iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VI **Month of publication:** June 2019

DOI: http://doi.org/10.22214/ijraset.2019.6338

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

Design and Development of Silent Speech Recognition System for Monitoring of Devices

Mr. Dilip R[1], Dr. Ramesh K. B[2]
[1]Assistant Professor, Department of Mechatronics Engineering, Acharya Institute of Technology, Bengaluru, Karnataka India
[2]Associate Professor, Department of Instrumentation Technology, RV College of Engineering, Bengaluru, Karnataka India

Abstract: This project indicates the method of monitoring EMG-based Silent Speech Interfaces include robust, confidential, non-disturbing speech recognition for human-machine interfaces and transmission of articulatory parameters and Our approach to capture silent speech relies on surface Electromyography (EMG), which is the process of recording electrical muscle activity using surface electrodes for example by a mobile telephone for silent human-human communication
Keyswords: MATLAB, Arduino, EMG, Feature SSI.

I. INTRODUCTION

Humans Automatic Speech Recognition (ASR) has now reached a level of precision and robustness which allows its use in a variety of practical applications. Notwithstanding, speech recognition suffers from several drawbacks which arise from the fact that ordinary speech is required to be clearly audible and cannot easily be masked: on the one hand, recognition performance degrades significantly in the presence of noise. On the other hand, confidential and private communication in public places is difficult if not impossible. Even when privacy is not an issue, audible speech communication in public places frequently disturbs bystanders. Both of these challenges may be alleviated by Silent Speech Interfaces (SSI). A Silent Speech Interface is a system enabling speech communication to take place without the necessity of emitting an audible acoustic signal, or when an acoustic signal is unavailable. Since speech is produced by the activity of the human articulatory muscles, the EMG signal measured in a person's face may be used to retrace the corresponding speech, even when this speech is produced silently, i.e., articulated without any vocal effort.

II. OBJECTIVE

To design and develop a system that is used to silent speech recognition system for monitoring of devices using the action of the jaw movement of the human body for controlling external devices using wireless technology

III. PROBLEM WITH EXISTING SYSTEM

Previous EMG-based speech recognition systems were usually limited to very small tasks and vocabularies. The main reason for this limitation was that those systems were usually session-dependent, i.e., they used training and test data from only one speaker and only one recording session.

IV. PROBLEM DEFINITION

The term recording session means that during the recording, the EMG electrodes were not removed or reattached. We show that a system trained on multiple recording sessions of one and the same speaker yields a reasonable performance and that a session-independent system recognizes test data from unseen sessions more robustly than a similarly large recognizer trained on data from just one session. We additionally prove that the increased robustness of a session- independent system also helps to cope with the difference between normal and silently articulated speech. Finally, we investigate how the system copes within creasing recognition vocabulary sizes and present results on an EMG-based speech recognition system with a vocabulary of more than 2000 words, which to the best of our knowledge is the largest vocabulary which has ever been used for recognizing speech based on EMG signals.

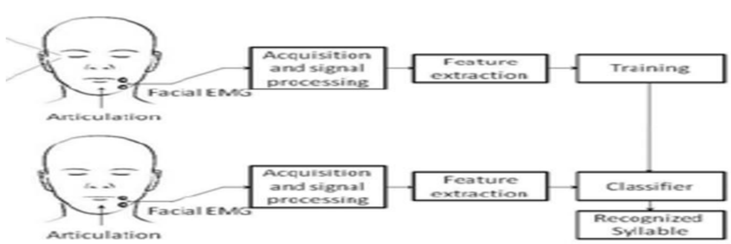

Figure 1: Representation of Basic Block Diagram

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

V. DIFFERENT TECHNOLOGIES OF SPEECH COMMUNICATIONS

Speech recognition system can be separated in different classes by describing what type of utterances they can recognize.

A. Isolated Word

Isolated word recognizes attain usually require each utterance to have quiet on both sides of sample windows. It accepts single words or single utterances at a time. This is having the “Listen and Non-Listen state”. Isolated utterance might be a better name of this class.

B. Connected Word

The connected word system is similar to isolated words but allows separate utterance to be run together a minimum pause between them.

C. Continuous Speech

Continuous speech recognizers allow the user to speak almost naturally, while the computer determines the content. Recognizer with continues speech capabilities is some of the most difficult to create because they utilize a special method to determine utterance boundaries.

D. Spontaneous Speech

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed ASR System with spontaneous speech ability should be able to handle a variety of natural speech feature such as words being run together.

