TERM PAPER PRESENTATION: ELL788

UNLEASH YOUR PSYCHE : A SYSTEMATIC REVIEW ON IMAGINED SPEECH RECOGNITION

AADARSH GUPTA	ASHWANI KUMAR
2019EE10451	2020EET2755

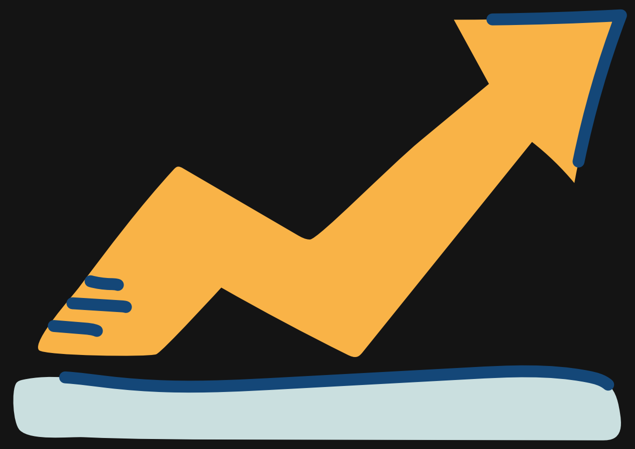
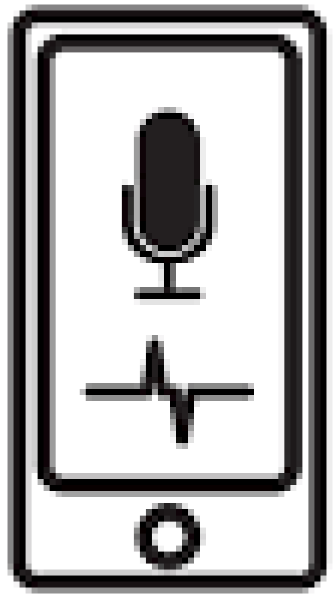
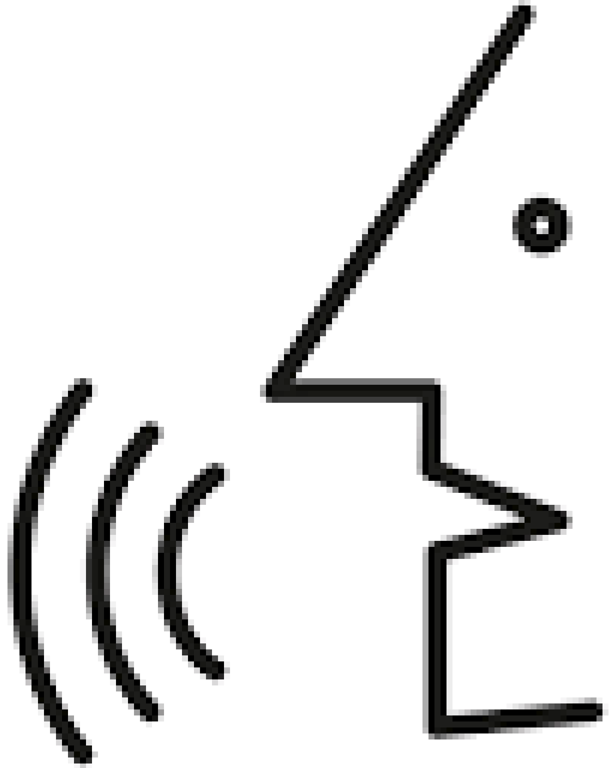


TABLE OF CONTENT

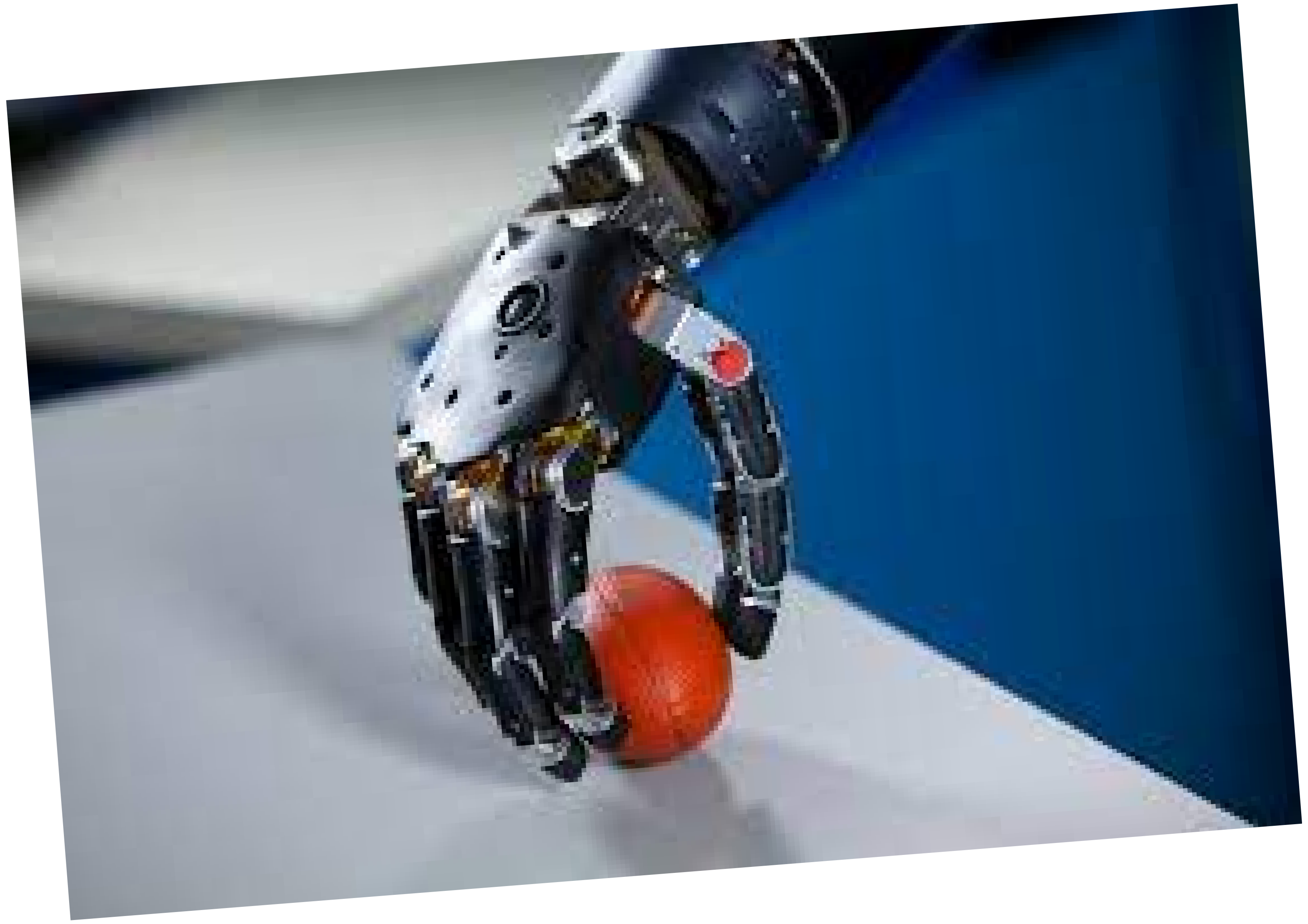
INTRODUCTION

Speech-driven interfaces have gained widespread acceptance and are used by a multitude of individuals in a vast variety of real-world applications and devices.





