)

$$Squash(r_i) = \frac{\| r_i \|^2}{1 + \| r_i \|^2} \cdot \frac{r_i}{\| r_i \|} \tag{4}$$

Where, $r_i \in \Re^{U \times V}$ represents the reshaped output of size $U \times V$.

C. *Secondary or Class capsules*

The secondary or class capsules are generated using the concept of dynamic routing by agreement. This procedure is used instead of a pooling operation that usually eliminates the local information, thus aiding in enhancement of network's robustness and classification accuracy.

In the first step, between two neighboring layers L and $L+1$ (i.e. between a primary capsule layer and a secondary capsule layer), a prediction vector is evaluated that denotes the strength of contribution of primary capsule i to the secondary capsule j using equation (5)

$$\hat{u}_{i|j} = w_{ij} u_i \tag{5}$$

Where, $\hat{u}_{i|j} \in \Re^{Q \times T \times C}$ represents the prediction vector where each vector has a size of $Q \times T$ and $w_{ij} \in \Re^{S \times T}$ is the weight matrix having size of $S \times T$.

The prediction vectors computed by the primary capsule layer are utilized to generate an input vector q_j of the j^{th} capsule in layer $L+1$ by the following equation

$$q_j = \sum_{i=1}^{N} c_{p_{ij}} \hat{u}_{i|j} \tag{6}$$

Where, $c_{p_{ij}}$ represents the coupling coefficient whose function is to provide assurance regarding the prediction of association of primary capsule i in layer L and with secondary capsule j in layer $L+1$. This coefficient is evaluated in the classical Capsnet [6] using Softmax function as shown below.

$$c_{p_{ij}} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{7}$$

where, b_{ij} indicates the log prior probability that a primary capsule i is coupled to a secondary capsule j and is updated by the routing algorithm. In our proposed model, we utilize a new function Sigsoftmax [9] instead of the classical Softmax to compute the coupling coefficient as shown by equation (8)

$$c_{p_{ij}} = \frac{\exp(b_{ij})\, \sigma(b_{ij})}{\sum_k \exp(b_{ik})\, \sigma(b_{ik})} \tag{8}$$

Where, $\sigma(\cdot)$ denotes a sigmoid function.

The use of classical Softmax diminishes the likelihood of various high level probabilities of feature association amongst the adjacent capsule layers. In other words, Softmax function evaluates almost similar association probabilities of primary capsules with that of class capsules [13]. Thus, this characteristic adversely affects the classification accuracy. Additionally, it possesses a poor representation capability leading to a condition called Softmax bottleneck [9]. On the other hand, the computation of Sigsoftmax function provides

Algorithm 1: Sigsoftmax Routing Procedure

1: Input to Routing Mechanism ($\hat{u}_{i|j}, n, L$)
2: for all capsules i in layer L and capsules j in layer $(L+1)$:
$b_{ij} \leftarrow 0$
3: **for** n iterations **do**
4: for all capsules i in layer L: $c_{p_i} \leftarrow Sigsoft\max(b_i)$
5: for all capsules j inlayer $(L+1)$: $q_j \leftarrow \sum_{i=1}^{N} c_{p_{ij}} \hat{u}_{i|j}$
6: for all capsules j in layer $(L+1)$: $o_j \leftarrow Squash(q_j)$
7: for all capsules i in layer L and capsules j in layer
$(L+1) : b_{ij} \leftarrow b_{ij} + \hat{u}_{i|j} \cdot o_j$
return o_j.

greater differences between the probabilities associated with class capsule for every primary capsule. Thus, this function provides an optimal segregation amongst capsules in neighboring layers. Moreover, this function has a better representation capacity than Softmax without the use of any auxiliary parameter or increase in computational cost. Hence, we use the Sigsoftmax based routing mechanism to reward the accordance between the output of the class capsules o_j and prediction vector $\hat{u}_{i|j}$. The detail of this procedure is illustrated by **Algorithm 1.**

Finally, a $\| L_2 \|$ or Euclidean norm of the output vector is evaluated to specify a distinct class label.

