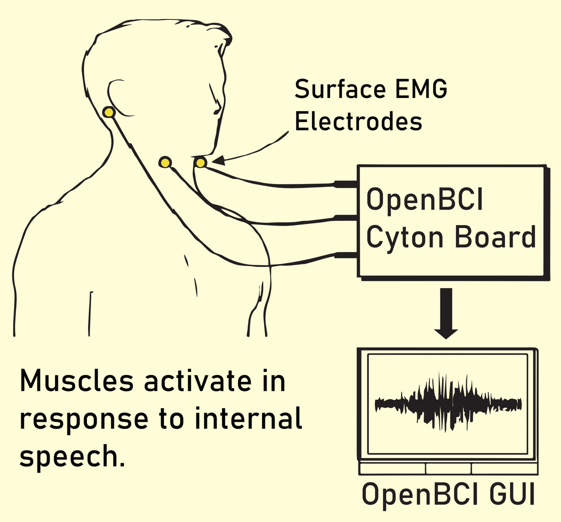
Understanding Subvocal Speech with ML Algorithms

Nivriti Bopparaju, Christopher Bjordahl, Ben Cooper, Dennis Koshta, Rico Passanisi-Rodriguez, Mathew Sarti
University of California Santa Cruz, CA 95064

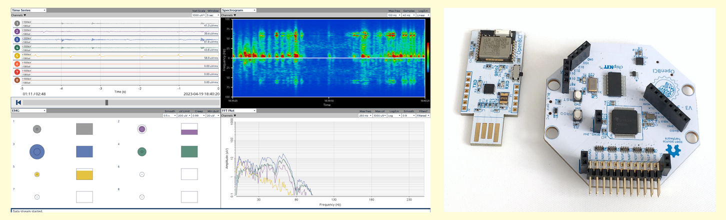
Abstract

Subvocalization refers to the internal speech that occurs while reading or thinking without producing any audible sound. It is accompanied by the activation of the muscles involved in speech production, generating electrical signals known as electromyograms (EMGs). These signals can potentially be used for silent communication and assistive technologies, improving the lives of people with speech impairments and enabling new forms of human-computer interaction. However, the accurate recognition of subvocal EMGs remains a challenge due to the variability in signal patterns and electrode placement. Our proposed solution is to develop an advanced subvocal EMG recognition system using machine learning techniques. By analyzing the EMG signals, our system will be able to identify the intended speech content and convert it into text. This technology will be applicable in various fields, including silent communication for military or emergency personnel, assistive devices for people with speech impairments, and hands-free control of computers and other electronic devices.

Background



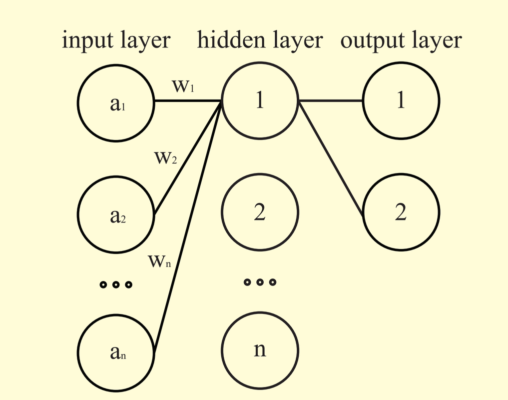
Subvocal Speech can be observed and processed with the use of open source hardware and software such as the OpenBCI Cyton Board & GUI.



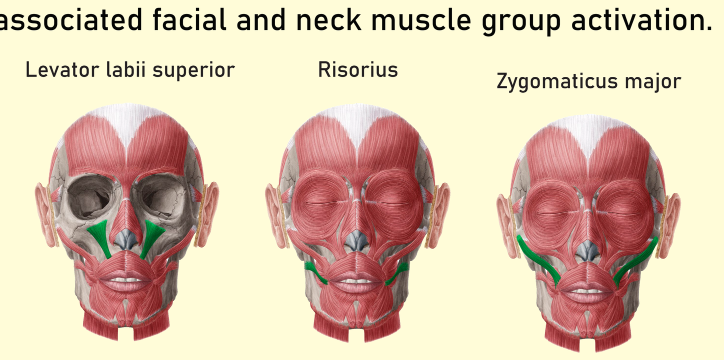
Muscles activate in response to internal speech.