VI.ARCHITECTURE OF SYSTEM

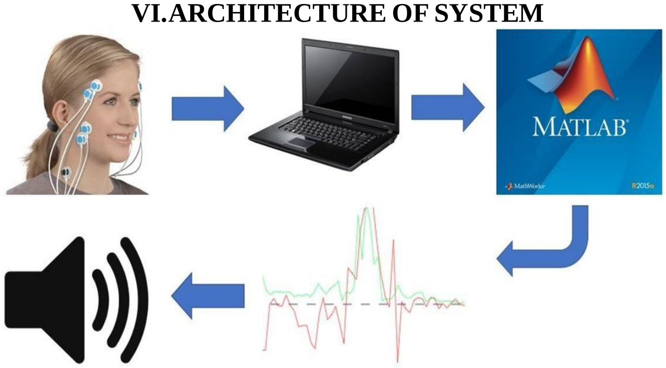

Fig 2: Block Diagram Of Slient Speech System

The Arduino UNO is an open-source microcontroller board based on the Microchip ATmega 328P microcontroller and developed by Arduino.cc. The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards(shields) and other circuits. The board has 14 Digital pins, 6 Analog pins, and programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable. The UNO board is the first in a series of USB Arduino boards, and the reference model for the Arduino platform. The ATmega 328 on the Arduino Uno comes preprogrammed with a bootloader that allows uploading new code to it without the use of an external hardware programmer.
The AD 8232 Single Lead Heart Rate Monitor is a cost-effective board used to measure the electrical activity of the heart. The AD8232 is designed to extract, amplify, and filter small bio potential signals in the presence of noisy conditions, such as those created by motion or remote electrode placement. This electrical activity can be charted as an ECG or Electrocardiogram and output as an analog reading. ECGs can be extremely noisy, the AD8232 Single Lead Heart Rate Monitor acts as an op amp to help obtain a clear signal from the PR and QT Intervals easily.
The AD8232 is an integrated signal conditioning block for ECG and other bio potential measurement applications. It is designed to extract, amplify, and filter small bio potential signals in the presence of noisy conditions, such as those created by motion or remote electrode placement. Surface EMG can be recorded by a pair of electrodes or by a more complex array of multiple electrodes. More than one electrode is needed because EMG recordings display the potential difference (voltage difference) between two separate electrodes. The lower motor neurons are the actual instigators of muscle movement, as they innervate the muscle directly at the neuromuscular junction. This innervations causes the release of Calcium ions within the muscle, ultimately creating a mechanical change in the tension of the muscle. Limitations of this approach are the fact that surface electrode recordings are restricted to

superficial muscles, are influenced by the depth of the subcutaneous tissue at the site of the recording which can be highly variable depending of the weight of a patient, and cannot reliably discriminate between the discharges of adjacent muscles. Surface EMG is used in a number of settings; for example, in the physiotherapy clinic, muscle activation is monitored using surface EMG and patients have an auditory or visual stimulus to help them know when they are activating the muscle(biofeedback).

VII. VII DESIGN AND IMPLEMENTATION

Fig. 3 special design mask with muscle/emg sensor

A specially designed face mask is used to place the electrodes on the user's face. It has holes at specific locations through which electrodes can be attached. The images showed above are 3D renders of the mask that we have specifically designed to be used along with this device. The locations of the electrode placements on the mask are carefully adjusted to give the most stable and best EMG signals. This mask will help in keeping the electrodes properly attached to the skin and also receive a stable signal from the electrodes.