Most of the research is based on motor imagery-based control of external devices, which uses imagined hand, arm, or foot movements to deliver directional commands



WHAT??

- **Imagined speech is the internal pronunciation of phonemes, words, or sentences, without the movement of the phonatory apparatus or any audible output.**

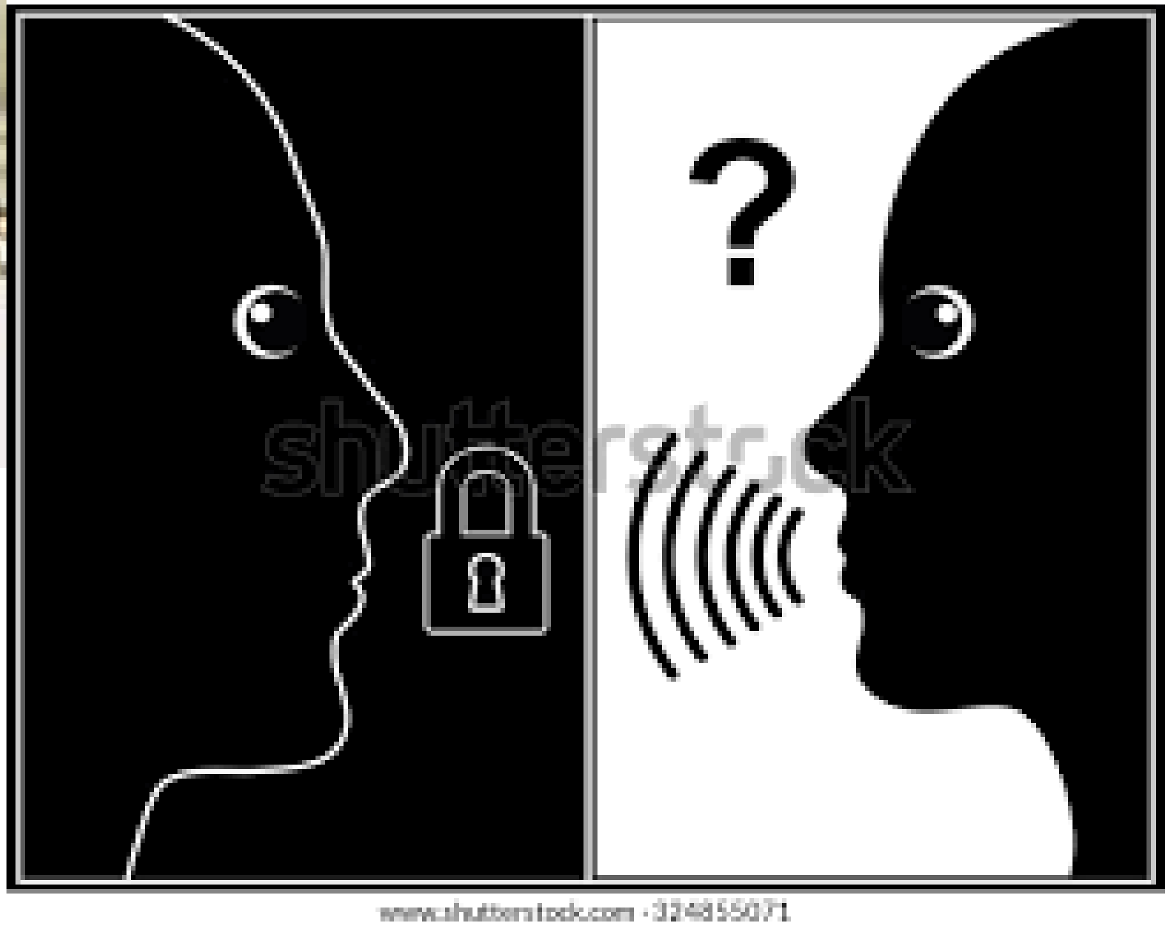
- **Imagining a word/vowel has effect on EEG alpha, beta and theta bands.**

WHY??

