An Augmentative Phonation and Articulation System using Advanced Signal Processing Techniques as an Alternative Communication Device

By

Uzma Shafiq

Master of Sciences in Biomedical Engineering

Fall 20-MSBME-NUST00000330042

Supervised by:

Prof. Dr. Javaid Iqbal

School of Mechanical and Manufacturing Engineering (SMME)

National University of Science and Technology (NUST)

Islamabad, Pakistan

July 2023

An Augmentative Phonation and Articulation System using Advanced Signal Processing Techniques as an Alternative Communication Device

A thesis submitted in partial fulfillment of the requirement for the degree of master's in biomedical Engineering.

By

Uzma Shafiq

Master of Sciences in Biomedical Engineering

Fall 20-MSBME-NUST00000330042

Supervised by:

Prof. Dr. Javaid Iqbal

Signature

School of Mechanical and Manufacturing Engineering (SMME)

National University of Science and Technology (NUST)

Islamabad, Pakistan

July 2023

FORM TH-4

National University of Sciences & Technology

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Regn No.) **Uzma Shafiq (330042)**

Titled: **An Augmentative Phonation and Articulation System using Advanced Signal Processing Techniques as an Alternative Communication Device** be accepted in partial fulfillment of the requirements for the award of **MS Biomedical Engineering** degree.

Examination Committee Members

1. Name: Dr. Asim Waris Signature:____________

2. Name: Dr. Syed Omer Gilani Signature:____________

3. Name: Dr. Adeeb Shehzad Signature:____________

Supervisor's name: Dr. Javaid Iqbal Signature:____________

Date: 25/7/23

Head of Department

25-7-23

Date

COUNTERSINGED

Date: 25-7-23

Dean/Principal

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by **Miss. Uzma Shafiq (Registration No. 330042),** of SMME (School of Mechanical and Manufacturing Engineering) has been vetted by undersigned, found complete in all respects as per NUST Statutes / Regulations, is within the similarity indices limit and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature:______________________

Name of Supervisor: Dr. Javaid Iqbal

Date: 25/7/23

Signature (HOD):______________________

Date: 25-7-23

Signature (Dean/Principal):______________

Date: 25-7-23

Proposed Certificate for Plagiarism

It is certified that MS Thesis Titled **An Augmentative Phonation and Articulation System using Advanced Signal Processing Techniques as an Alternative Communication Device** by **Miss Uzma Shafiq** has been examined by us. We undertake the follows:

a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.

b. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.

c. There is no fabrication of data or results which have been compiled/analyzed.

d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Dr. Javaid Iqbal

Signature : 25/7/23

Dr. Javaid Iqbal
Principal & Dean
School of Mechanical &
Manufacturing Engineering
(SMME) NUST Islamabad

CERTIFICATE OF ORIGINALITY

Declaration

I hereby declare that the research work presented in this thesis has been generated by me because of my own research work. Moreover, none of its contents are plagiarized or submitted for any kind of assessment or higher degree. I have acknowledged and referenced all the main sources of help in this work.

Uzma Shafiq

Fall 2020-MS BME 00000330042

Dedication

I dedicate my thesis to my Late Father and my beloved Mother.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to all those who have supported me throughout this challenging yet fulfilling journey of completing my thesis.

First and foremost, I dedicate this thesis to my late father, whose unwavering love, encouragement, and belief in my abilities continue to inspire me every day. Though he is no longer with us, his guiding presence remains an indelible source of strength and motivation.

To my loving mother, thank you for being my constant pillar of support. Your sacrifices, boundless love, and unending encouragement have been the driving force behind my academic achievements. Your belief in me has been my guiding light, and I am forever grateful for your presence in my life.

I extend my sincere appreciation to my supervisor, Dr. Javaid Iqbal and my co supervisor Dr. Asim Waris, for their invaluable guidance, mentorship, and expertise throughout this research endeavor. Their dedication and encouragement have played a pivotal role in shaping the outcome of this thesis.

