# WEARABLE SILENT SPEECH INTERFACE FOR AUGMENTED REALITY APPLICATIONS

**Riley Flanagan, Tania Rivas**

Advisor: Dr. Christine Walck

*Department of Mechanical Engineering, Embry-Riddle Aeronautical University*

## BACKGROUND

With the advent of heads up displays (HUD) and AR systems in next generation spacesuits and military equipment like the Integrated Visual Augmentation System, no current effective means of controlling display elements has emerged. Existing input solutions like acoustic speech recognition, physical gestures, and buttons are unlikely to resolve users' issues of privacy, difficulty of movement, and unintentional input. To meet these demands, new seamless input models need to emerge to fully utilize the capabilities presented by the new technology while maintaining a low profile. The team posits subvocal input as the answer. Use of subvocalizations for input as a concept has been previously demonstrated with success. Research at NASA JPL successfully demonstrated effectiveness of subvocal input for prototype lunar rover control [1], and the Massachusetts Institute of Technology have demonstrated word recognition accuracy upwards of 90% using similar techniques for controlling household devices [2]. Silent speech devices have yet to see utilization as a prime means of communicating with HUDs. Development of a silent speech interface as a hand-free input scheme for HUD communication is a crucial step in realizing the interfaces necessitated by future use cases.

## METHODOLOGY

Four Delsys Avanti Mini electromyography (EMG) sensors are applied to the areas of interest as shown in Figure 1. Each sensor measures muscle activation signal of a specific muscle (Table 1). For gathering training data, words are subvocalized when hearing an audio cue. The model is trained on a set of words from a command library, including: "Up", "Down", "Left", "Right", "Yes", and "No". The entire word library is subvocalized per trial for forty trials. In the post-processing phase, data from trials is parsed for word components and compiled into a full speech recognition training dataset using Python. This data is used for training a Convolutional Neural Network as a classifier that makes up the silent speech recognition model. The trained speech recognition model will be used as a live input mechanism for an interface constructed on the Microsoft HoloLens.

| Sensor Number | Targeted Muscle |
|---|---|
| 1 | Digastric |
| 2 | Stylohyoid |
| 4 | Sternohyoid |
| 5 | Cricothyroid |

*Table 1: Targeted muscles and sensor numbers*

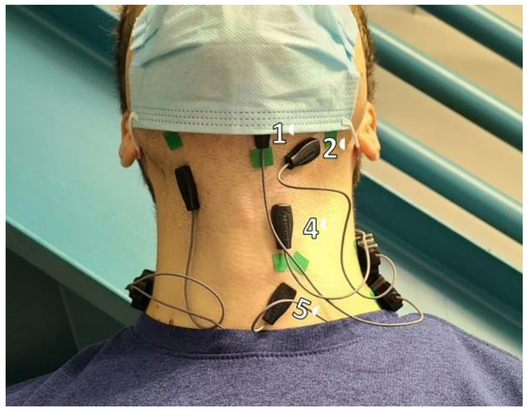


*Figure 1: Sensor placements with corresponding numbers*

## ABSTRACT

Adoption of Augmented and Virtual Reality (AR and VR) interfaces in the aerospace and defense fields has been inhibited by conspicuous and cumbersome input mechanisms like gestures and spoken voice recognition. Silent speech interfaces using non-invasive EMG electrodes are posited as a means for controlling AR and VR interfaces with potential for inconspicuous and high bandwidth input. The objective of the team is to develop a silent speech interface that receives input from subvocalizations via skin surface EMG electrodes and decodes this input into commands to interact with a heads-up-display or similar AR system. Collected EMG sensor data from the neck surface is used to train a convolutional neural network that functions as a classifier to determine the subject's subvocal input against a word library. The user will equip the wearable interface and use it to silently send commands through subvocalizations to control an Augmented Reality device. Word recognition accuracies in trials using the current command library demonstrate effectiveness of the model. Future work includes expanding the dataset used to train the recognition model and live demonstration in controlling an augmented reality interface.

