Unlocking Communication: Neural Underpinnings To Verbal Speech Synthesis.

Owais Mujtaba Khanday[1], Asmaa Sbaih[1], Jose L. Perez-Cordoba[1], Laura Miccoli[2], Marc Ouellet[3], Alberto Galdon[4], Gonzalo Olivares[4], Jose A. Gonzalez-Lopez[1]

[1] CITIC-UGR, Centro de Investigación en Tecnologías de la Información y las Comunicaciones, University of Granada, 18014, Granada, Spain

[2] CIMCYC, Mind, Brain, and Behaviour Research Center, University of Granada, 18071, Granada, Spain.

[3] Departamento de Psicología Experimental, Facultad de Psicología, University of Granada, 18017, Granada, Spain

[4] Unidad de Epilepsia Refractaria (CSUR), Hospital Universitario Virgen de las Nieves, 18014, Granada, Spain.

Introduction

Neural substrates supporting auditory linguistic processing, production, and comprehension have been elucidated through the application of various neuroimaging techniques, like magnetoencephalography (MEG), electroencephalography (EEG), and functional magnetic resonance imaging (fMRI). The intricate interaction among these brain regions during speech tasks highlights the complexity of language processing mechanisms. Functional connectivity analyses, integrated with artificial intelligence (AI) methodologies, reveal the nuanced dynamics by which divergent cerebral regions synchronize and interact to facilitate the seamless execution and comprehension of speech. This heightened understanding augments interpretive capacity concerning neural activation patterns associated with language pathologies like aphasia, dysarthria, and stuttering, thereby refining diagnostic protocols and therapeutic interventions. Furthermore, machine learning and deep learning-based algorithms expedite the deciphering of these neural activity patterns, thereby propelling the advancement of speech synthesis technologies. Deciphering neural signals pertinent to speech holds paramount significance by transducing neural signals into synthesized speech or text, furnishing a pivotal means of facilitating verbal expression for those grappling with articulatory disabilities. Our research endeavours to narrow the communication gap among individuals experiencing articulatory disabilities by decoding verbal speech directly from neural activity.