Mute and deaf persons

Physically disabled

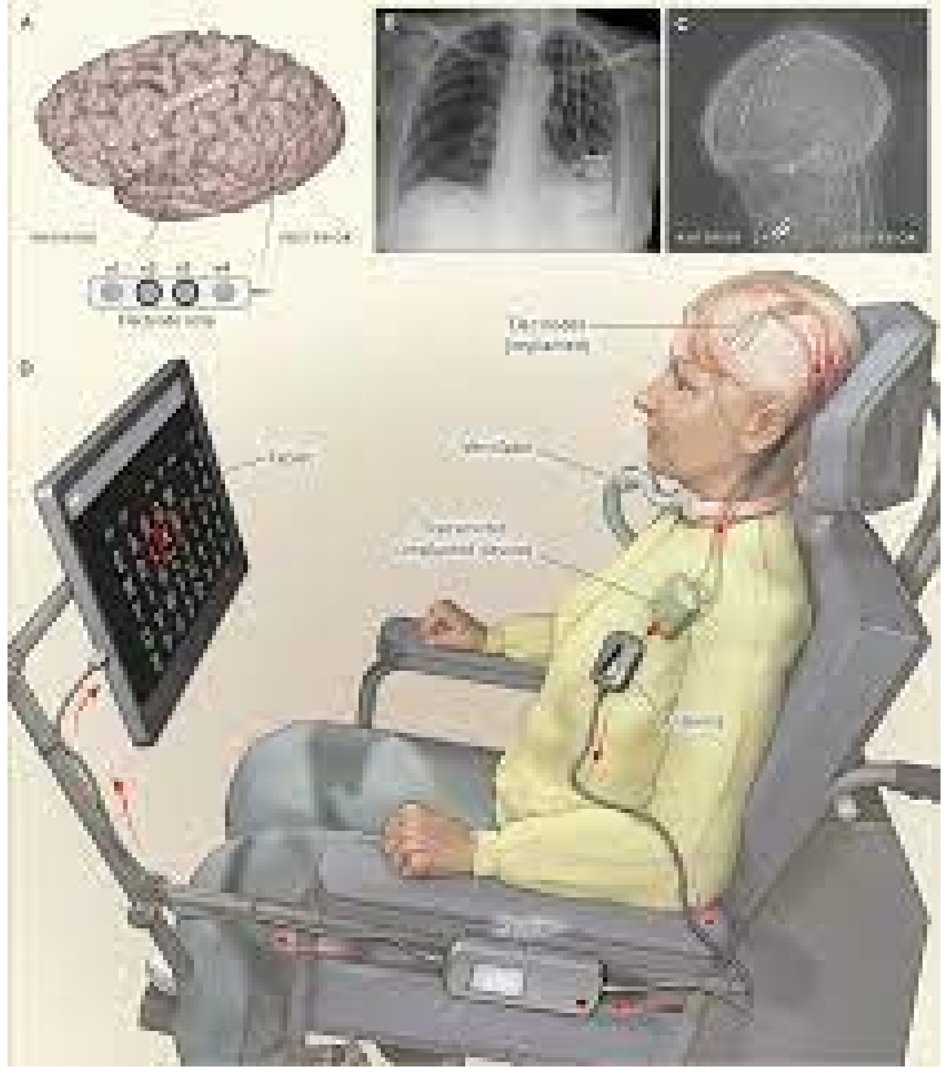
Silent communication



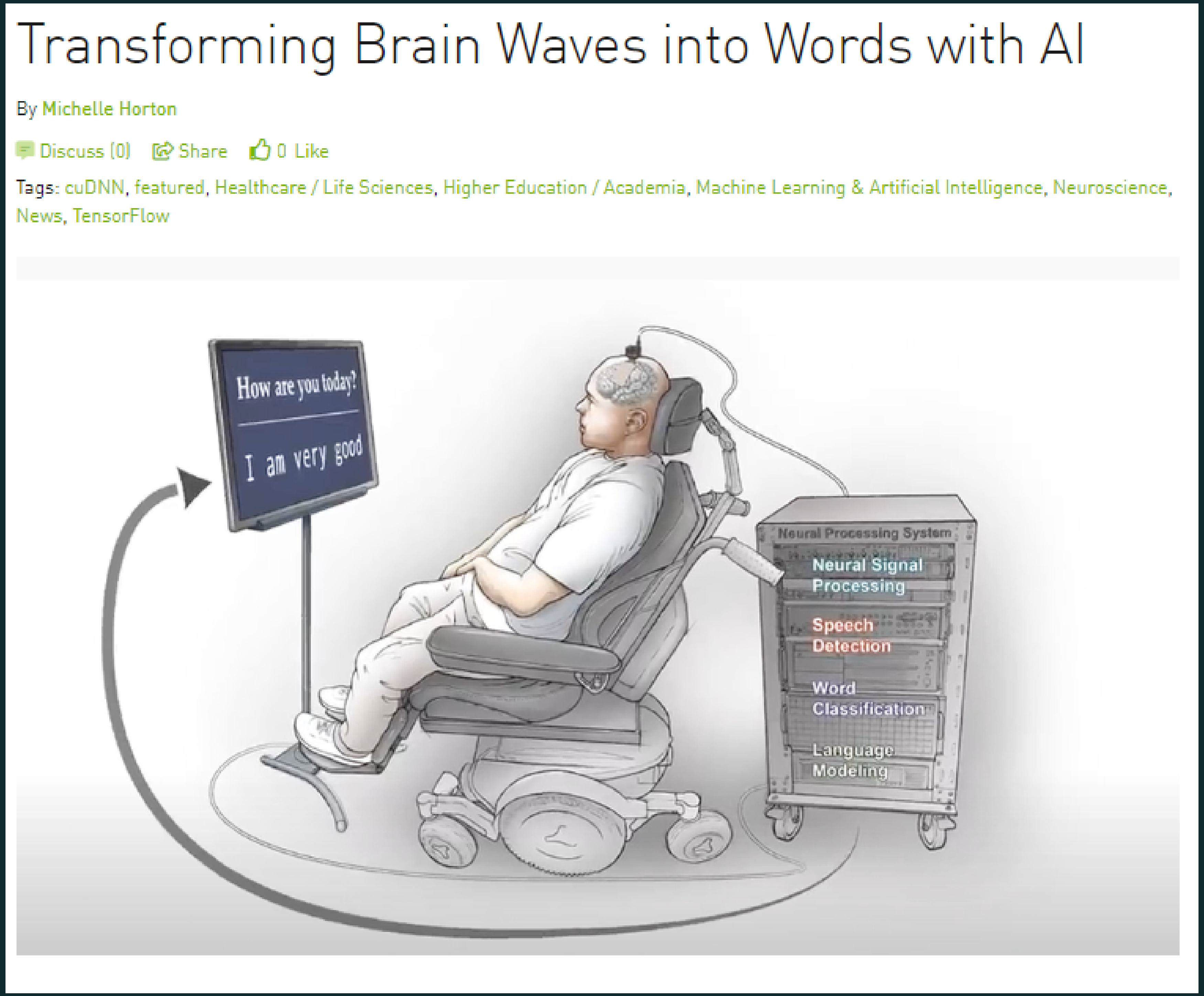
HOW??

- Brain Computer Interface (BCI), also referred to as human-machine interfaces, are systems that use brain signals to control computers or hardware devices.

- Such systems can equip the users with a medium to communicate and express their thoughts, thereby improving the quality of rehabilitation and clinical neurology.



TECHNICAL BLOG BY NVIDIA



Transforming Brain Waves into Words with AI

By Michelle Horton

Discuss (0) Share 0 Like

Tags: cuDNN, featured, Healthcare / Life Sciences, Higher Education / Academia, Machine Learning & Artificial Intelligence, Neuroscience, News, TensorFlow