Lastly, I am thankful to Scholl of Mechanical and Manufacturing Engineering, NUST for providing me with the resources and opportunities to pursue my research aspirations.

In conclusion, this thesis stands as a testament to the collective support, love, and encouragement I have received from my late father, my loving mother, my supervisor, and co-supervisor, as well as my friends and family. It is with deep appreciation and a sense of achievement that I present this work, which would not have been possible without all of you.

With heartfelt gratitude,

Uzma Shafiq

Contents

LIST OF FIGURES

LIST OF TABLES

Abstract

Speech recognition systems utilize acoustic signals collected using a microphone. But in individuals with speech disorders such as who have undergone laryngectomy or have vocal cord paralysis, it is not possible to collect the acoustic signals. For rehabilitation of such individuals an alternate method of communication has to be devised which is independent of acoustic signals. Facial muscles of such individuals remain intact and can be thus used for speech recognition purposes. The limited research on subvocal voice recognition demands the need to develop robust methods which can aid in rehabilitation of the effected individuals. This study aims at filling the gaps left by the previous literature. The EMG signal filtration is carried out to denoise the signals using ana advanced signals processing technique called Variational mode decomposition (VMD). VMD decomposes the input signals into its sub signals in different frequency spectra. Each frequency spectrum undergoes Iterative interval thresholding (IIT) using SIFT operator. These filtered signals are then used to extract crucial information from the signals using a novel feature extraction technique that utilizes the VMD method along with singular vector decomposition. These extracted features are utilized to classify the isolated words using Random Forest classifier. The results demonstrate the superiority of this technique, achieving an accuracy of 98.6% and 92% for a vocabulary set of 70 words and 96 words respectively. The final objective of this study was to develop a novel speech activity detection method using IMU signals as an alternate to EMG used previously in the literature. The proposed activity detection algorithm is carried out in four stages. The algorithm provides significantly better results as compared to EMG based activity detection method. The mean activation error rate (AER) for IMU based algorithm was 0.138 and for EMG based activity detection method was 0.275. These findings demonstrate the superiority of the proposed feature extraction and speech activity detection methods over alternative techniques. In conclusion, subvocal voice recognition is vital for rehabilitation, enabling silent communication and enhancing independence. Limited research and lack of IMU utilization pose challenges, but the thesis proposes a novel approach integrating spatiotemporal feature extraction and IMU data for improved accuracy and robustness. Results demonstrate the superiority of the proposed methods.

Chapter 1: Introduction

Speech is a critical a part of human verbal exchange, that plays a critical role in each day lifestyles activities. It is the most primary means for expressing our thoughts, thoughts, emotions, and dreams, allowing us to connect with others and shape massive bonds. We communicate records, interact in discussions, try to find aid, and percentage our experiences via speech, in the end converting the fabric of our society. Through speech, we trade data, hold discussions, are seeking assistance, and percentage our experiences, in the end changing the corporation of our society.

Automatic Speech Recognition (ASR) systems are smart gears that turn spoken words into written text without any human involvement. These structures have substantially stepped forward over the years, thanks to the development in techniques like device studying and deep learning. ASR structures are used in lots of different regions, like digital assistants, transcribing services, voice-controlled devices, and more. In this article, we can discover the principal components and techniques utilized in ASR structures, in conjunction with a few critical assets inside the discipline.