Experimental Setup

Non-invasive Approach

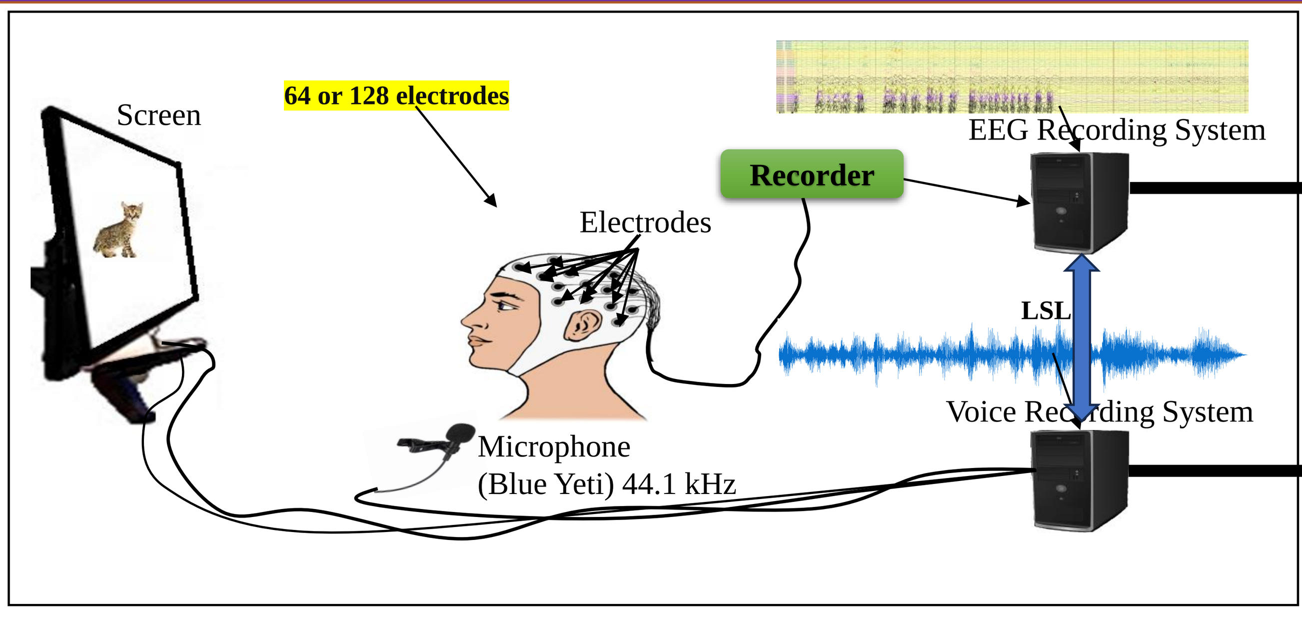

Figure 1. The experimental setup for collection of neurological data using non-invasive technique for speech synthesis

Invasive Approach

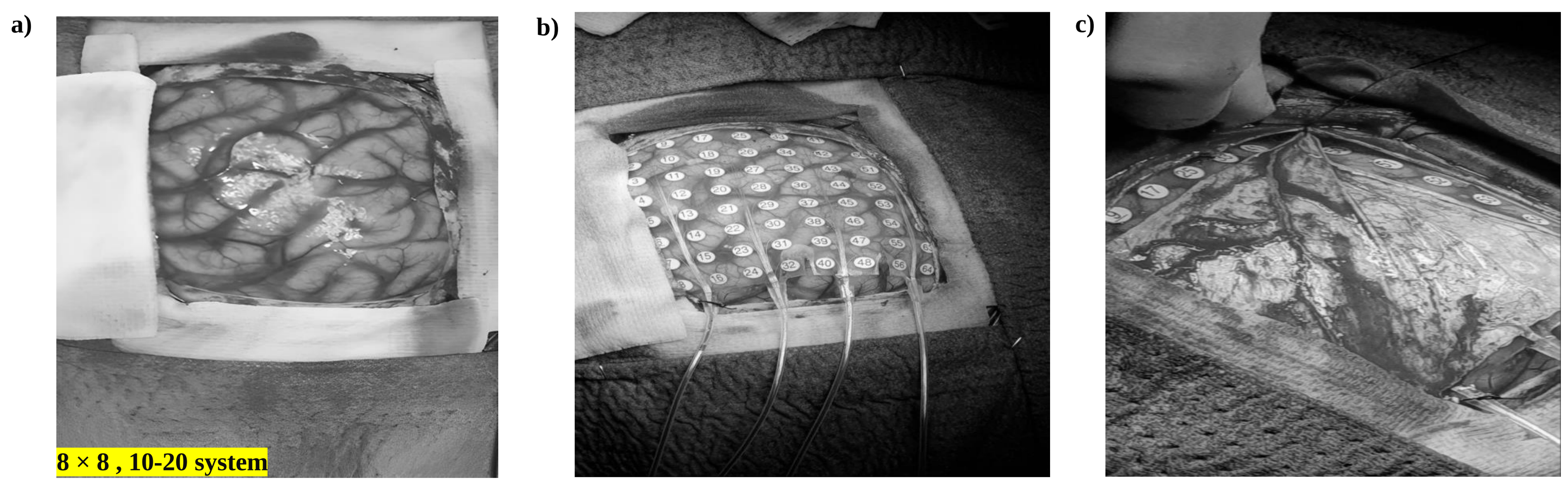

Figure 2. a) opening for invasive data collection subdural electrodes b) Electrode placement c) closing of wound