D. Loss Function

The loss function aims to evaluate the precision of prediction made by the network. It is defined as

$$\Im_k = Z_k \max(0, e^+ - \| o_k \|)^2 + \beta(1 - Z_k) \max(0, \| o_k \| - e^-)^2 \quad (9)$$

where, Z_k is an indicator variable that helps to denote the absence/ presence of an active class label, e^+ and e^- are the upper and lower boundary parameters respectively and β is the regularization parameter.

IV. Experiments And Results

This section includes the following experiments to obtain the analytical results for vowel sound classification problem from the brain response of subjects. Experiment 1 deals with selection of active brain regions for seven vowel sounds using eLORETA software. Experiment 2 performs filtering and artifact removal. Experiment 3 provides the spectrogram analysis of the artifact free EEG signals. Finally the output of this analysis is transferred to the classifier module for classification of seven class labels.

A. Experimental framework.

The experiment has been executed in Artificial Intelligence laboratory of Jadavpur University, Kolkata, India. The experimental setup has been shown in Fig. 3. A 21 channel EEG system, manufactured by Nihon-Kohden has been used to capture the electrical activity of the brain [14]. Out of ten healthy volunteers, six female and four male have volunteered for the said experiment. The experiment has been performed with 21 electrodes which include temporal (T_1, T_2, T_3, T_4, T_5 and T_6), frontal (F_3, F_4, F_7 and F_8), motor cortex (C_3, C_4), parietal (P_3, P_4), occipital (O_1, O_2) and prefrontal (Fp_1 and Fp_2,) electrodes. Each subject is requested to sit in a comfortable resting position to eschew any probable contamination of muscle artifacts [15].

B. Stimuli preparation and presentation

The experiment comprises 5 sessions, where each session includes 5 trials. In each trial, the subject receives visual stimulus of 5 seconds duration containing instruction to mentally utter the vowel sound presented on the computer screen. A sufficient time-gap of 15 seconds is mainained between 2 consecutive presentation to avoid the residual effect of previous stimulus. Thus, for ten subjects, 10 × 7 vowel sounds × 5 session/stimulus × 5 trials/sessions = 1750 training instances are generated for the said purpose. Fig.4 illustrates a single sequencial trial of stimulus portrayed to each subject.

C. Experiment 1: Source Localisation using eLORETA Software

In this experiment, the highly enhanced brain lobes are chosen using exact low resolution brain electromagnetic topographic analysis (eLORETA) technique [16], [17]. eLORETA is a linear inverse solution method which helps to analyze the distribution of electrical activity within the cortical regions of the brain [18]. In this experiment, EEG data has been collected from the scalp of the subject for 5 seconds (= 5,000 milliseconds) duration when he/she is involved in mental utterance of vowel sounds. eLORETA software segments this duration into a fixed 640 time frames. Thus topographic solutions are provided during a period of 7.81 milliseconds for each time frame. Fig. 5 illustrates the eLORETA solution of vowel sound imagery. It is evident from the figure that the occipital region is highly active for the first four time-frames, representing the visual perception of vowel sound stimuli for approximately 32 milliseconds duration. The parietal and pre-frontal regions are highly active bilaterally, for the remaining time-frames, which signify the involvement of vowel sound imagination for the present experiment.

D. Experiment 2: Pre-processing and Eradication of Artifacts

This experiment aims at removing physiological artifacts such as eye blinking, respiration, muscles artifacts etc [19]. An Infinite Impulse Response (IIR) filter has been selected for the present study instead of a Finite Impulse Response (FIR) design as it utilizes a meagre number of filter coefficients for a particular filter order. This methodology is employed by utilizing an Elliptical Band pass filter of order 10 with pass band frequency in the range of 3 to 13 Hz. The Elliptical Band pass filter has been chosen for its better roll-off and stop band attenuation as compared to other standard filters [20].