DATASETS

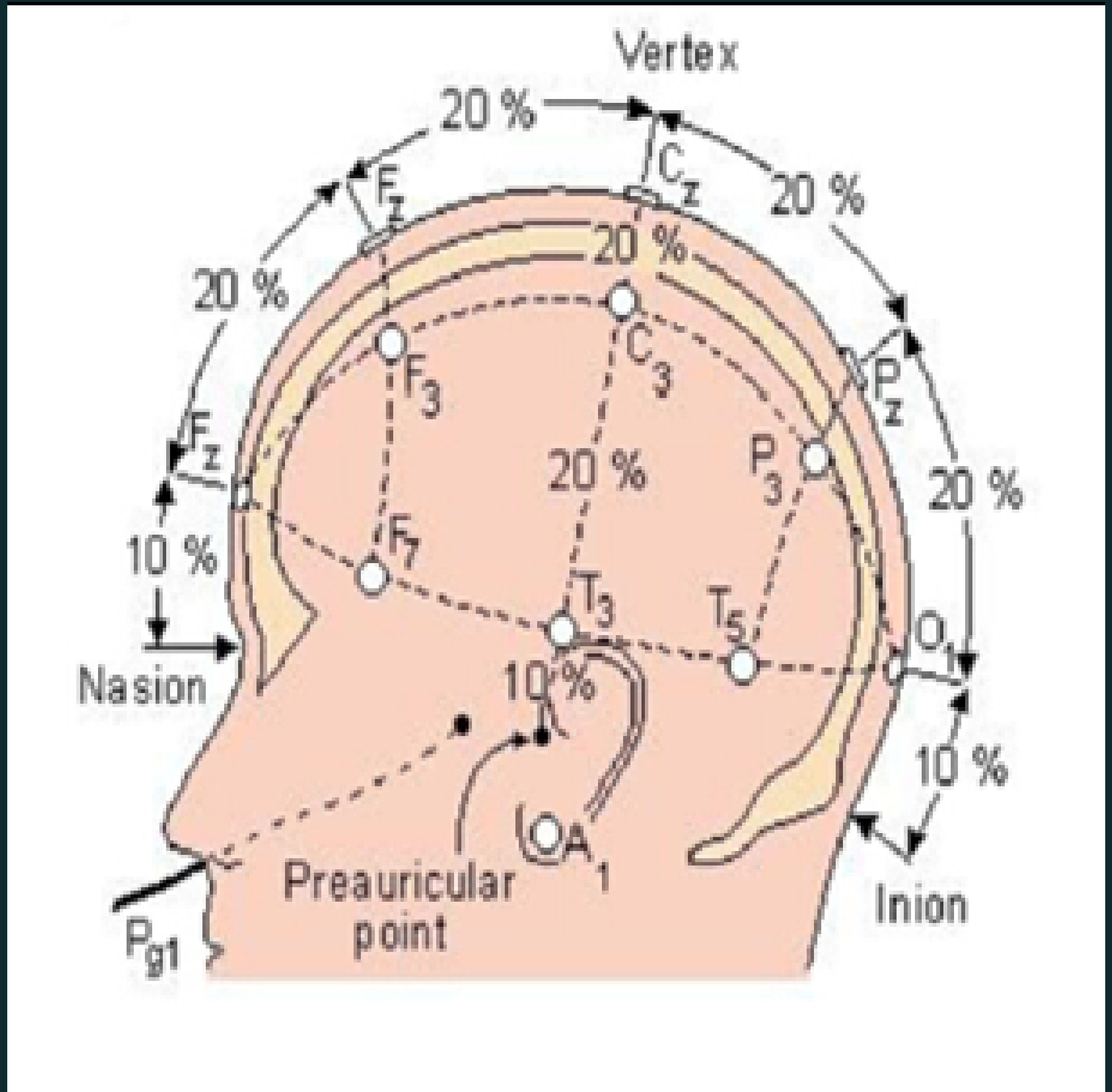
BD1

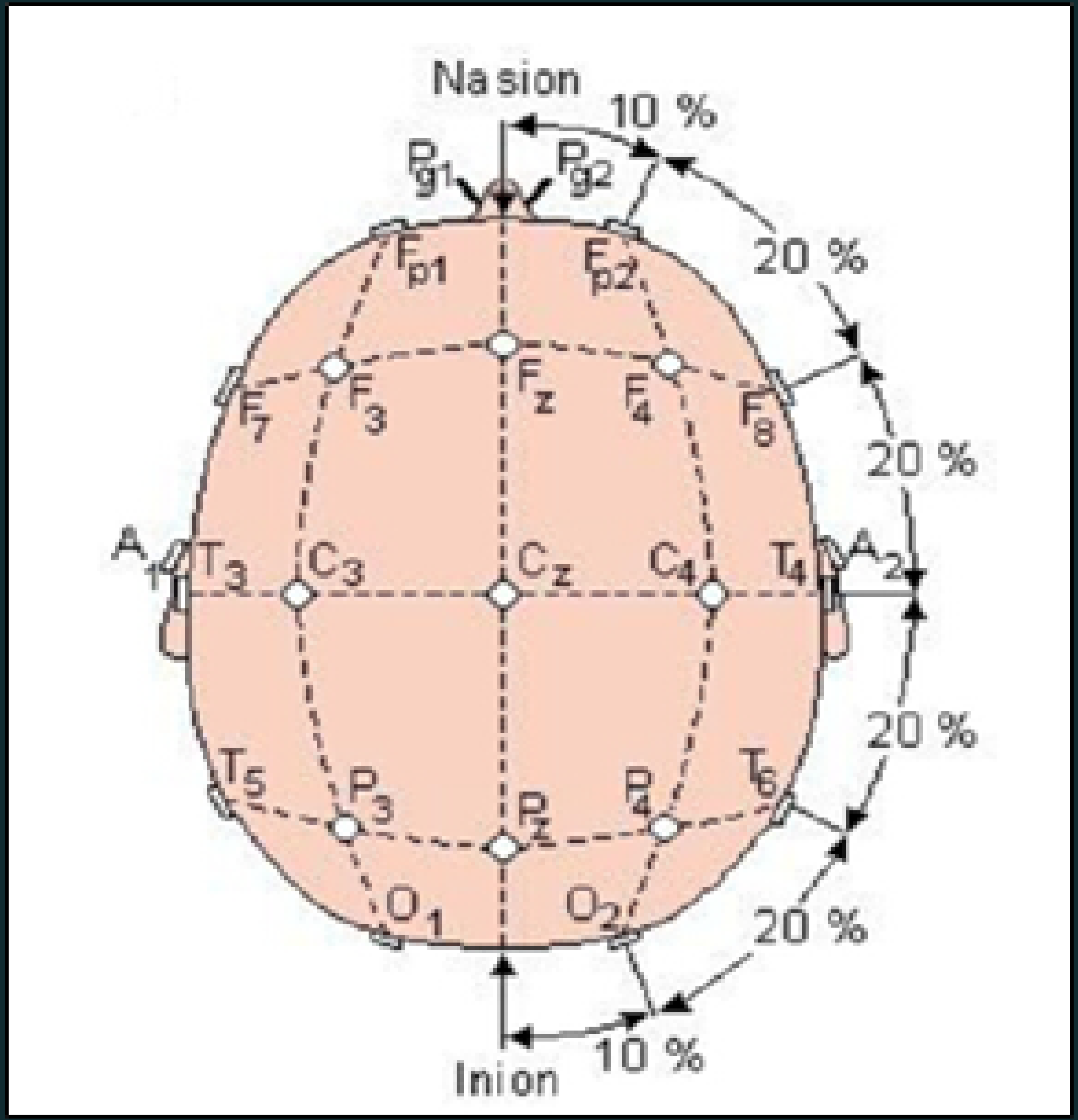
- Developed by Coretto et al. which records imagined speech tasks with 5 vowels and 4 words.
- Warning for 3 seconds before starting, then they were shown the vowel they had to imagine for two seconds. Next, they imagined the vowel continuously for four seconds .
- This procedure was repeated 40 times for each imagined vowel.

DATASETS

BD1

- EEG electrodes were located according to the international 10–20 system and the database contains information from six electrodes F3, F4, C3, C5, P3, and P4.