Electrode plantation locations

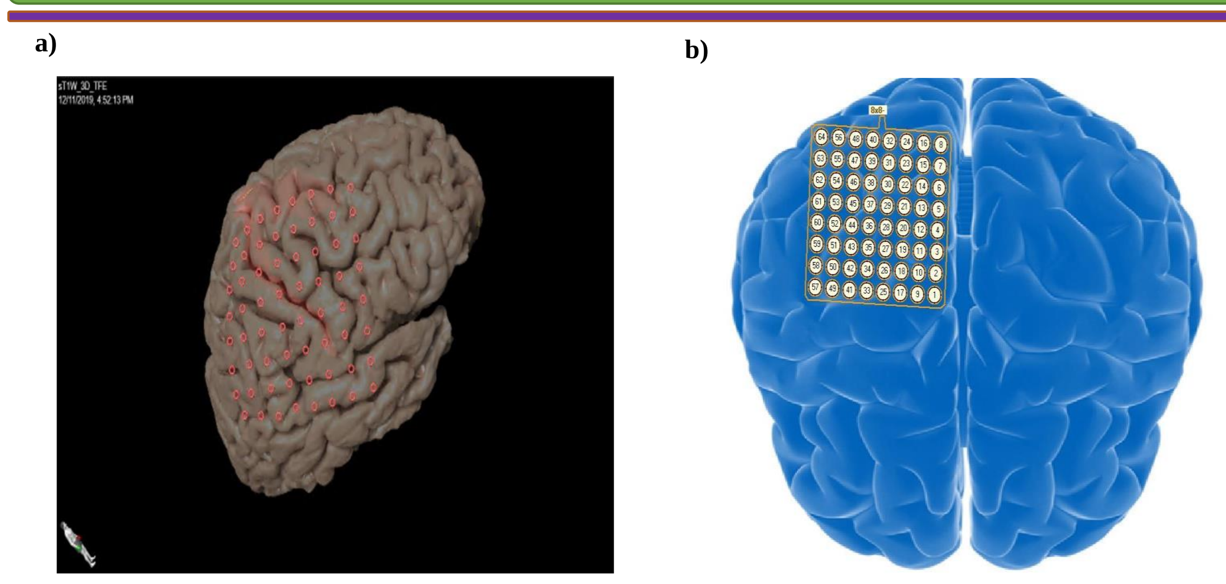

Figure 3. a) and b) Areas of the brain where electrodes are placed

Participant Info

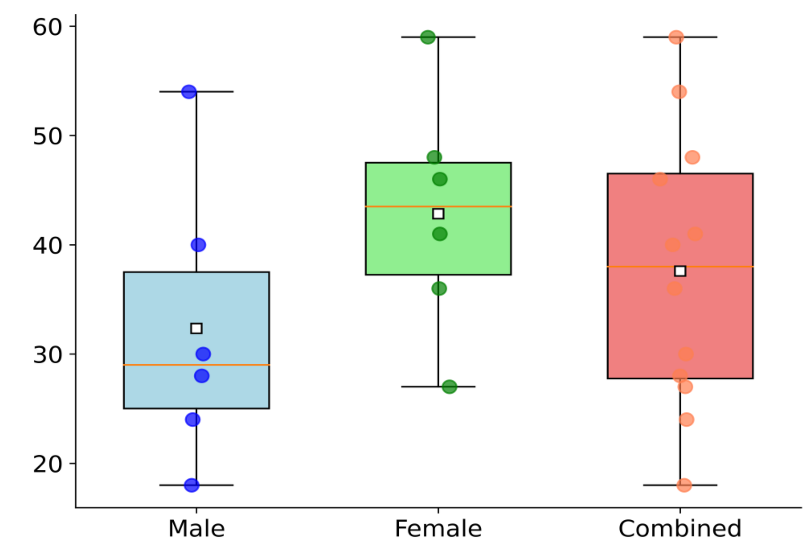

Figure 4. Participants Age Distribution

Stimulus Setup

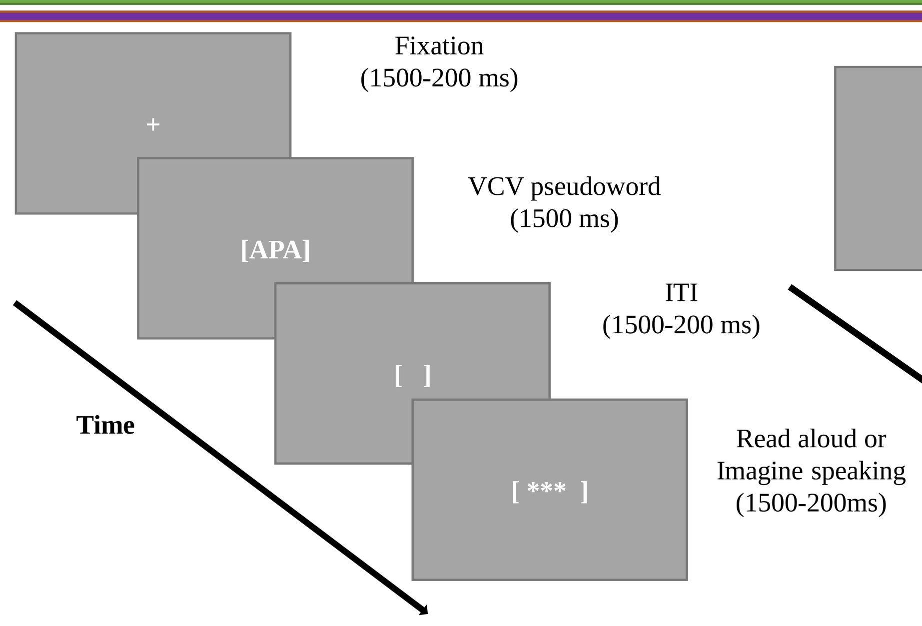

Figure 5. Vowel Consonant Vowel (VCV) task for language production

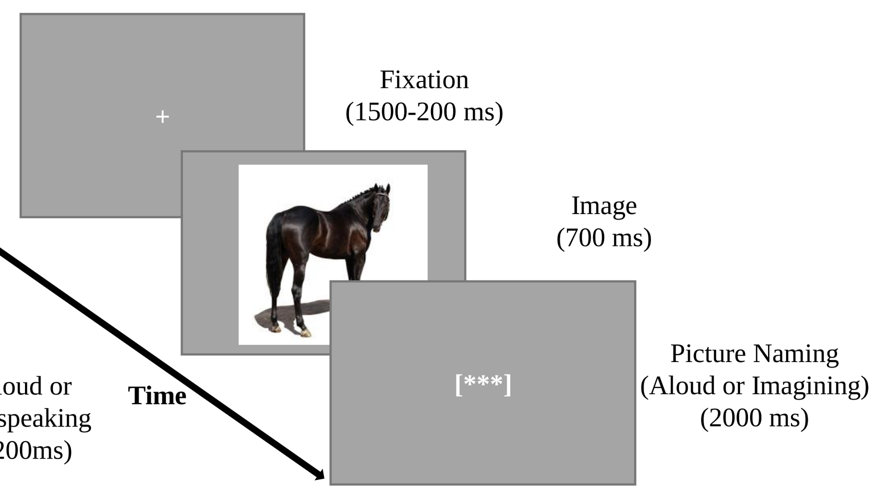

Figure 6. Picture naming task

Methodology

- 50 VCV syllables
- 5- Spanish vowels (a, e, i, o, u)
- 10- Spanish consonants (p, m, f, t, n, r, s, l, n, k, x)
- 6 semantic categories
- 5 images in each category
- 6 times each image is recorded
- Total of 6 × 5 × 6 = 180 image recordings per session. (spoken)
- Total of 6 × 5 × 6 = 180 image recordings per session. (imagined)