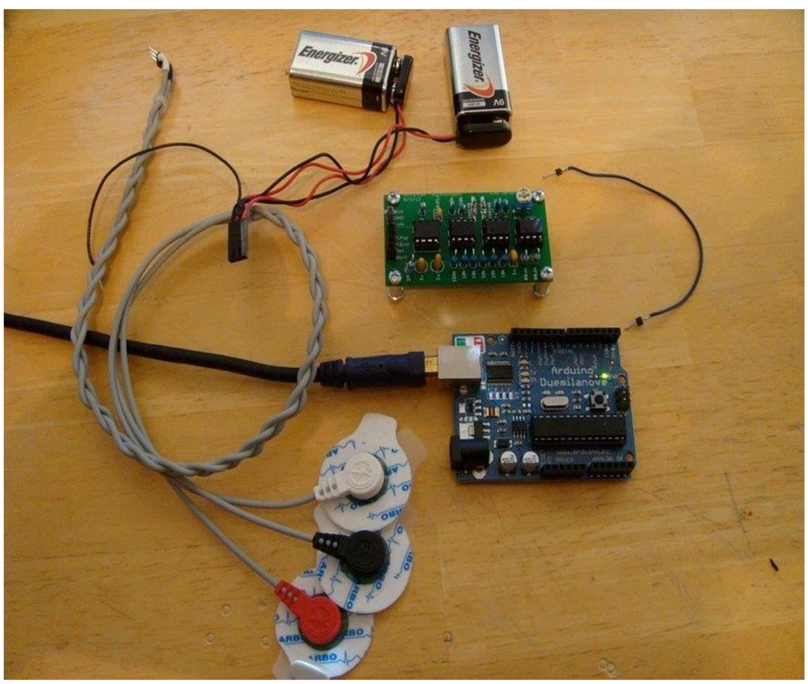

Fig. 4 Muscle/EMG sensor

Measuring muscle activation via electric potential, referred to as electromyography (EMG), has traditionally been used for medical research and diagnosis of neuromuscular disorders. However, with the advent of ever shrinking yet more powerful microcontrollers and integrated circuits, EMG circuits and sensors have found their way into prosthetics, robotics and other control systems.

VIII. WORKING OPERATION

A. Connect the Power Supply (two 9V Batteries)

Connect the positive terminal of the first 9V battery to the +Vs pin on your sensor.

Connect the negative terminal of the first 9V battery to the positive terminal of the second 9V battery. Then connect to the GND pin on your sensor.

Connect the negative terminal of the second 9V battery to the –Vs pin of your sensor.

B. Connect the Electrodes

After determining which muscle group you want to target (e.g. bicep, forearm, calf), clean the skin thoroughly.

Place one electrode in the middle of the muscle body, connect this electrode to the RED Cable's snap connector.

Place a second electrode at one end of the muscle body, connect this electrode to the Blue Cable's snap connector. Place a third electrode on a bony or non-muscular part of your body near the targeted muscle, connect this electrode to the Black Cable's snap connector.

1) Connect to a Microcontroller (e.g. Arduino)
2) Connect the SIG pin of your sensor to an analog pin on the Arduino (e.g. A0)
3) Connect the GND pin of your sensor to a GND pin on the Arduino.

Electrical Specifications

Parameter	Min	TYP	Max
Power Supply Voltage (Vs)	±3.5V	±5V	±18V
Gain Setting, Gain = 207*(X /1 kΩ)	0.01 Ω (0.002x)	50 kΩ (10,350x)	100 kΩ (20,700x)
Output Signal Voltage (Rectified & Smoothed)	0V	--	+Vs
Differential Input Voltage	0 mV	2-5mV	+Vs/Gain

Fig 5: Electrical Specifications

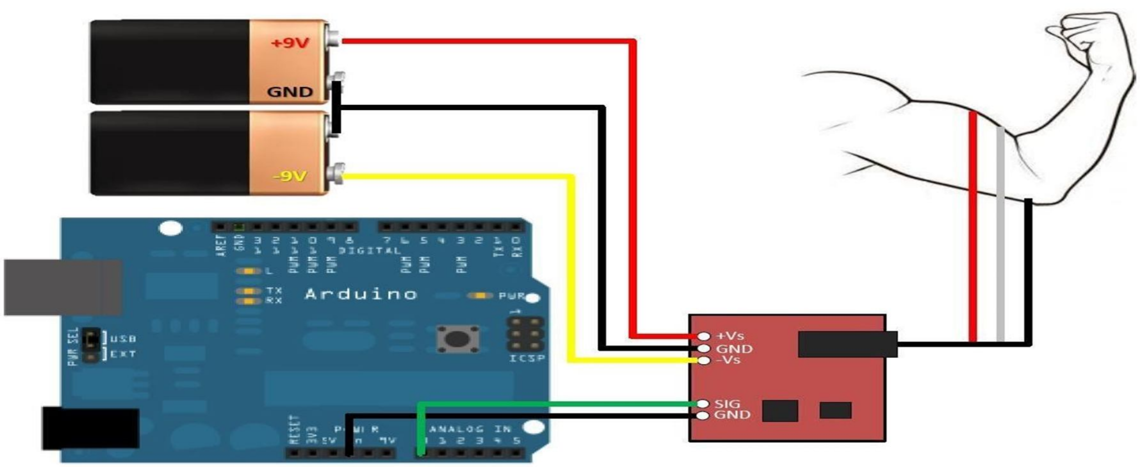

Figure 6: Overall Working Operation

IX. RESULTS AND DISCUSSION

A. Hidden Markov Models

The one main problem that we can solve using HMMs.:

Given an observation sequence O and the dimensions N and M, find the model λ = (A, B, π) that maximizes the probability of O. This can be viewed as training a model to best fit the observed data. Alternatively, we can view this as a (discrete) hill climb on the parameter space represented by A, B and π. Consider the problem of speech recognition (which just happens to be one of the best-known applications of HMMs). We can use the solution of the Problem to train an HMM, say, λ0 to recognize the spoken word "no" and train another HMM, say, λ1 to recognize the spoken word "yes". Then given an unknown spoken word.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

B. Graphical User Interface

MATLAB has a very useful and powerful tool called App Designer. App Designer is a rich development environment that provides layout and code views, a fully integrated version of the MATLAB editor, and a large set of interactive components. Two different apps are created to be interfaced with the device.

One app is to record the facial muscle signals and the other one is to use the device in real time which will compare the live signal with previously recorded ones using our algorithms and give out the outputs both in the form of text and as well as audio output from the PC.

C. Signal Recording App

The recording app has an input field which is used to type the word that the user would speak and another number input field to type the number of trials for each word. After the 'START' button is clicked, the user will speak the word and the signals received from the Arduino that are captured using the EMG sensor and the electrodes stuck to the user's facial muscles. This signal is then plotted on the graph that is displayed on the recording app. This process is repeated until the specified number of trials is complete. All the values of these signals are stored in an array until the number of trials are finished.

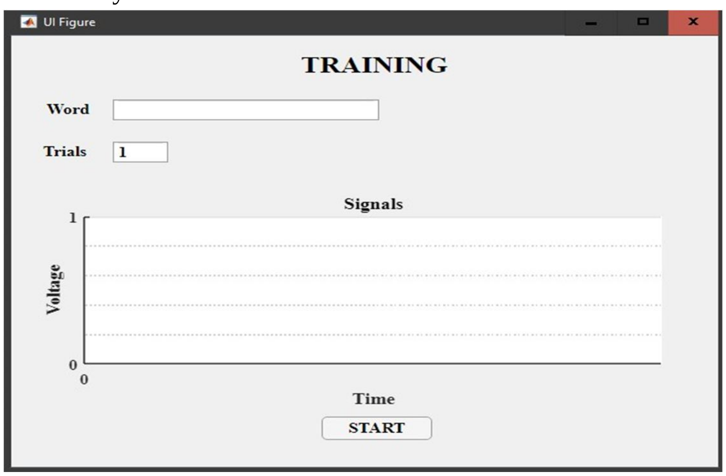

Fig. 6 Signal Recording App

Once all the trials are finished, the values in the array are stored into an Excel file in which each column would describe a separate trial. The same process is repeated for various words. This is how the Signal Recording App works.

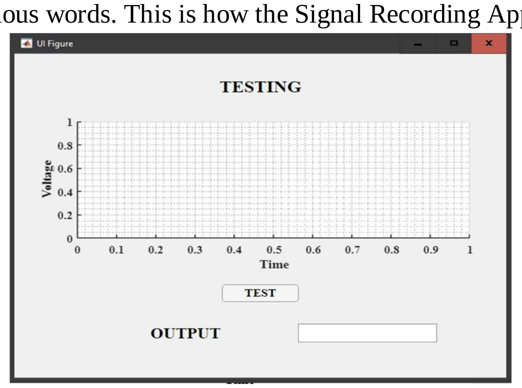

Fig.7 Signal Testing App

The Testing App is specifically designed for the users of this device to be able to use it in real time and get the proper output with the highest possible accuracy. The Testing Application has a button named 'Test'. When this button is clicked, the system will record the facial muscle movements in real time with minimal possible lag from the electrodes attached to the user's skin. This signal is then plotted on the graph found on the Testing App. The app then compares the signal that is received from the sensors with the signals that are already stored in the database and provide the output using a method that will be discussed in the next section.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