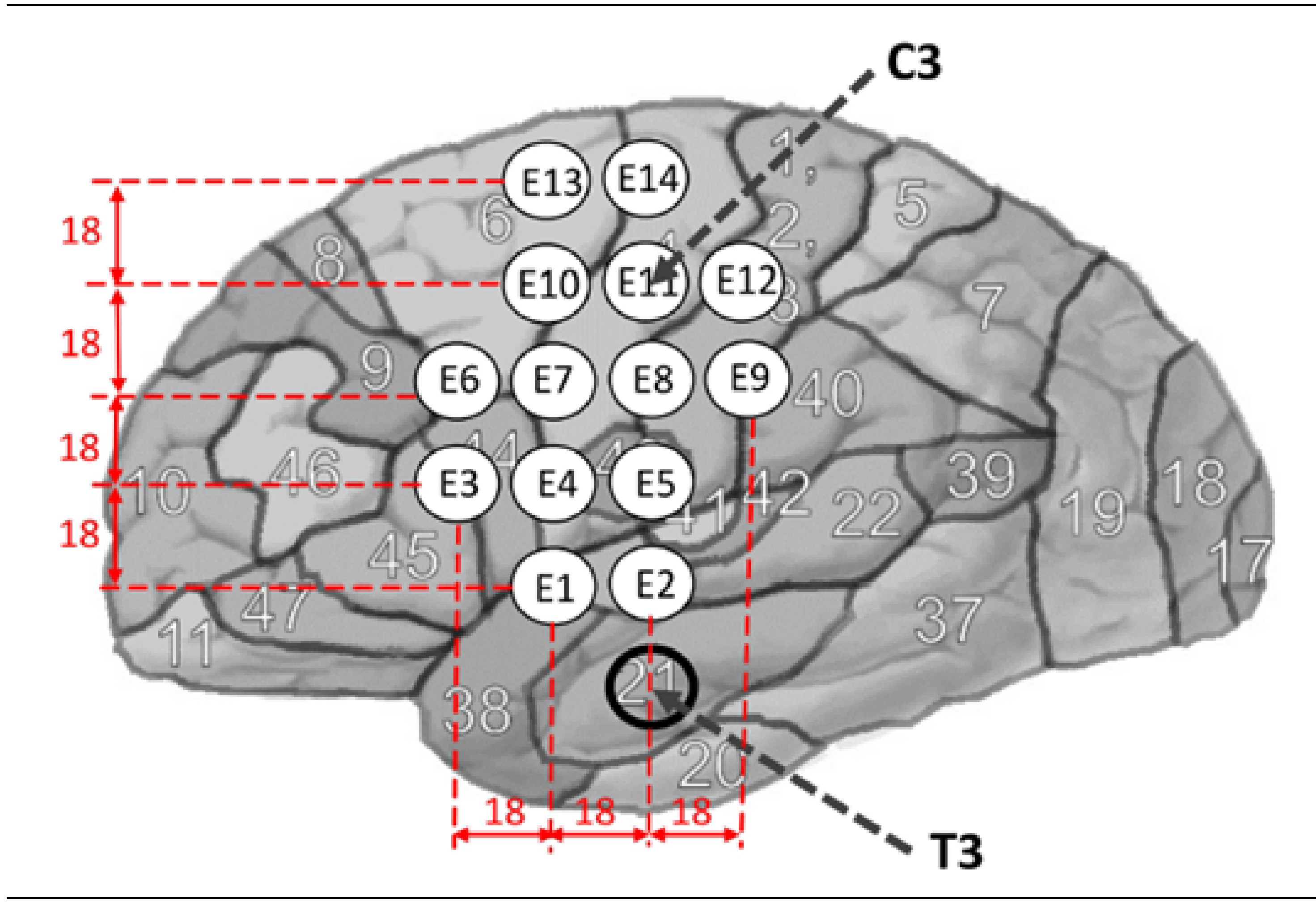




BD2

- Created specifically for the study, (20 women and 30 men) (M = 24.76, SD = 7.66).

- The neuroheadset has 14 electrodes located covering the language area. Two reference electrodes were located on the forehead.



KARA One

- **Database combines three modalities i.e. EEG, face tracking & audio during imagined and vocalized phonemic and single-word prompts.**
- **A 64- channel Neuroscan Quick-cap was used where the electrode placement follows the standard 10-20 system.**
- **A 5-second rest state, stimulus state, text would appear on the screen and its associated auditory utterance was played on speakers, followed by 2-second period relax.**
- **A 5-second imagined speech state, in which the participant imagined speaking the prompt without moving followed by speaking state,**

EFFICIENT ARCHITECTURES

SPEAK YOUR MIND!
Towards Imagined Speech Recognition With Hierarchical Deep Learning

Pramit Saha[1], *Muhammad Abdul-Mageed*[2], *Sidney Fels*[1]

[1]Human Communication Technologies Lab, University of British Columbia [2]Natural Language Processing Lab, University of British Columbia

pramit@ece.ubc.ca, muhammad.mageed@ubc.ca, ssfels@ece.ubc.ca

Abstract

Speech-related Brain Computer Interface (BCI) technologies provide effective vocal communication strategies for controlling devices through speech commands interpreted from brain signals. In order to infer imagined speech from active thoughts, we propose a novel hierarchical deep learning BCI system for subject-independent classification of 11 speech tokens including phonemes and words. Our novel approach exploits predicted articulatory information of six phonological categories (e.g., nasal, bilabial) as an intermediate step for classifying the phonemes and words, thereby finding discriminative signal responsible for natural speech synthesis. The proposed network is composed of hierarchical combination of spatial and temporal CNN cascaded with a deep autoencoder. Our best models on the KARA database achieve an average accuracy of 83.42% across the six different binary phonological classification tasks, and 53.36% for the individual token identification task, significantly outperforming our baselines. Ultimately, our work suggests the possible existence of a brain imagery footprint for the underlying articulatory movement related to different sounds that can be used to aid imagined speech decoding.

Index Terms: Brain Computer Interface, hierarchical deep neural network, phonological categories, Imagined Speech recognition, spatio-temporal CNN, deep autoencoder.

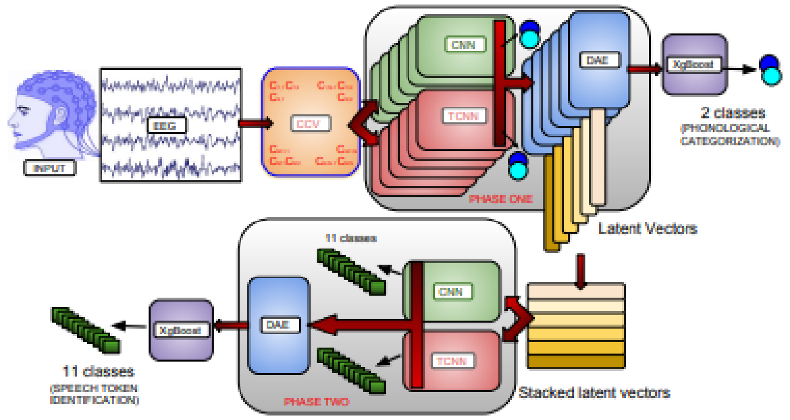


Figure 1: *Overall framework of the proposed approach*

Among the various brain activity-monitoring modalities in BCI, Electroencephalography (EEG) [3, 4] has been demonstrated as carrying promising signal for differentiating different brain activities (through measurement of related electric fields). However, these are high dimensional, and have poor Signal-to-Noise ratio, low spatial resolution, and plenty of artifacts. Besides, it is not entirely clear how to decode the desired information from the high-dimensional raw EEG signals.

sensors

MDPI

Article

Recognition of EEG Signals from Imagined Vowels Using Deep Learning Methods

Luis Carlos Sarmiento [1,*], Sergio Villamizar [2], Omar López [1], Ana Claros Collazos [1], Jhon Sarmiento [1] and Jan Bacca Rodríguez [2]