These technologies can record muscle activation and export this processed data as a csv file or allow a data stream for a variety of applications.

Neural Network



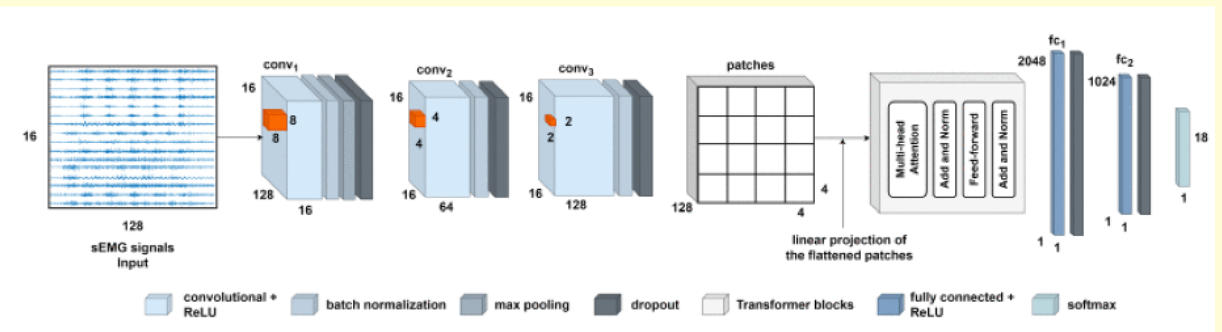
Phonemes are the perceptually distinct units of sound in a specified language. For the english language there are 44 unique phonemes each with an associated facial and neck muscle group activation.



Large data sets allow for the machine learning technique known as artficial neural network algorithms.

For example, the labiodental fricative phonemes /f/ & /v/ require contraction of the above facial muscle groups.

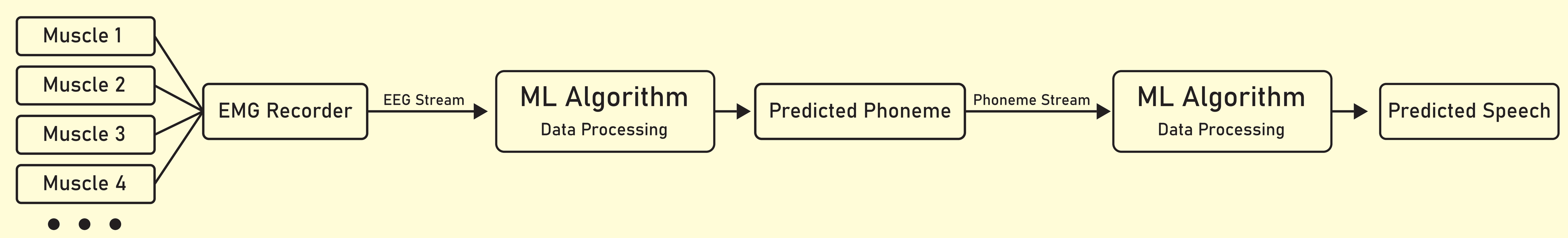
TMC-ViT Architecture



The Temporal Multi-Channel Vision Transformer architecture for raw sEMG channels. Conv stands for convolutional, fc for fully-connected, Norm for normalization, and Add for residual connection. The TMC-ViT is trained with processed sEMG data, and after the convolution blocks, the data is passed to the ViT. In this way, a model that makes robust and accurate predictions of phonemes based on sEMG signals is possible.

Approach

Overview:

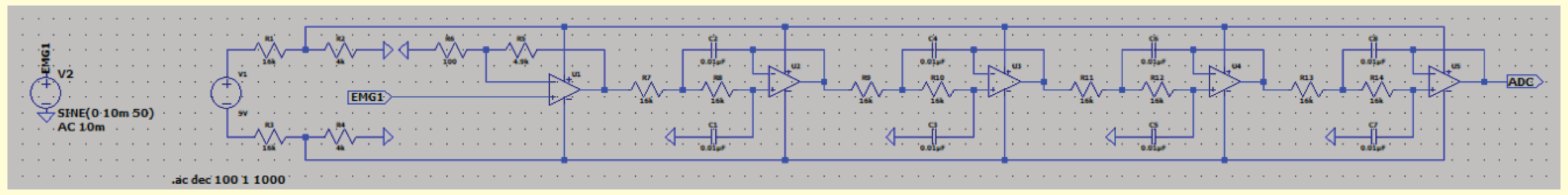


Our end goal is to have a speech prediction machine which acts essentially as a subvocal recognition device.