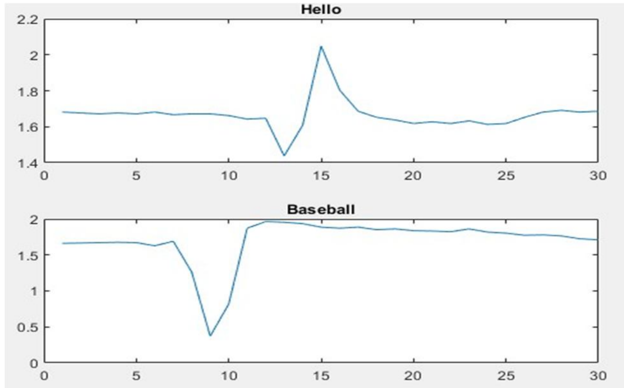

Fig. 8 Utterances of the words ‘Hello’ and ‘Baseball’

The above image shows the two different utterances of the words ‘Hello’ and ‘Baseball’. These two signals can be visually interpreted as different but we need to programmatically infer that these two signals are indeed different and the words uttered are different. Dynamic Time Warping function will find the best optimal path between two signals. This is shown by a diagonal line. The lesser the deviation from the diagonal, the higher is the similarity. By finding this distance, we can calculate the similarity between signals and we set a threshold for the deviation from the diagonal. If the deviation is less than the threshold, the two signals are said to be similar.

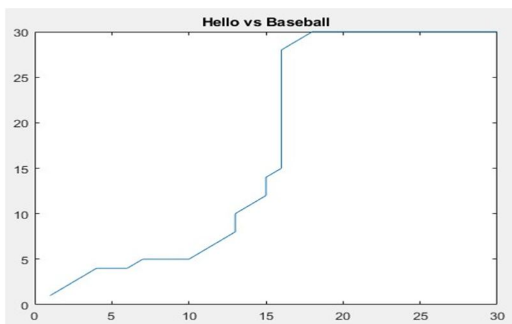

Fig. 9 The warping path for the words ‘Hello’ and ‘Baseball’

The above image shows the warping path of words ‘Hello’ and ‘Baseball’. The warping path of these two words as shown in the graph is very far from the diagonal. This means that the two signals are not of the same word and hence, the algorithm will predict that these two are not the same signal. We then compared two different signals of the same word and the results of this are as follows.

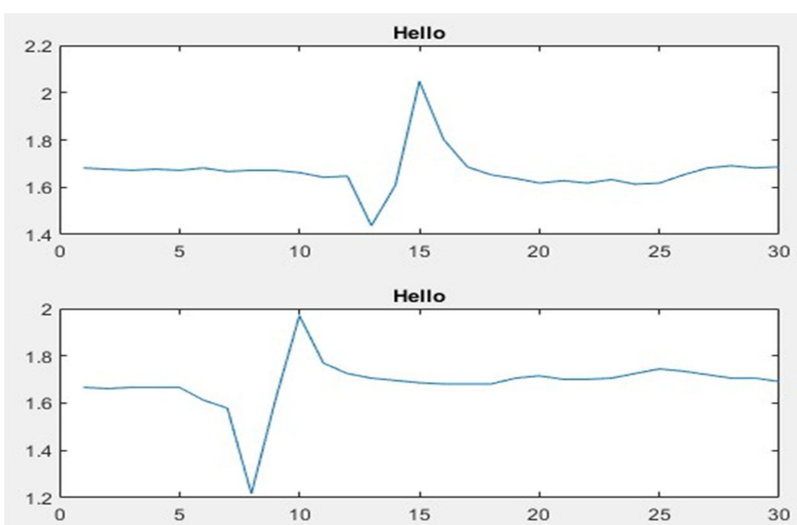

Fig.10 Two different utterances of the word ‘Hello’

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

The above image shows the two different utterances of the word 'Hello'. These two signals are similar visually, but we need to find the similarity between these signals programmatically. Dynamic Time Warping function will find the best optimal path between two signals. This is shown by a diagonal line. The lesser the deviation from the diagonal, the higher is the similarity. By finding this distance, we can calculate the similarity between signals and we set a threshold for the deviation from the diagonal. If the deviation is less than the threshold, the two signals are said to be similar.