1 Departamento de Tecnología, Universidad Pedagógica Nacional, Bogotá 111321, Colombia; olopezv@pedagogica.edu.co (O.L.); asclarosc@upn.edu.co (A.C.C.); jfsarmientov@pedagogica.edu.co (J.S.)
2 Department of Electrical and Electronics Engineering, School of Engineering, Universidad Nacional de Colombia, Bogotá 111321, Colombia; sivillamizard@unal.edu.co (S.V.); jbaccar@unal.edu.co (J.B.R.)
* Correspondence: lsarmientov@unal.edu.co; Tel.: +57-1-594-1894

Abstract: The use of imagined speech with electroencephalographic (EEG) signals is a promising field of brain-computer interfaces (BCI) that seeks communication between areas of the cerebral cortex related to language and devices or machines. However, the complexity of this brain process makes the analysis and classification of this type of signals a relevant topic of research. The goals of this study were: to develop a new algorithm based on Deep Learning (DL), referred to as CNNeeg1-1, to recognize EEG signals in imagined vowel tasks; to create an imagined speech database with 50 subjects specialized in imagined vowels from the Spanish language (/a/,/e/,/i/,/o/,/u/); and to contrast the performance of the CNNeeg1-1 algorithm with the DL Shallow CNN and EEGNet benchmark algorithms using an open access database (BD1) and the newly developed database (BD2). In this study, a mixed variance analysis of variance was conducted to assess the intra-subject and inter-subject training of the proposed algorithms. The results show that for intra-subject training analysis, the best performance among the Shallow CNN, EEGNet, and CNNeeg1-1 methods in classifying imagined vowels (/a/,/e/,/i/,/o/,/u/) was exhibited by CNNeeg1-1, with an accuracy of 65.62% for BD1 database and 85.66% for BD2 database.

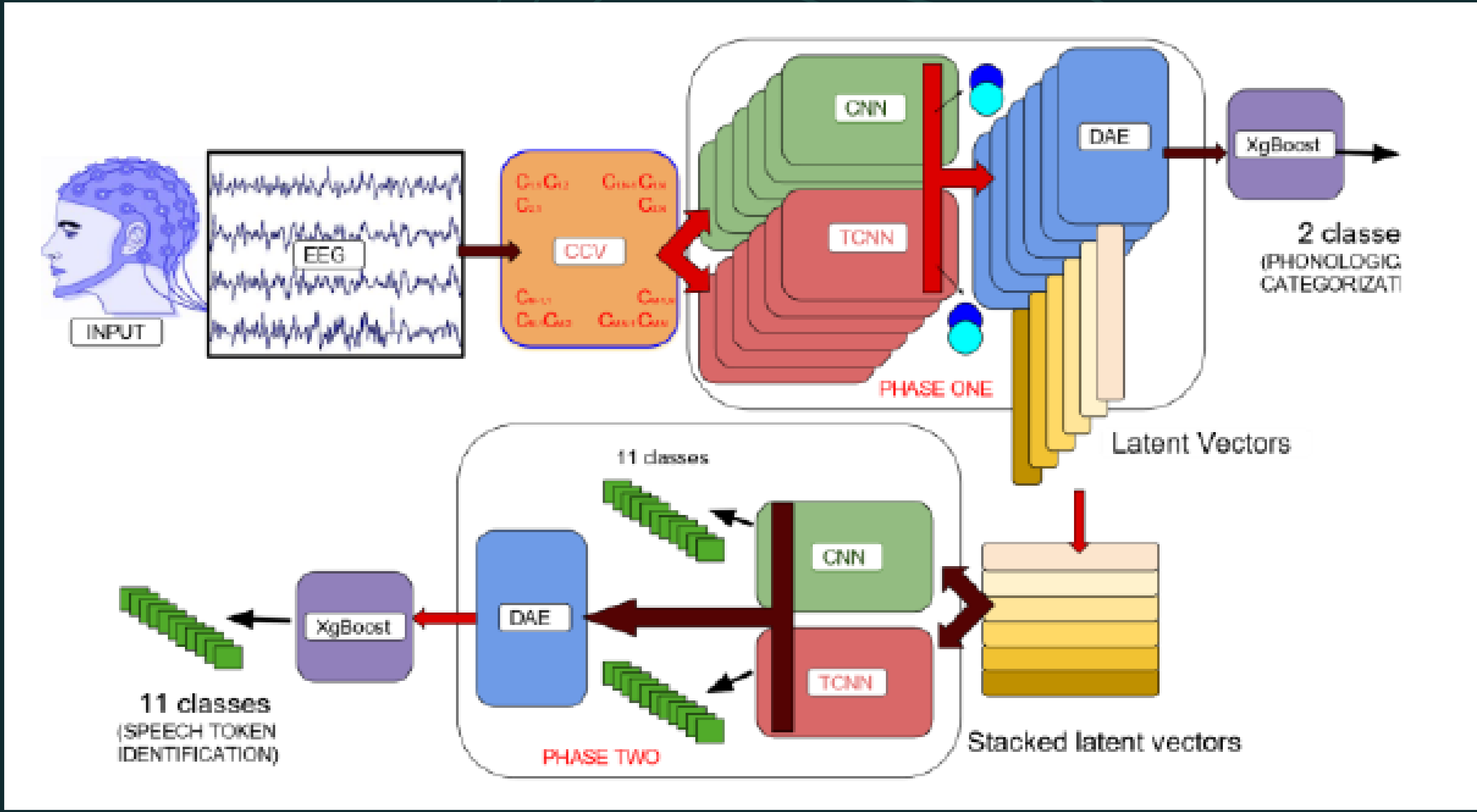
Keywords: imagined speech; electroencephalography; brain-computer interface (BCI); deep learning; convolutional neural networks (CNN); vowels

check for updates

Citation: Sarmiento, L.C.; Villamizar, S.; López, O.; Collazos, A.C.; Sarmiento, J.; Rodríguez, J.B. Recognition of EEG Signals from Imagined Vowels Using Deep Learning Methods. *Sensors* **2021**, *21*, 6503. https://doi.org/10.3390/s21196503

NETWORK ARCHITECTURE

CNN + TCNN + DAE



WORKING

- **Detect speech tokens from speech imagery based on KARA ONE database**

- **A 4-layered 2D Convolution Neural Networks (CNNs) has been used to extract spatial features from the covariance matrix to predict phononical categories, while 6-layered temporal CNN (TCNN) has been used in parallel on the channel covariance matrices to investigate long-term dependencies and temporal correlations of the signal.**

- **To lower the dimensionality of the spatial temporal recordings and remove background noise effects, an unsupervised deep autoencoder (DAE) on the fused heterogeneous features obtained by the CNN and TCNN was employed.**