First we must train our EMG to phoneme ML Algorithm. To do this we must compile a large enough data set to train a neural network to accurately predict given the same muscle group readings. Online data sets are also available to test and strengthen our experimental results.

Secondly, produce a phoneme to speech ML Algorithm. This can be done with online data sets along with our experimental phoneme streams.

Hardware:

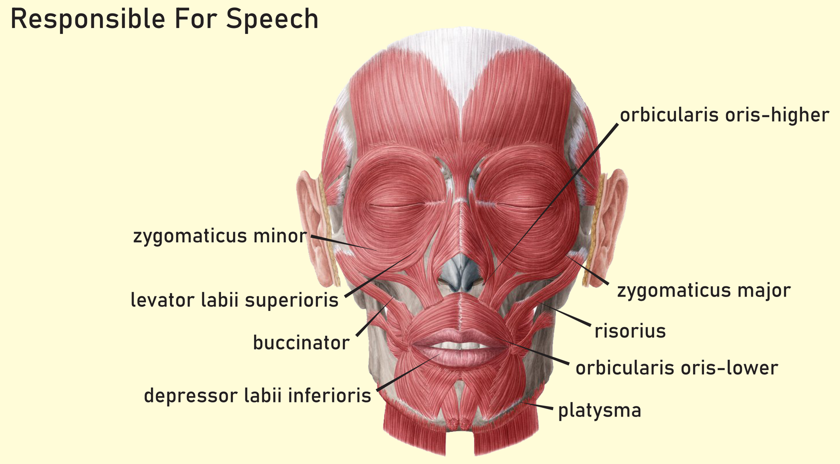


Prototype of the analog sEMG amplification and filtration stage using an active butterworth filter topology for pre-FPGA signal processing

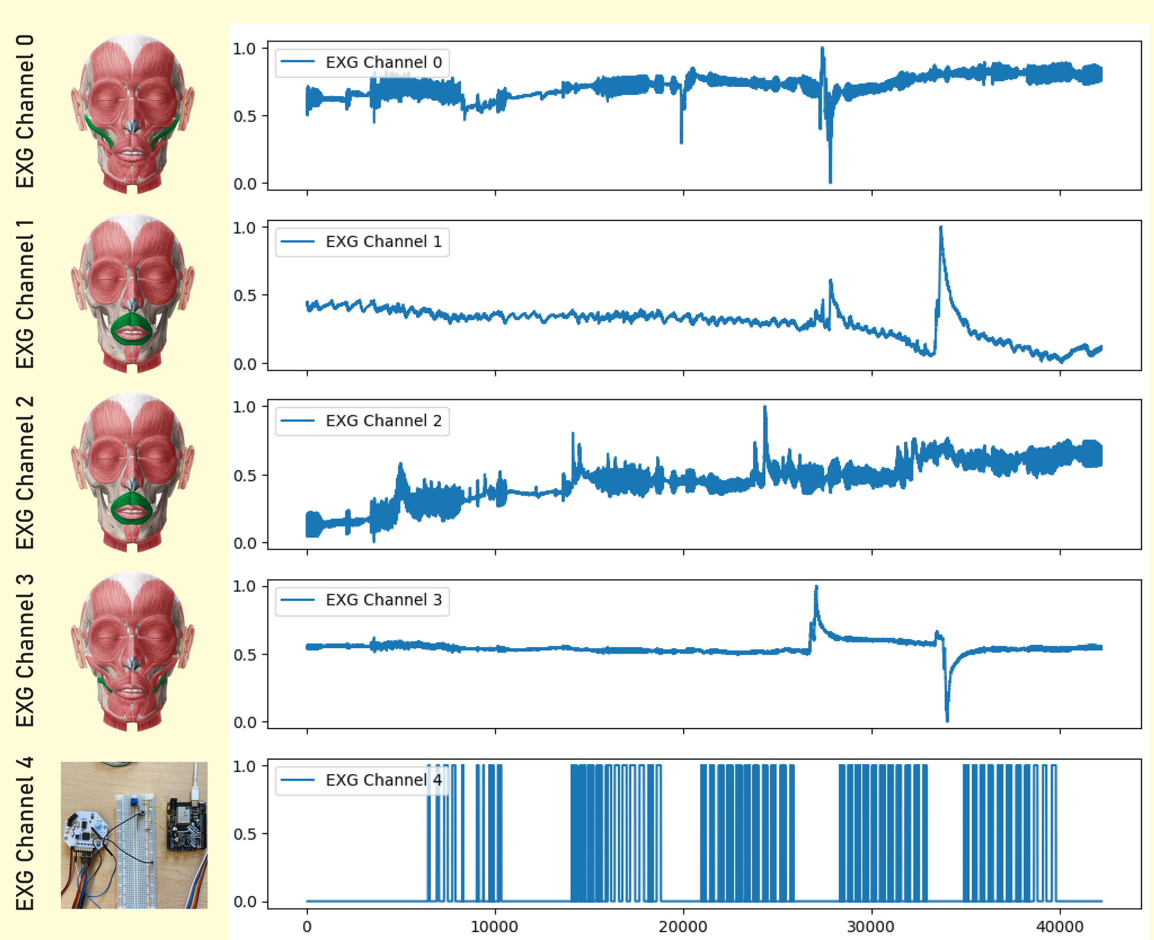
Data Collection:

Trial Type	b Boat	v Video	i: shEEp	u: shOOt	uə tOUrist
Audible	b_aud	v_aud	i_aud	u_aud	ue_aud
Whisper	b_whi	v_whi	i_whi	u_whi	ue_whi
Silent	b_sil	v_sil	i_sil	u_sil	ue_sil

Part of the Facial Muscles Responsible For Speech



Normalized EXG Channels



EMG data collected from multiple subjects as audible, whisper, and silent under distinct phonemes. EMG signals will be recorded using surface electrodes placed on skin over the speech-related muscles. With OpenBCI Cyton Board & GUI at amplification factor of 1170, 16 bits A/D conversion, resolution of 0.033 microvolts per bit, and a frequency range of 0.9-295 Hz. EMG sampled at 600 Hz. Recorded in a push-to-talk setting (channel 4) in quiet rooms, without electrical shielding.

Machine Learning & Data Processing:

To preprocess the .csv files containing raw sEMG data, we used Pandas to manipulate and visualize the data to make useful inferences about it. We first clean the EXG Channel 4 by replacing outliers with respective max/mins. Then we segment the phonemes and individual recordings based on channel 4's activity. We plan to split this data into training and test sets, then train a TMC-ViT model using TensorFlow and Keras to predict the five different phonemes.

Future Directions:

We plan on expanding the phonemes used from the current set to one that is more diverse. Additionally, we intend on integrating our subvocal EMG recognition system into suitable applications, such as live transcription or voice synthesis, and demonstrate its potential for a variety of use cases. We will use containerization and a microservice approach to design these tools.

References

Wand, M., Janke, M., Schultz, T. (2014) The EMG-UKA corpus for electromyographic speech processing. Proc. Interspeech 2014, 1593-1597, doi: 10.21437/Interspeech.2014-379

Epstein, Melissa, et al. Dissection of the Speech Production Mechanism. University of California, 2002. Appendix B: Muscles of the Speech Production Mechanism

Godoy, R. V., Lahr, G. J., Dwivedi, A., Reis, T. J., Polegato, P. H., Becker, M., Caurin, G. A., & Liarokapis, M. (2022). Electromyography-based, robust hand motion classification employing temporal multi-channel vision transformers. 2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob). https://doi.org/10.1109/biorob52689.2022.9925307

Acknowledgements

Thank you to the members of our club that have persevered through this challenging year. We applaud your patience with our newly found cohort and congratulate you on your victories in the past and the many more to come. Additionally, we would like to thank Chris Toukmaji, former president, co-founder, and current project advisor for all of his guidance to NeurotechSC.

UNIVERSITY OF CALIFORNIA SANTA CRUZ