ASR structures are now usually utilized in different industries and regions. Their ability to change spoken language into written textual content automatically brings a whole lot of advantages, making it possible to create new applications and improve existing strategies. Automatic speech popularity systems have several critical uses. Virtual assistants like Siri, Alexa, Google Assistant, and Cortana depend on Automatic Speech Recognition (ASR). These smart assistants use a technology known as ASR to recognize and reply to what customers say. This we could users do such things as set reminders, take a look at the climate, make telephone calls, and manipulate clever domestic gadgets. ASR permits you to manipulate gadgets like smart speakers, smart TVs, and smartphones for the usage of just your voice. People can speak to those gadgets to get information, play songs, manipulate matters in their domestic, and do matters without using their arms. ASRs also are very crucial in making customer support interactions automatic. Interactive Voice Response (IVR) systems that use Automatic Speech Recognition (ASR) generation allow customers to apply a menu and complete responsibilities without speaking to a person. ASR enables agencies to recognize what their customers are announcing all through calls via mechanically transcribing and analyzing those conversations for beneficial statistics. ASR generation can assist with translating languages in real-time. By changing spoken language into written text, ASR systems may be combined with an era that knows and interprets language, allowing for instant translation services. The ASR generation allows people with listening to troubles or disabilities to have higher get entry to things. Subtitles and captions created by way of ASR help human beings in accessing audiovisual materials, making them extra inclusive and accessible to a wider variety of people. ASR systems are also commonly used to automatically transcribe audio and video recordings. ASR-powered transcription offerings make it less difficult and quicker to alternate spoken content (like interviews, lectures, and conferences) into written text. It saves time and effort compared to doing it manually.

The significance of speech becomes even more apparent when we don't forget individuals with speech issues. Speech disorders encompass many situations that influence a person's potential to make sounds or speak definitely or fluently. These disorders can happen in a variety of methods, which include fluency, voice problems, and speech problems. They may be present at birth (as with developmental language problems) or obtained later in life via factors inclusive of neurological situations, harm, and trauma.

People with speech disorders have problems that cross past just talking to others. These humans often have problems in social conditions, faculty, and paintings. When human beings are not capable of proportioning their mind and emotions properly, it may lead them to feel annoyed, lonely, now not assured, and restriction their possibilities to study and develop in faculty and of their non-public lives.

But improvements in technology and verbal exchange equipment have made it higher at helping humans with speech troubles. Augmentative communique systems that use strategies like electromyography (EMG) and inertial size gadgets (IMU) have ended up beneficial gear for recognizing and detecting speech. EMG sensors are used to measure the muscular tissues worried in speaking, at the same time as IMUs file movement and alignment records, which makes those systems more accurate and bendy.

These communication systems help human beings who have trouble speaking. They deliver them a way to talk and have conversations. These systems help humans who have problems communicating by way of turning their moves or gestures into speech or textual content. This allows them to interact with others and participate in specific components of existence.

Additionally, augmentative conversation systems have blessings for greater than simply the person that has trouble talking. These systems help all people apprehend and aid people with speech problems, like their own family, buddies, instructors, and healthcare professionals. They also assist people recognize and experience others in society. They assist people to speak better and accept and respect differences.

Chapter 2: Literature Review

Electromyography (EMG) is a valuable technique to measure and record the electrical signals created by muscular tissues in a human body. This helps researchers and docs study how muscle groups operate. It can be used to recognize how muscular tissues move, diagnose muscle problems, song development in rehabilitation, and create devices to assist with movement. EMG measures the electric signals produced with the aid of muscular tissues once they agree. When a muscle tightens up, a few devices within the muscle begin operating and send out electric indicators called action potentials. You can discover these activations with the aid of placing sensors on the surface skin or with the aid of setting small needles into your muscle groups. The recorded EMG sign tells us about how our muscle mass contracts, their time of activation, the way strength of these activations [1].

EMG is commonly used to diagnose musculoskeletal disorders such as muscle dysfunction, arthritis, and muscle tension syndrome. Abnormal electrical signals from nerves can provide important information for early diagnosis and especially treatment of problems [2]. Rehabilitation and physical therapy use EMG to assess muscle activity and monitor how well patients are doing during their treatment. It helps create a personalized exercise program and regulates increased muscle strength and coordination [3]. EMG signals can help people control artificial limbs and other devices that assist with movement [4]. This allows people who have lost limbs or have trouble moving to do different tasks using signals from their muscles.

EMG is used in studying how muscles work during different tasks in ergonomics and biomechanics research [5]. This helps us understand why muscles get tired, find out about risks to our body, and make our work and movement better.