Next, in this experiment the filtered EEG signals are further evaluated using Independent Component Analysis (ICA) [21] to restore 19 independent components of the 19

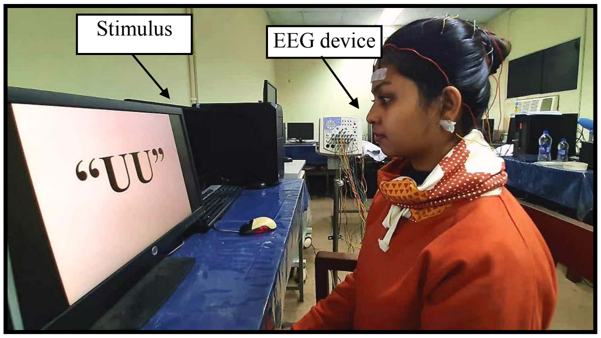

Fig 3 . An Experimental set up: EEG signal acquisition from the scalp of human subject S_6 during experiment

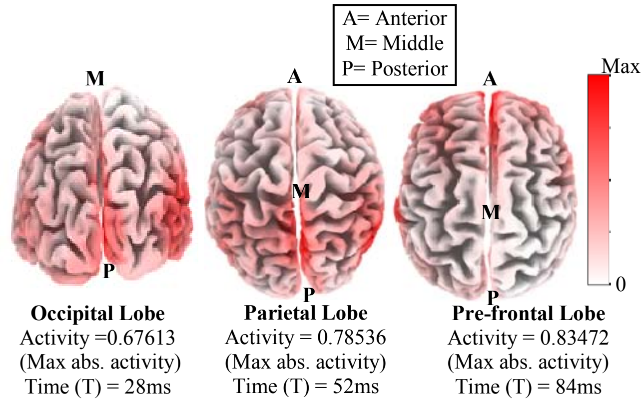

Fig. 5. 3D surface plot for vowel sound imagery

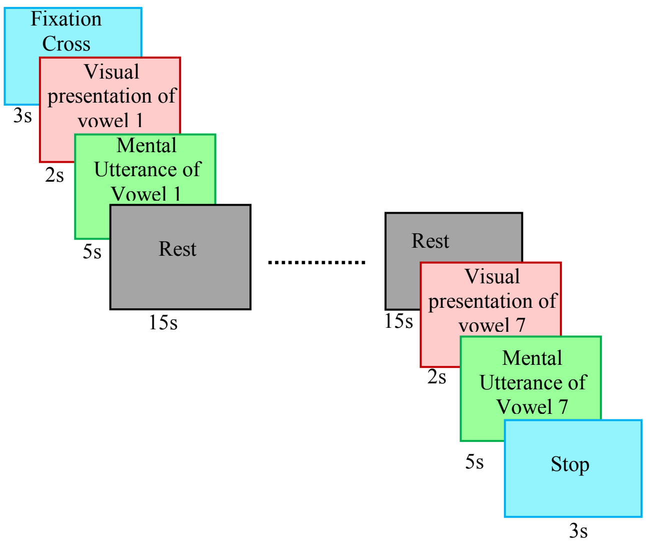

Fig. 4. Structure of stimulus presentation for vowel sound imageries for a single trial