## RESULTS

Collected data from subvocalization trials is band-pass filtered at 2-200 Hz to remove artefacts such as the heartbeat and environmental noise. Individual subvocalized words are extracted from EMG signal data and compiled into separate training and testing datasets. The compiled training dataset is used to train a 1-D Convolutional Neural Network built with the Keras library in Python. The model is 7 layers deep, with max pooling and dropout layers to reduce overfit.

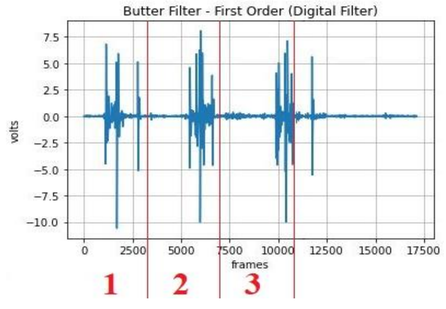


*Figure 2: Filtered signals from three subvocalizations of the word "Right"*

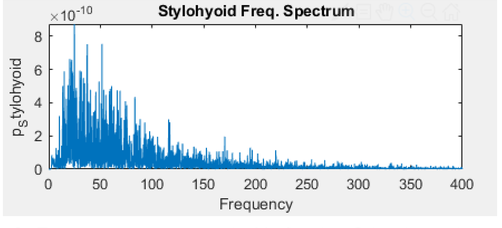


*Figure 3: Frequency spectrum in Hz for the Cricothyroid muscle in subvocalized trial*

## DISCUSSION

Ten trained models were tested on a four-trial testing dataset, with classification accuracies for each listed in Table 2.

| Trained Model | Accuracy |
|---|---|
| 1 | 66.66% |
| 2 | 83.33% |
| 3 | 91.66% |
| 4 | 83.33% |
| 5 | 91.66% |
| 6 | 83.33% |
| 7 | 83.33% |
| 8 | 100.00% |
| 9 | 66.66% |
| 10 | 75.0% |
| MEAN | 82.5% |

*Table 2: Trained models and classification accuracies when tested against the test dataset.*

Successful classification is shown in Table 2, with one trained model showing a testing accuracy of 100%, which offset by models testing at a significantly lower 66.66%. An average word classification accuracy of 82.5% between all models is observed. While all models were trained and tested on the same datasets, the stochastic nature of the model has significant effects on output, with the dropout layer adding artificial noise to training, and the gradient-descent based optimization algorithm adding random variance to the completed models. Further concerns in accuracy values include potential for overfit from the small dataset (12 trials) used for these preliminary assessments. Some values, such as trials 2, 6, and 7 show repeated values in accuracy, which is likely another result of the smaller dataset.

In future work, the team hopes to expand the training and testing datasets to over 400 trials from a larger subject pool. A diversity in subjects used for training will allow user-independent use of the interface, instead of being limited to the single user it is currently trained on. In addition, using more trials should help to reduce the overall variance in accuracy shown across many tests.

## CONCLUSIONS

More trial data will be necessary to improve the accuracy and usability of the interface, and collecting that data from a large subject pool represents the first step in improving classification accuracies. While the current interface functions as a proof-of-concept for classifying previously gathered input in a controlled setting, the team is looking to expand the functionality of the interface to classify input in a live setting. After live input is demonstrated, integration to control a Microsoft HoloLens represents the next challenge in demonstrating feasibility of subvocal input for AR applications.

## REFERENCES

[1] J. Bluck, "NASA - NASA Develops System To Computerize Silent, "Subvocal Speech"", *Nasa.gov*, 2020. [Online]. Available:https://www.nasa.gov/home/hqnews/2004/mar/HQ_04093_subvocal_speech.html. [Accessed: 02- Mar- 2020].

[2] A. Kapur, "AlterEgo | 23rd International Conference on Intelligent User Interfaces", *Dl.acm.org*, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3172944.3172977. [Accessed: 02- Mar- 2020].