In sports science, EMG is used to study how muscles work during activities like running, jumping, and throwing [5]. This information helps us understand how muscles are used and how to improve performance.

The wide use of EMG in rehabilitation of individuals suggests its use for rehabilitation of individuals with speech disorders. The EMG signals from facial muscles can provide vital information about the words being uttered by the subject. Squamous cell carcinoma is the most common type of cancer after lung cancer in men and breast cancer in women. The frequent use of cigarettes, bidis, betel palm and betel nut are the leading causes of rise in squamous cell carcinoma of mouth and vocal tract in the subcontinent. Cancer results in laryngectomy of individuals in most cases. Each year thousands of individuals undergo laryngectomy leaving them with no way of communication. Moreover, other medical conditions paralyzed vocal cords can also lead to phonological disorders leading to an inability to communicate properly. Thus, we need to devise an alternate method of communication for such individuals. In all the previously mentioned cases the facial muscles remain healthy and intact and are not damaged in any manner. Thus, the facial muscles can be used to devise an alternate communication method.

Surface electromyography (sEMG) is being used to study silent speech and understand how our muscles work when we talk without making any sounds. Subvocal speech means silently saying words and sentences without making any sound. This process includes making the speech muscles work, and sEMG is a way to record and study the electricity in these muscles while someone speaks in their mind.

Surface EMG is used to find and record the electrical signals made by the muscles used in silent speech. By attaching electrodes on certain muscles in the face and throat, like the ones responsible for lip movement, chewing, and voice box control, scientists can track the muscle's movements when a person talks without making any sound.

We examine the sEMG signals from the muscles used for speaking to find important information. These characteristics may include the strength, speed, and timing of the muscle movements. Feature extraction is important for discovering unique patterns connected to various speech sounds and movements [6].

Pattern recognition and decoding is a process where we use algorithms and techniques to figure out and understand patterns in subvocal speech after extracting its features. By converting the muscle signals from the mouth into sounds or written words, the system can make speech that is not spoken out loud turn into written text or make it audible through computer-generated speech [7].

Subvocal speech recognition systems that use Surface EMG technology have great potential in creating communication devices to help people who have difficulty speaking or when they cannot or do not want to use their voice [8]. These devices can help people communicate without making noise, which makes life better and helps them join in social situations.

Brain-Computer Interfaces (BCIs) are devices that can connect our brains with external devices. We can use Surface EMG to help create these direct communication pathways [6]. By analyzing and understanding the signals produced by our muscles when we silently speak in our minds, brain-computer interfaces (BCIs) can enable people to communicate or operate devices without speaking out loud.

A lot of research has been carried out on subvocal speech recognition using sEMG. Jennifer Vojtech classified continuous sentences using 8 channels and achieved an accuracy of 95% [9], but the study failed to report any speech detection algorithm. Anat Ratnovsky classified 7 words using 3 channels and achieved an accuracy of 50% [10]. There are two major drawbacks, first being the low accuracy and second being the scarce vocabulary of isolated words. Kumar et al classified the 5 vowels with an accuracy of 92% [11]. Ki Seung lee classified 60 words using 3 channels and achieved an accuracy of 88.13% [12]. Although the vocabulary is much larger than [10, 11] but the study failed to report any classification or activity detection methods for continuous sentences. Mathias Janke classified 50 continuous words but failed to report any activity detection method [13]. Meltzner classified 65 isolated words with an accuracy of 92.3% and carried out classification of continuous sentences [14]. Moreover [14] also reported a speech activity detection method.

Chapter 3: Objectives and Problem Statement

With all the previously mentioned studies following gaps have been identified:

1. Isolated words vocabulary limited to 65 words. [14] reported the maximum isolated vocabulary set of 65 words and achieved a maximum accuracy of 92.3%.
2. Maximum accuracy achieved for maximum vocabulary set of 65 words was 92.3%.
3. None of the literature reported filtering the data with advanced techniques.
4. The number of channels used for highest reported accuracy was 11, which is computationally very expensive.
5. Speech onset and offset faults can lead to alignment issues and, eventually, recognition mistakes. A purely sEMG-based approach has been used in the past by Geoffrey Meltzner which is insufficient to distinguish between speech and non-speech-related muscle activity.