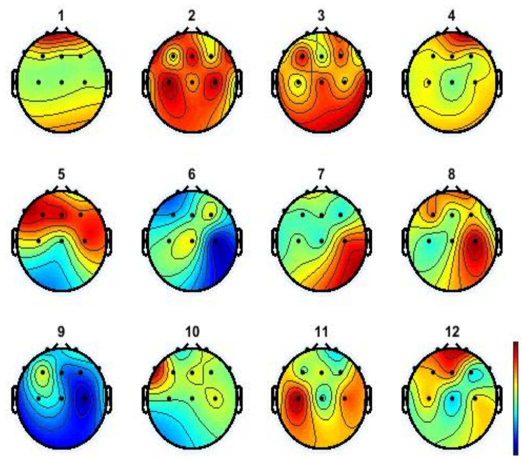

Fig. 6. Scalp map for components using ICA

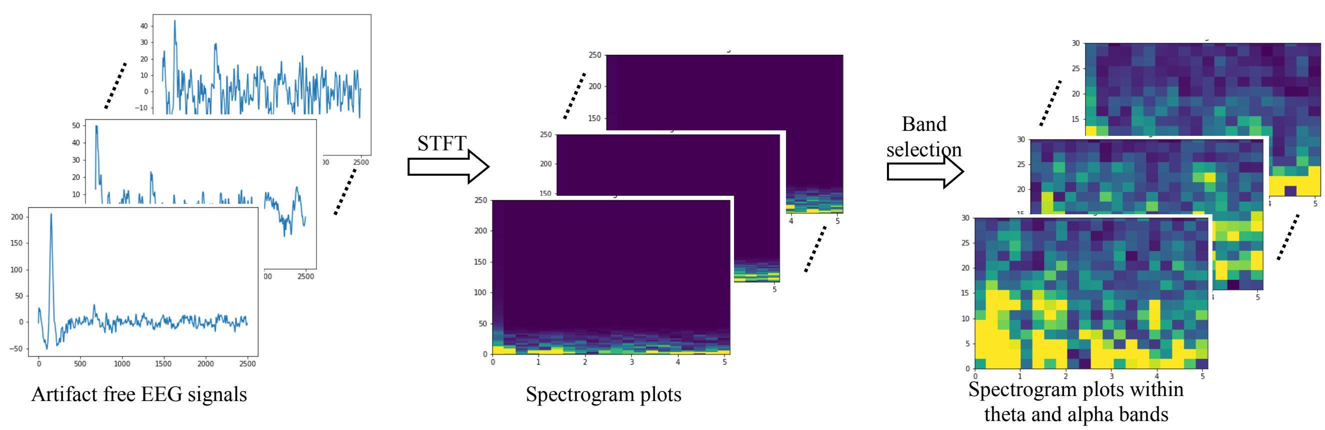

Fig. 7. Spectrogram analysis collected from active electrodes

EEG signals. 12 scalp maps are selected for further processing amongst the 19 procured maps as depicted in Fig. 6.

E. Experiment 3: EEG Spectogram analysis using STFT Algorithm.

The prime motivation of this experiment is to abstract the spectrogram images from artifact free EEG signals for uplifting the classification accuracy of the classifier based on deep-learning approach. Here, Short Time Fourier Transform (STFT) algorithm [22] has been adopted to translate the time-domain EEG signals to time-frequency domain signals. For each trial of a subject, a 3D data array has been constructed with the dimension of 129 × 21 × 12 feature vector, by utilizing STFT method consisting of Hamming window length of 256 with overlap length of 100 amongst consecutive window frames. Next, the procured feature set within the

TABLE I
RELATIVE STUDY OF CLASSIFICATION ACCURACY (STANDARD DEVIATION) FOR VOWEL SOUND AA

Classifiers with optimal parameter settings	CA	Specificity	Sensitivity	Run-time complexity
LDA [23]	71.58	0.69	0.72	51.85 ms
KNN [24]	72.15	0.75	0.73	52.44 ms
BPNN [25]	75.97	0.72	0.76	66.28 ms
LSVM [26]	78.74	0.79	0.78	58.15 ms
KSVM -RBF kernel [27]	81.96	0.78	0.82	55.74 ms
Type-1 fuzzy [28]	85.23	0.86	0.83	37.96 ms
Interval Type-2 Fuzzy Sets (IT2FS) [29]	87.52	0.88	0.86	41.82 ms
General Type-2 Fuzzy Sets (GT2FS) [30]	89.89	0.89	0.88	92.55 ms
CNN [31]	90.87	0.90	0.89	111 ms
Capsule network [6]	91.12	0.93	0.88	76.84 ms
Proposed Capsule network	**94.86**	**0.95**	**0.93**	**68.29 ms**

range of theta and alpha power has been selected since, they provide maximum power for mentally uttered vowel sounds. Thus, the feature space gets reduced to 21×21×12 dimensional vector which is finally fed to the classification module to classify seven vowel sound imageries. Fig. 7 provides the spectrogram analysis of the filtered EEG signals collected from 12 electrodes for a vowel sound (AA) of a single subject.