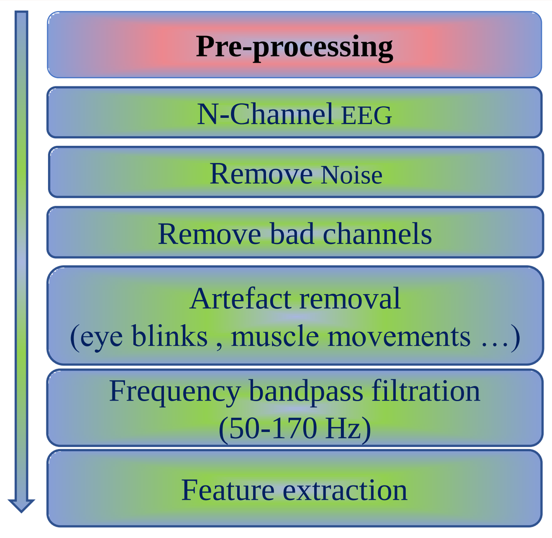

Figure 7. Pre-processing of EEG Flowchart

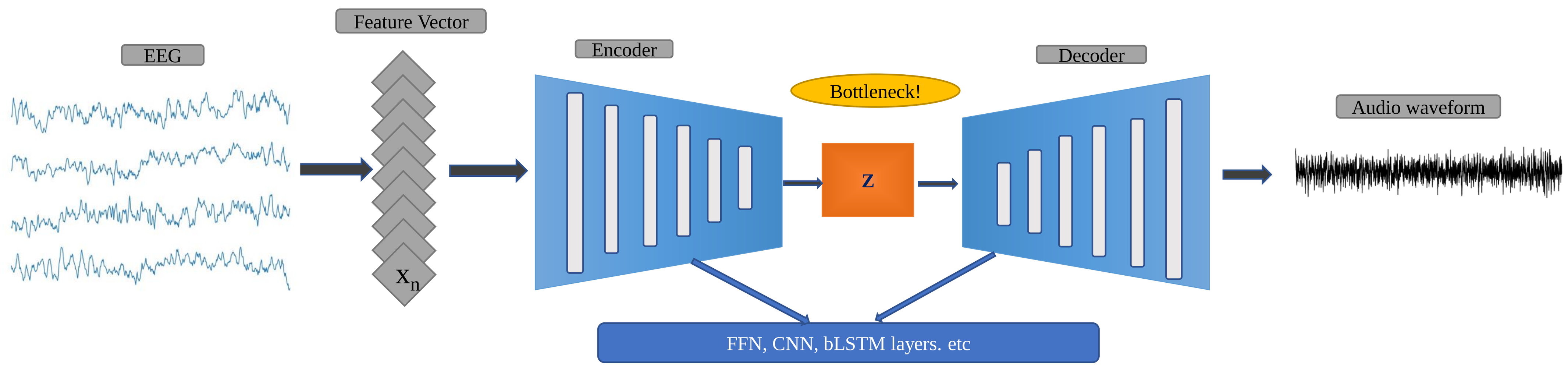

Figure 8. Autoencoder based DL model for speech synthesis

Results and Summary

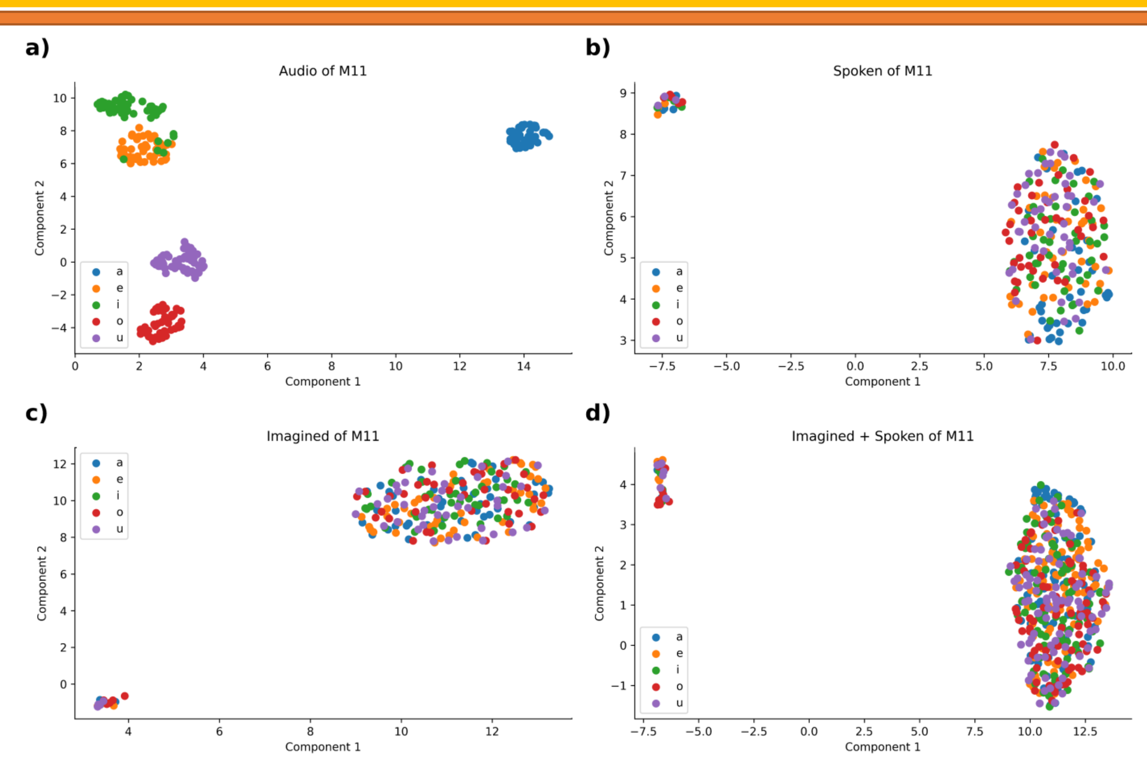

Figure 9. UMAP Visualation of Patient M11 for vowels.

- Spoken EEG data shows some separation between vowels but may be less distinct than MFCC due to overall neural activity.
- Imagined EEG data exhibits even less defined clusters compared to spoken vowels due to weaker brain activity.
- Although we cannot see the grouping in a 2d plane as it is hard to separate the neural activity using simple models, we will use deep learning based models to decode the activity into their repective vowels and words etc