Previous literature does not report any advanced filtering techniques. Speech signals can be contaminated by noise. Due to their lower amplitude, it can be difficult to separate noise from the speech signal. The precision of an EMG-based application can be harmed by unwanted noise. Different kinds of sounds make it hard to use EMG signals because they mess with the signal's quality. Reducing unnecessary noise from EMG signals is a big problem that needs to be addressed. Regular filters are commonly used to reduce unwanted noise in a signal. Examples of these filters include low pass, high pass, adaptive, Wiener, and Kalman filters [15]. Traditional filters can reduce noise in a signal, but they also remove some of the signal's original qualities if it shares frequencies with the noise.

Dragomiretskiy and his colleagues in 2014, a new method called variational mode decomposition (VMD) was created [16]. It is an improved version of the Wiener filter, and it works by dividing the data into different adaptive bands. VMD is a method that breaks down a signal into its components using a process called signal decomposition. Each VMF that is obtained after demodulation has a distinctive central frequency. To prevent different aspects of a signal from blending in VMD, the various parts of each VMF are determined based on the signal's frequency range. Such denoising has been used effectively for analyzing seismic time-frequency, diagnosing gearbox faults [17], predicting wind speeds [18], diagnosing bearing faults [19], cleaning up biomedical images [20], and removing noise from ECG and EMG signals [21,22]. In this study we will be utilizing VMD combined with three thresholding operators and two thresholding techniques to carry out denoising. The thresholding techniques utilized will be Iterative thresholding (IT) and Iterative interval thresholding (IIT). The efficacy of each operator with each technique will be tested by carrying out a comparison with the state of the denoising techniques using advanced signal processing. These techniques will be based on EMD [23], wavelet denoising [24], non-local means filtering [25] and VMD based [26].

The other major issue that was observed in the previous literature was the inadequate accuracy obtained for larger vocabulary sets. Phoneme based classification is one of the methods this issue can be resolved to some extent. But with phoneme-based classification the issue arises that certain phonemes have very similar activations of the muscle. This can lead to false recognition of the words. Moreover, we need to devise a new feature extraction technique that can be used to extract vital information from the signals and can be later used for robust classification.

VMD can be used to extract these features for a more robust classification of higher number of words in vocabulary set. VMD is resistant to noise and resolves the issue of modal aliasing [26]. VMD decomposes the signal into its sub signals. These resulting sub signals can be further processed to achieve a single representation of each frequency range. This single representation can be calculated by taking the square root of the eigen values extracted from different sub signals.

One of the major issues that literature fails to address is the development of speech activity detection systems. [14] reported a method for detection using EMG signals. But EMG signals are not a viable means of activity detection. This is due to various reasons, 1. EMG based activity detection systems assume rest when the muscle activation falls below a certain threshold which is not true for all cases because in some cases the activation levels are not significant when the words are articulated. 2. Subconscious haw movements can result in activation of facial muscles that can be considered as speech activity thus leading to faulty recognition of the words. Thus, we need to devise a new method independent of the EMG signals that addresses all these issues and successfully detects the activation region of speech. Inertial Measurement Units (IMUs) are electronic gadgets that have different motion sensors, like devices that measure how fast something is moving (accelerometers), ones that detect how something is rotating (gyroscopes), and ones that measure magnetic fields (magnetometers). These sensors work together to measure how an object is positioned, how fast it is moving, and how quickly its speed is changing. IMUs are used in various areas like robots, virtual reality, tracking movement, navigation, and wearable gadgets.