V. PERFORMANCE ANALYSIS AND STATISTICAL VALIDATION

A. Relative Performance Analysis

The relative performance of the proposed classifier has been evaluated on the basis of three metrics- Classification Accuracy (*CA*), Sensitivity and Specificity which are defined by equations (10), (11), (12).

$$CA = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{11}$$

$$Specificity = \frac{TN}{TN+FP} \tag{12}$$

where, *TP*, *TN*, *FP*, *FN* indicates the number of true positives, true negatives, false positives and false negatives respectively. Table I depicts the comparative analysis of the proposed classifier with other primitive ones. It is quite evident from this table that the performance of the proposed classifier surpasses other classifiers by an enhanced marginal level. Hence, this classifier is able to provide a precise and accurate categorization of seven specific vowel sound imageries.

TABLE II
COMPARATIVE STUDY OF CLASSIFICATION ACCURACY (STANDARD DEVIATION) WITH VARYING ACTIVATION FUNCTION FOR VOWEL SOUND AA

Activation function	CA (%)
Sigmoid [11]	85.76
Hyperbolic Tangent (Tanh) [32]	86.45
Rectified Linear Unit (RELU) [33]	89.13
Leaky RELU [34]	90.20
Exponential Linear Unit (ELU) [10]	92.15
Scaled Exponential Linear Unit (SELU) [35]	92.88
Swish [12]	93.75
Exponential Linear Sigmoid SquasHing (ELiSH) [8]	**94.86**

TABLE III
STATISTICAL ANALYSIS OF WITH THE PROPOSED CAPSULE NETWORK BASED CLASSIFIER WITH REFERENCE ALGORITHMS

Classifier algorithm used for comparison using desired features	Parameter used for Mc Nemar's test		z	Comments on acceptance/rejection of hypothesis
	n_{01}	n_{10}		
LDA	25	89	34.81	Rejected
KNN	11	51	24.53	Rejected
BPNN	14	36	8.82	Rejected
LSVM	17	59	22.11	Rejected
KSVM -RBF kernel	12	36	11.02	Rejected
Type-1 fuzzy	21	66	22.25	Rejected
Interval Type-2 Fuzzy Sets (IT2FS)	24	42	4.37	Rejected
General Type-2 Fuzzy Sets (GT2FS)	15	63	28.32	Rejected
CNN	36	77	14.16	Rejected
Capsule network	35	78	15.61	Rejected
Proposed Capsule network	**9**	**16**	**1.44**	**Accepted**

B. Performance Analysis on the basis of Activation functions

To further analyse the performance of the proposed classifier, the classification accuracy of the model is evaluated by varying the activation function in the first Convolution layer. This comparison is depicted in Table II. It can be clearly inferred from the table that the classification accuracy of the proposed ELiSH induced network outperforms the same model comprising of classical activation functions by a significant margin.

C. Statistical Analysis using McNemar's Test

The proposed classifier has been validated by utilizing the well known McNemar's test [36]. The z value utilized in the test has been illustrated by equation (13) where the all the terms in the equation convey the same significance as indicated in [36]. Two classifier algorithms X and Y, have been considered, where X denotes the proposed capsule network based classifier algorithm while the other standard classifiers listed in table-III are indicated by Y . We define two parameters n_{01} and n_{10}, where n_{01} depicts the number of classes misclassified by X but not by Y and n_{10} represents the number of classes misclassified by Y but not by X. Thus, we define z by the following equation

$$z = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \tag{13}$$

This analysis confirms that the null hypothesis for the primitive classifiers have been declined since the Z-score for all of them is beyond $\chi^2_{1,0.95} = 3.84$.