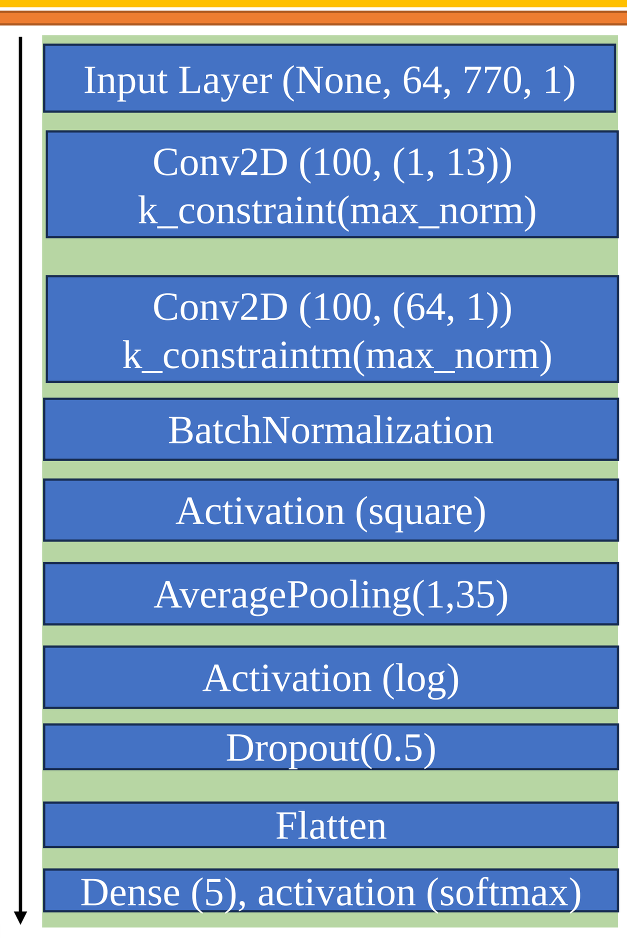

Figure 10. DL based architecture for VCV Decoding

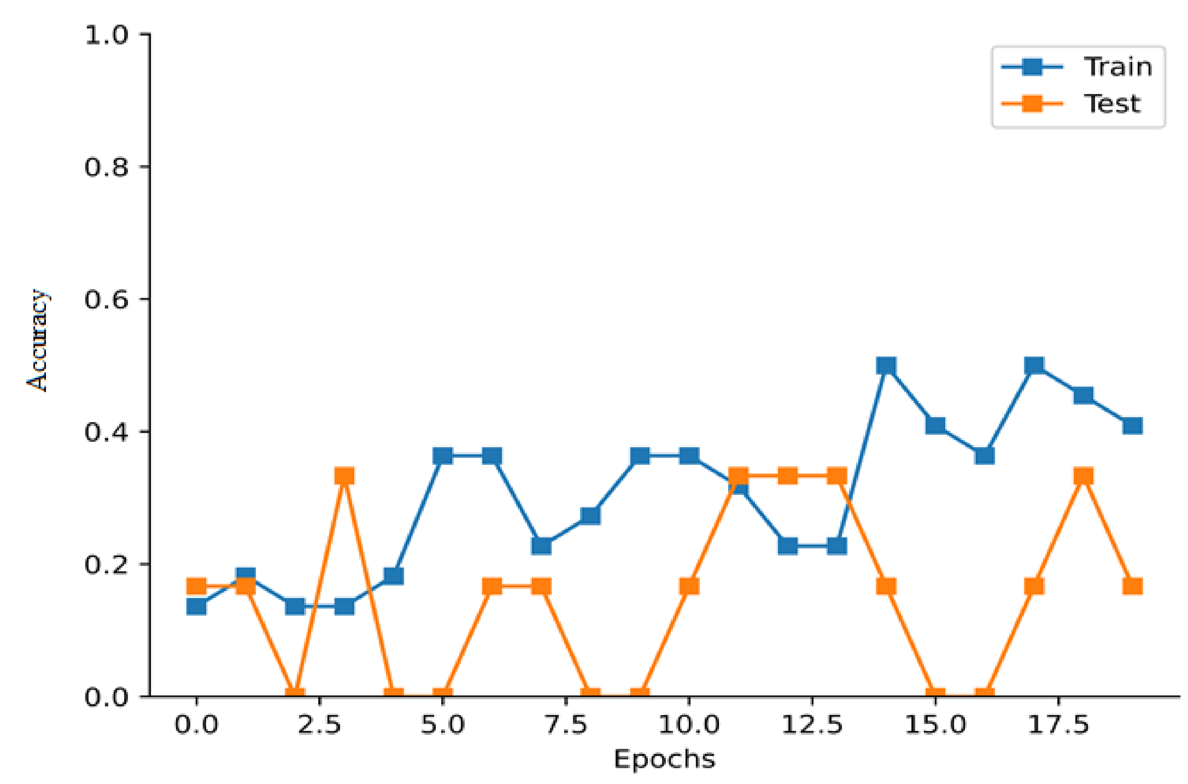

Figure 11. Performance of ShallowCNN model on 5 CVC task decoding

- The performance is low due to less data available at the moment.
- The model was trained on 4 samples from each category and tested on 2 samples from each category

Acknowledgements

Grant PID2022-141378OB-C22 (PI Jose A. Gonzalez-Lopez) funded by MCIN/AEI/10.13039/501100011033 and, as appropriate, by "ERDF A way of making Europe", by "ERDF/EU", by the "European Union" or by the "European Union NextGenerationEU/PRTR

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN

UNIVERSIDAD DE GRANADA

Correspondence email: joseangl@ugr.es