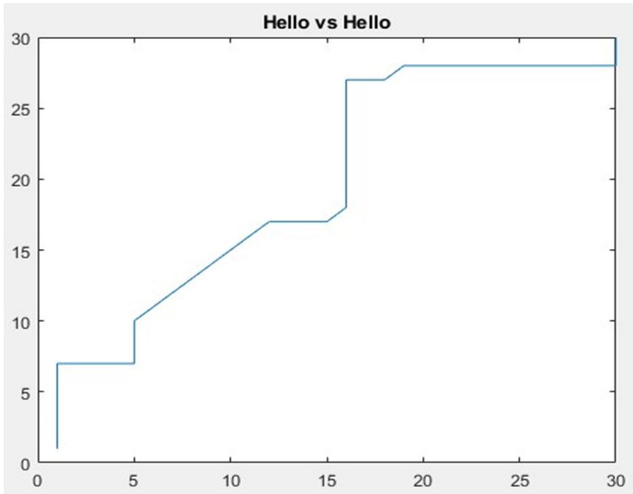

Fig. 10 The warping path of two different utterances of the word 'Hello'

The above image shows the warping path of two different utterances of the word 'Hello'. Although the diagonal is not a perfectly straight line, it is still below the threshold. Hence, it is found to be similar to each other and that both the signals are of the same word. The output is then printed on the screen showing which word was spoken by the user.

X. CONCLUSION AND FURTURE WORK

This report briefly describes the process of designing and implementing a silent speech recognition device. Which is a Human machine interface, enabling the wearer to communicating with a computer. But in our project the computer acts as an intermediary device which assess the signals and converts it into understandable words. We have experimented with various techniques to differentiate words from the signals, such as Dynamic time warping (DTW) and using Neural Networks. DTW has been selected in the current implementation of our project due to the ease of usability. MATLAB with its Arduino support package and signal processing toolbox has eased the process of acquiring the signals and manipulating them via the hardware device. The APP designer toolkit has provided us an easy way to implement a simple Graphical user Interface. More complexity was involved in software rather than hardware. Our current implementation involved using the AD8232 IC module which provides for amplifying and filtering of signals to appropriate levels for the Arduino. Our current implementation had the limitation of using only three electrodes, thus our hardware only provided for distinct signals for a certain range of words. Our purpose for designing a silent speech recognition interface was to aid wheelchair bound persons with speech impediments. Although during the development of this project we had multitudes of success, but there is still a long way to go for this device to be a product that will actually aid the differently abled. We look forward to that day.

There are endless possibilities for improvements but some of which that have come necessary to us during the development phase are as follows. In the future we would be able to add more electrodes. Addition of electrodes would make the device more precision. It would also help make the signals more accurate. We can implement Artificial Neural Networks. This will make it future proof and accuracy of classification of signals will happen. We can add more voices to Text-to-Speech so that it can understand, sense and process various voices of different intensities and pitch. It gets better with use, i.e. it learns from the user and becomes more compatible. We can enable the IoT device which will help us to control several devices at the same time. It helps us save time and also the work gets done in a jiffy. This device will make it easier for people with speech disorder (stammering, lisp, etc.) to control or activate voice search or voice assistant like Google home, Alexa etc.

These are some of the advancements we would like to do on our project in the future. We hope that this device helps improve lives of many people who face problems related to speech.

REFERENCES

[1] Mr. Dilip R , Dr. Ramesh K. B, 2019, design and development of virtual instrumentation system for measurement of human body parameters, international journal of engineering research & technology (ijert) volume 08, issue 06 (june 2019),

[2] Kapur, S. Kapur, and P. Maes, "AlterEgo: A Personalized Wearable Silent Speech Inter-face." 23rd International Conference on Intelligent User Interfaces (IUI 2018), March 5,2018

[3] Michael Wand, and Tanja Schultz “Session-independent EMG-based Speech Recognition.”(2011)

[4] NASA’s Subvocal Speech Recognition Research

[5] EMG Sensor Guide: https://www.sparkfun.com/tutorials/ad8232-heart- ratemonitor-hookup-guide

[6] “G. S. Meltzner, J. T. Heaton, Y. Deng, G. D. Luca, S. H. Roy, and J. C. Kline, “Silent Speech Recognition as an Alternative Communication Device for Persons with Laryngectomy,” IEEE/ACM Transactions on Audio, Speech, and Language Processing

[7] Janke, Matthias, and Lorenz Diener. “EMG-to-Speech: Direct Generation of

[8] Speech from Facial Electromyographic Signals.” IEEE/ACM Transactions on Audio,

[9] Speech, and Language Processing, vol.25,no.12, 2017, pp. 2375 2385.,doi:10.1109/taslp.2017.2738568.

[10] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, Vol. 77, No. 2, February 1989, https://www.cs.sjsu.edu/~stamp/RUA/Rabiner.pdf .