VI. Conclusion

This work provides an innovative approach to model and design a mind driven type writing system by utilizing the EEG signals acquired from subjects during the mental enunciation of seven vowel sounds. Experimental analysis performed by eLOERTA software infers the participation of occipital, pre-frontal and parietal lobes for this cognitive activity. The acquired signals are preprocessed and fed to a spectrogram analyzer module to abstract the most important features required for categorization of seven vowel sound imageries. These extracted features are then transmitted to the proposed classifier model which is able to successfully classify the seven class labels. Moreover, a codebook has also been developed to encode the consonants by utilizing two vowel sounds separated by a space. Since, the proposed capsule network based classifier outperforms the state-of-the-art techniques with respect to classification accuracy, it can serve as an important tool for the design of next generation mind driven type writers to aid patients suffering from various communication related impairments. Additionally, this technique can be utilized in military applications under precarious circumstances that demand non-verbal communication. Besides, the proposed approach can also be employed in various gaming applications to elate the experience of the users.

Acknowledgment

The authors are immensely thankful for the financial support provided by Ministry of Human Resource Development (MHRD) sponsored JU-RUSA 2.0 project to Jadavpur University, Kolkata- 700032, India.

References

[1] C. S. DaSalla, H. Kambara, M. Sato, Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Netw.*, vol. 22, no. 9, pp. 1334-1339, 2009.

[2] S. Iqbal, Y. Uzzaman Khan, O. Farooq, "EEG Based Classification of Imagined Vowel Sounds," *in proc. 2nd IEEE conference of Computing for Sustainable Global Development(India COM)*,2015.

[3] B. M. Idrees, O. Farooq, "Vowel classification using wavelet decomposition during speech imagery," *SPIN 2016*, pp. 636-640, 2016.

[4] N. Hashim, A. Ali, W. N. Mohd-Isa, "Word-based classification of imagined speech using EEG," *International Conference on Computational Science and Technology*, pp. 195-204, 2017.

[5] P. Saha, S. Fels, "Hierarchical deep feature learning for decoding imagined speech from eeg," *AAAI*, 2019.

[6] S. Sabour, N. Frosst, G. E. Hinton, "Dynamic routing between capsules," *Advances in Neural Information Processing Systems*, pp. 3859-3869, 2017.

[7] H. Chao, L. Dong, Y. Liu, B. Lu, "Emotion recognition from multiband EEG signals using CapsNet," *Sensors*, vol. 19, no. 9, pp. 2212, May 2019.

[8] M. Basirat, P. M. Roth, "The Quest for the Golden Activation Function," *arXiv* , 2018.

[9] S. Kanai, Y. Fujiwara, Y. Yamanaka, S. Adachi, "Sigsoflmax: Reanalysis of the Softmax Bottleneck," *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, 2018.

[10] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv*, 2015.[Online]. Available: http://arxiv.org/abs/1511.07289.

[11] J. Turian, J. Bergstra, and Y. Bengio, "Quadratic features and deep architectures for chunking," *in Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, vol. Companion Volume:, pp. 245–248, 2009.

[12] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," *ArXiv*, 2017.

[13] Z. Zhao, A. Kleinhans, G. Sandhu, I. Patel, K.P. Unnikrishnan, "Capsule Networks with Max-Min Normalization," *arXiv preprint arXiv:1903.09662*, 2019.

[14] A. Saha, A. Konar, A. Chatterjee, A. L. Ralescu & A. K. Nagar, "EEG analysis for olfactory perceptual-ability measurement using recurrent neural classifier," *IEEE Trans. Human-machine systems,* vol. 44, no. 6, pp. 717-730, Dec. 2014.

[15] A. Khasnobish, A. Konar, D. Tibrewala, and A. K. Nagar, "Bypassing the natural visual motor pathway to execute complex movement related tasks using interval type-2 fuzzy sets," *IEEE Trans. Neural System and Rehabilation Engineering*, vol. 25, no. 1, pp.88-102,Jan 2017.

[16] M. Hata, H. Kazui, T. Tanaka, R. Ishii, L. Canuet, R. D. Pascual-Marqui, Y. Aoki, S. Ikeda, H. Kanemoto, K. Yoshiyama, and M. Iwase, "Functional connectivity assessed by resting state EEG correlates with cognitive decline of Alzheimer's disease–An eLORETA study," *Clinical Neurophysiology*, vol. 127, no. 2, pp.1269-1278, 2016.