IMUs are very important in robots and self-driving cars for figuring out how they move, controlling their movements, and finding their way. They help us know where we are and which direction we are facing. This is widely used in robotics to find the target [27]. In virtual reality and augmented reality systems, IMUs are used to identify the movement of user's head in real-time. This aids in the creation of a smoother experience. IMUs are integrated into head-mounted displays (HMDs) to enable easy and smoother interaction [28]. IMUs are used in gait analysis during walking, playing sports, and recovering from injuries. IMUs also play a crucial role in measuring the angles of joints, the velocity of motion, and it's rotation. This helps us understand how things move and evaluate someone's athletic ability [29]. IMUs are put into wearable gadgets to keep an eye on your health and track your movements. They can give important information for keeping track of fitness, detecting falls, analyzing posture, and monitoring patients with movement disorders [30]. IMUs are sometimes used together with GPS to make navigation more accurate in places where GPS signals are not very good, like indoors or urban canyons. IMU signals help to keep track of position when GPS is not available [31]. All these applications of IMU suggest that they can be a good alternative of EMG for speech activity detection.

The problem statement can be formulated based on previously mentioned gaps in the literature.

> "To design an alternate system for speech recognition that utilizes EMG technology and IMU data to capture and interpret electrical signals generated by the facial muscles involved in speech production, with the goal of achieving reliable and accurate speech recognition and speech activation detection in a variety of situations where traditional microphones are impractical or ineffective."

Chapter 4: Data Collection

4.1. Subjects

Data was gathered from 10 individuals who were in good health. Before starting the experiment, everyone was informed about the whole procedure. All the people involved agreed and signed a consent form giving their permission for the data collection procedure. The people chosen were in good health and didn't have any muscle injuries. The selection requirement was the person had to have strong and healthy muscles in their face. The person in question does not have any sickness or injury that affects the movement of their facial muscles. The subjects must not have undergone any type of surgery.

4.2. Muscles of Interest:

Data was gathered using 4 wireless electrodes called Delsys Trigno (Avanti). Each person had four electrodes used on their faces. An electrode was put on each muscle we wanted to study. Because facial muscles are small and not as strong as muscles in other parts of the body, we made sure to focus on the biggest facial muscles that are used when speaking. The following muscles were chosen because they fit the requirements.

1. Buccinator: Buccinator muscle is one of the large facial muscles that is involved in speech activity. Buccinator lies between the maxilla and the mandible at the side of the face. It forms the anterior part of the cheek or the lateral wall of the oral cavity.
2. Masseter: Masseter is the strongest muscle of the human body. It plays a great role in mastication. It is a superficial muscle present on the side of the face.
3. Depressor anguli oris: Depressor anguli oris is a paired triangular muscle that extends from the mental tubercle of mandible to the angle of the mouth.
4. Digastric Muscle: Digastric muscle is present just below the jaw. It helps in the opening and closing of the jaw. Speech activity involves the movement of the jaw to produce a large vocabulary. Thus, the activation of this muscle can give vital information.

4.3. EMG and IMU:

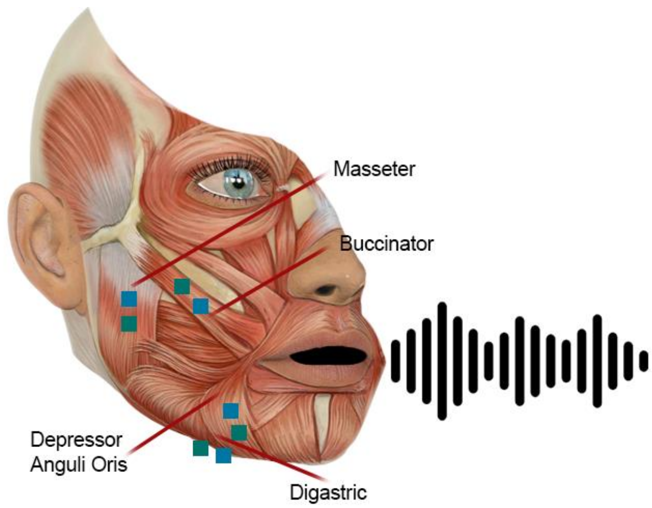

Figure 1: Electrode placement on the muscles of interest

The electrodes were used to collect muscle activity data from six people, while muscle activity and motion data were collected at the same time from four other people. The motion data is the IMU data which showed how the muscles were positioned in a 3D area. The IMU data comes from three types of sensors: accelerometers, gyroscopes, and magnetometers.