WORKING

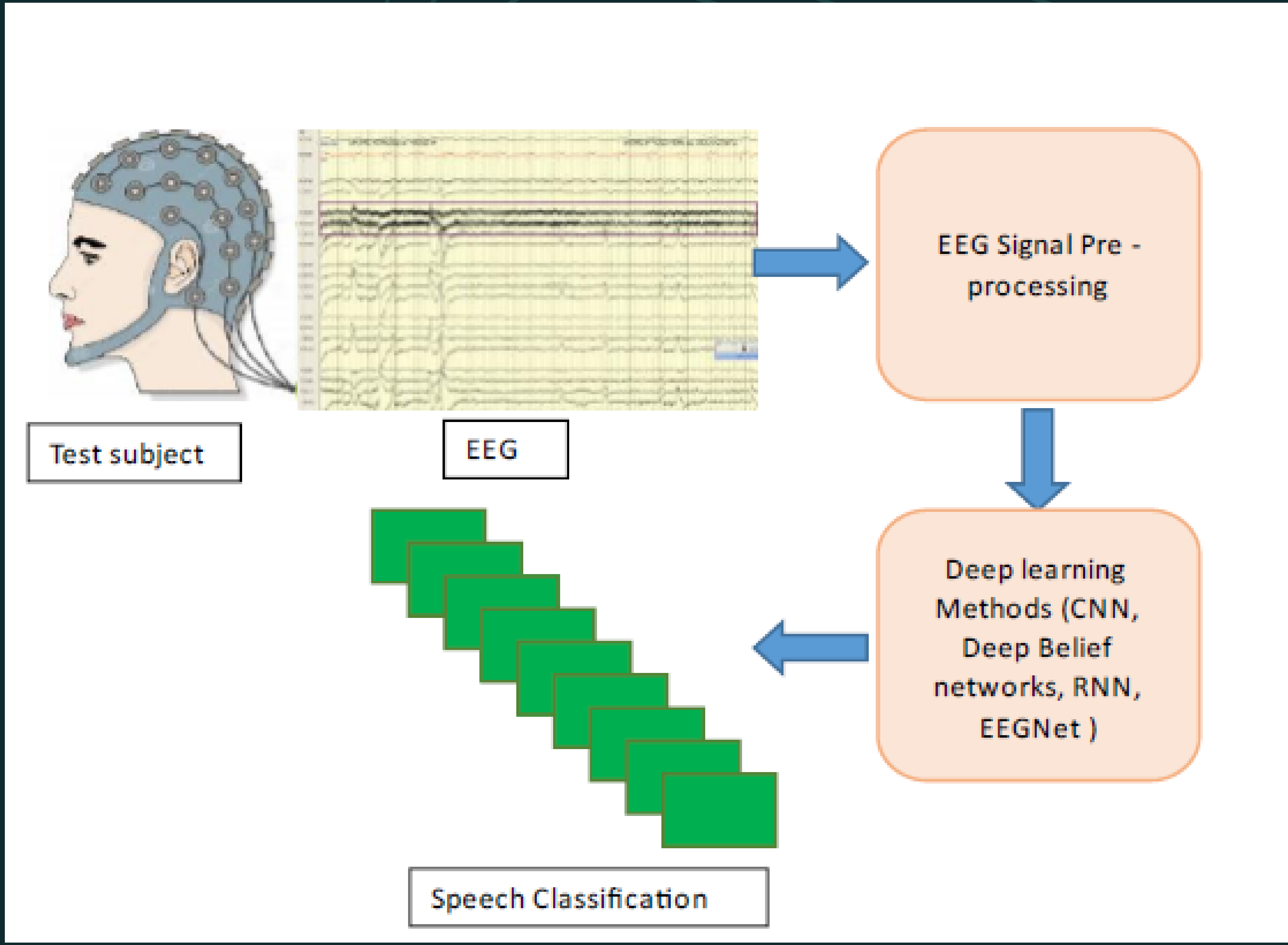
- **Achieves an accuracy of 83.42% across six different binary phonological classification and 53.36% for the individual token identification task.**
- **Comparison of accuracy for 10% test data for speech token prediction task**

Method	EEG Data	Phonological Features
LSTM	8.45	15.83
CNN	8.88	16.02
CNN+LSTM	12.44	22.10
CNN+LSTM+DAE	23.45	49.19
Proposed model	**28.08**	**53.36**

- **The bottleneck characteristics of the autoencoders are stacked into a 6x256 matrix that helps predict 11 unique speech tokens in the EEG dataset.**