[17] L. Canuet, I. Ryouhei i, D. Pascual-M. Roberto, I. Masao, K. Ryu, A. Yasunori, I. Shunichiro, T. Hidetoshi, N. Takayuki, and T. Masatoshi, "Resting-state EEG source localization and functional connectivity in schizophrenia-like psychosis of epilepsy," *PloS one*, vol. 6, no. 11, pp. e27863, 2011.

[18] R. D. Pascual-Marqui, D. Lehmann, T. Koenig, K. Kochi, M. C. Merlo, D. Hell, & M. Koukkou, "Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia," *Psychiatry Research: Neuroimaging*, vol. 9, no. 3, pp. 169-179, 1999.

[19] P. Manoilov, "EEG eye-blinking artefacts power spectrum analysis," in *Proc. Int. Conf. Comput. Syst. Technol.*, 2006, pp. 3–5.

[20] A. Widmann, E. Schröger, and B. Maess, "Digital filter design for electrophysiological data–a practical approach," *Journal of neuroscience methods*, vol 250, pp 34-46, 2015.

[21] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "ICA: a potential tool for BCI systems," *IEEE Signal Processing Mag.*, vol. 25, no. 1, pp. 57-68, 2008.

[22] K. W. Ha, and J. W. Jeong, "Motor Imagery EEG Classification Using Capsule Networks," *Sensors*, vol. 19, no.13, pp: 2854, 2019.

[23] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920-928, 2011.

[24] A. Ahangi, M. Karamnejad, N. Mohammadi, R. Ebrahimpour, and N. Bagheri, "Multiple classifier system for EEG signal classification with application to brain–computer interfaces," *Neural Computing and Applications*, vol. 23, no. 5, pp. 1319-1327, 2013.

[25] V. N. Dao, and V. R. Vemuri, "A performance comparison of different back propagation neural networks methods in computer network intrusion detection," *Differential equations and dynamical systems*, vol. 10, no. 1 and 2, pp. 201-214, 2002.

[26] F. Pan, B. Wang, X. Hu, & W. Perrizo, "Comprehensive vertical sample-based KNN/LSVM classification for gene expression analysis," *J. of Biomedical Informatics*, vol. 37, no. 4, pp. 240-248, 2004.

[27] U. Ravale, N. Marathe, and P. Padiya, "Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function," *Procedia Computer Science*, vol. 45, pp. 428-435, 2015.

[28] J. M. Mendel and R. I. B. John, "Type-2 Fuzzy Sets Made Simple," *IEEE Transactions on fuzzy systems*, vol. 10, no. 2, pp; 117-127, 2002.

[29] A. Halder, A. Konar, R. Mandal, A. Chakraborty, P. Bhowmik, N. R. Pal, and A. K. Nagar, "General and interval type-2 fuzzy face-space approach to emotion recognition," *IEEE Trans. Systems, Man, and Cybernetics: Systems, vol.* 43, no. 3, pp. 587-605, 2013.

[30] A. Saha, A. Konar, and A. K. Nagar. "EEG Analysis for Cognitive Failure Detection in Driving Using Type-2 Fuzzy Classifiers," *IEEE Trans. Emerging Topics in Computational Intelligence*, vol. 1, no. 6, pp. 437-453, 2017.

[31] A. Krizhevsky, I. Sutskever, G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv Neural Inform Process Syst*, pp. 1097–1105, 2012.

[32] B. Karlik and A. Vehbi, "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks," *International Journal of Artificial Intelligence and Expert Systems (IJAE)*, vol. 1, no. 4, pp. 111–122, 2011.

[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Haifa*, pp. 807–814, 2010.

[34] A. Maas, A. Hannun, and A. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *in International Conference on Machine Learning (icml)*, 2013.

[35] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks," *arXiv*, 2017.

[36] X. Sun and Z. Yang, "Generalized mcnemars test for homogeneity of the marginal distributions," *in SAS Global forum*, vol. 382, pp. 1–10, 2008.