NETWORK ARCHITECTURE

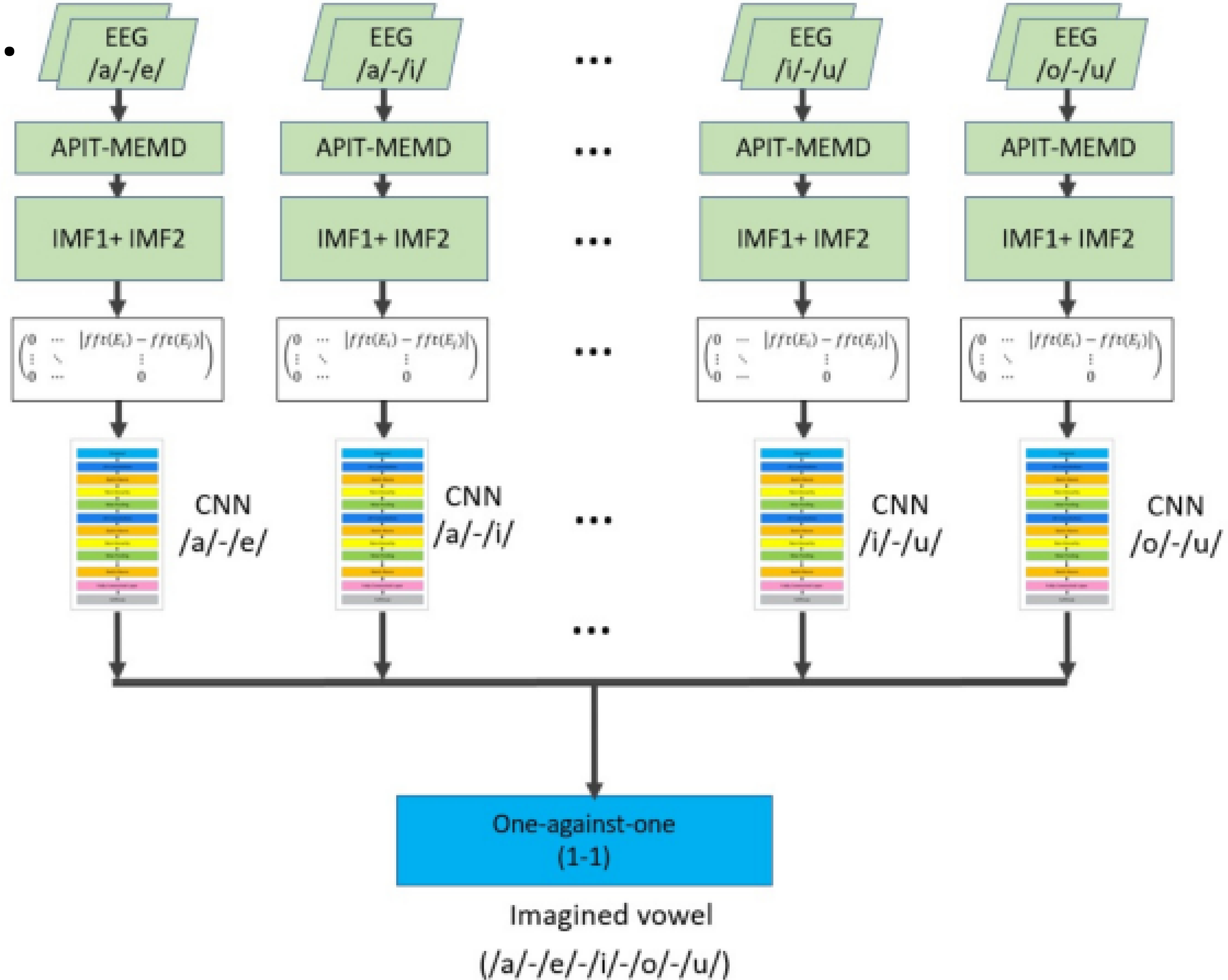
CNNeeg-1



WORKING

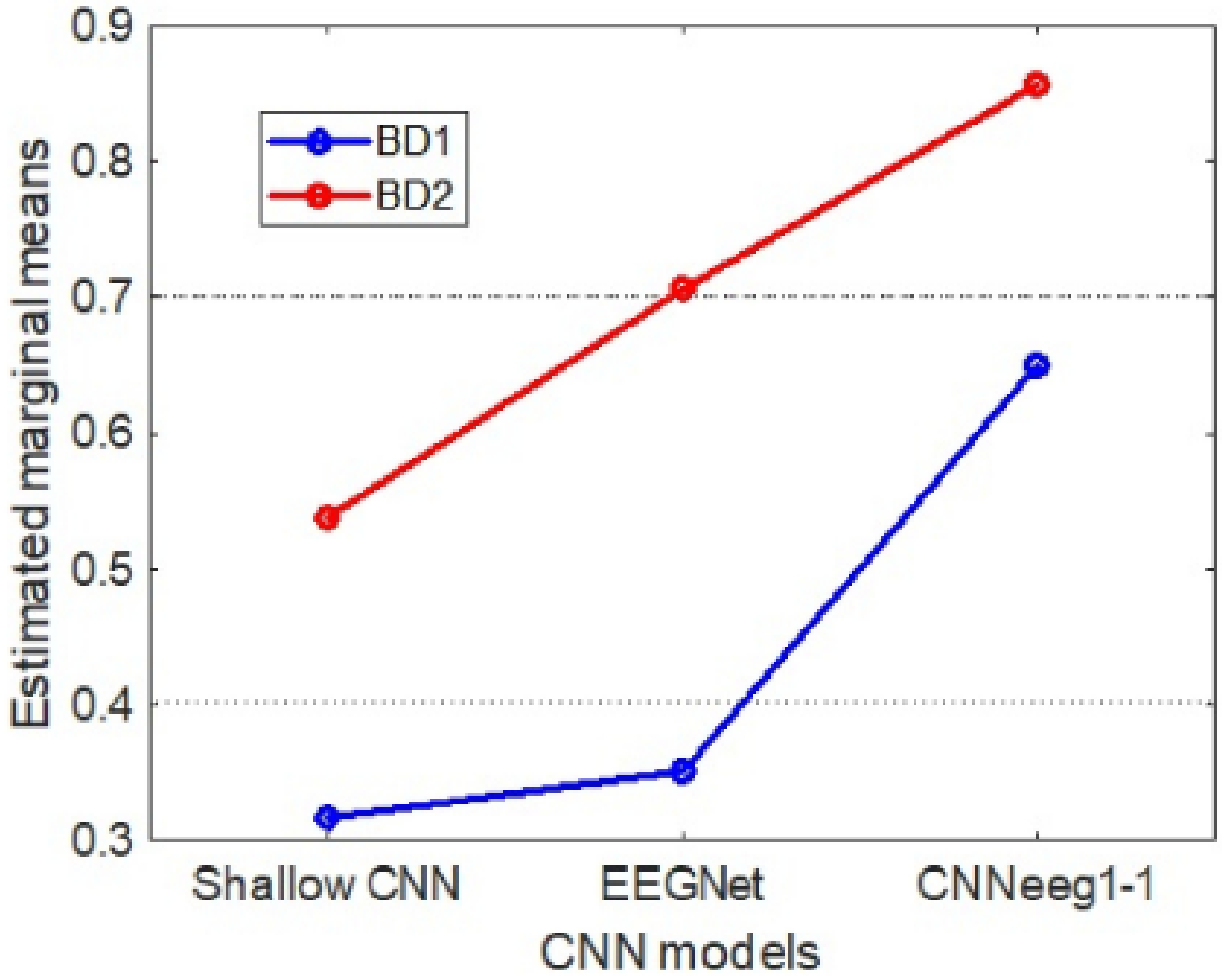
- The proposed architecture consists of 10 CNNs using deep learning for one of the imagined speech pairs: (/a/-/e/), (/a/-/i/), (/a/-/o/), (/a/-/u/), (/e/-/i/), (/e/-/o/), (/e/-/u/), (/i/-/o/), (/i/-/u/), (/o/-/u/).

- The input layer for each CNN receives the information of the images obtained from the EEG imagined vowels. It consists of a tensor of size 32 x 15 x 1 for database BD1 and 32 x 91 x 1 for database BD2



Results

- **The evaluated mean and standard deviation accuracy of the models on BD1 are:**
  a) Shallow-CNN: μ = 0.3171, σ = 0.0114
  b) EEGNet: μ = 0.3506, σ = 0.0133
  c) CNNeeg1-1: μ = 0.6562, σ = 0.0123



Conclusion

- The proposed model comprising of CNN+TCNN+DAE outperforms baselines on KARA ONE database for the speech token prediction task, when cross-covariance (CCV) of the EEG signals is evaluated.
- Statistical results were presented for CNNeeg1-1 based on deep learning architectures for EEG imagined vowel signal recognition using two different databases: BD1, with 15 subjects and BD2, with50 subjects.
- CNNeeg1-1 outperforms both Shallow CNN and EEGNet for EEG imagined vowel classification in intra-subject and inter-subject training analysis with both databases.

References

[1] Saha, P., Abdul-Mageed, M. and Fels, S., 2022. SPEAK YOUR MIND! Towards Imagined Speech Recognition With Hierarchical Deep Learning. [online] arXiv.org. Available at: <https://arxiv.org/abs/1904.05746> [Accessed 15 April 2022].
[2] Herff, C. and Schultz, T., 2022. Automatic Speech Recognition from Neural Signals: A Focused Review.
[3] Bakhshali, M., Khademi, M., Ebrahimi-Moghadam, A. and Moghimi, S., 2022. EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density.
[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct braincomputer communication," Proceedings of the IEEE, vol. 89, no. 7, pp. 1123–1134, 2001.
[5] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 16, no. 4, pp. 317–326, 2008.
[6] S. Machado, F. Araujo, F. Paes, B. Velasques, M. Cunha, ´ H. Budde, L. F. Basile, R. Anghinah, O. Arias-Carrion, M. Cagy ´ et al., "Eeg-based brain-computer interfaces: an overview of basic concepts and clinical applications in neurorehabilitation," Reviews in the Neurosciences, vol. 21, no. 6, pp. 451–468, 2010.
[7] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Ar- ´ naldi, "A review of classification algorithms for eeg-based brain– computer interfaces," Journal of neural engineering, vol. 4, no. 2, p. R1, 2007.
[8] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in ICASSP,

THANK YOU!