4.4. Vocabulary:

After picking the muscles we wanted to focus on, the next important thing was to choose the right words or set of data for the research. We chose words based on the classification and speech activity detection algorithm we are using. Using electrodes on the face to collect data for subvocal speech can make the subjects feel uncomfortable if the data recording is done all at once. However, gathering information on different days can cause differences in where the electrodes are put and how much the subject is stimulated, based on how much energy they have.

To deal with this, we chose a vocabulary size that would bring important and new results without making the subjects uncomfortable during data collection. 70 words were chosen for 6 out of the 10 subjects. These words were put into seven groups, and each group had ten words. There were different categories of words in the groups. One group had numbers from 0 to 9, another group had words we use to tell someone what to do in everyday life. There was also a group for words that give us directions, another for words that help us with our verbs, and another for words that replace nouns. The last two groups were for words that connect sentences together and for random words we use often.

For the other four subjects, we added 97 new words to the vocabulary in our dataset. Also, the researchers gathered ongoing spoken information from these participants. This information included 25 statements that were formed using the same set of 97 words.

4.5. Experimental Procedure:

The words were shown on the screen, one by one. Each word appeared for 5 seconds before being spoken. Only EMG data was collected for six subjects. For the remaining four subjects, both EMG and IMU data were collected from four muscles at the same time. The sampling rate for EMG was 2000 and 148 measurements per second for IMU. We gathered information using MATLAB 2020a. It took about 30 minutes to record 10 words articulated and repeated twenty times. For subjects 1-6, they articulated 70 words, and it took them a total of 4 hours. It takes 3. 5 hours for 7 groups to finish, with 30-minute breaks in between each ten words. The words used for subjects 7-10 got more, so the information and time also increased. The time it took to learn 97 vocabulary words for subjects 7-10 was about 6 hours. So, it took 48 hours to collect data from 10 subjects. The process of gathering information took place for a period of 10 days. Every day, information was gathered from just one person.

Chapter 5: Methodology: Denoising of EMG Signals

5.1. Data preprocessing

The EMG data that was gathered went through a process where a special filter was applied, allowing only frequencies between 10Hz and 450Hz to pass through. This filter was used to remove frequencies that were not within the desired range for the EMG data. We used a filter that eliminates electricity interference from the power lines, specifically at a frequency of 60Hz.

5.2. Data Filtration

The VMD method can separate harmonic signals that have similar frequencies without being affected by the sampling frequency. This helps prevent the mixing of different modes. VMD is a simplified version of the wiener filter that can be used in different frequency ranges. The model and center frequency are regularly updated together, which makes the model estimation changeable. After each guess, the model is changed into the time domain by using the inverse Fourier transform.
VMD breaks down the original signal into smaller sub-signals called VMFs.

$$s = \sum_{n=1}^{M} \mu^n \quad (1)$$

μ^n is the component of the signal and is given as:

$$\mu^n(t) = b^n(t) \cdot \cos(\theta^n(t)) \quad (2)$$

Hilbert transform is used to attain the marginal spectrum of the signal given by:

$$fs = [\delta(t) + j\pi t] * (t) \quad (3)$$

A squared norm of the signal is obtained giving (4).

$$min_{\{u_k\},\{w_k\}} \left| \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \right. \quad (4)$$

Where u_k is the sub signals obtained from f. The constrained problem in (4) is converted by Langrangian multiplication operator λ and quadratic penalty factor α as given in (5).

$$L(\{u_k\},\{w_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_k(t) \right] e^{ijw_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(r) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k \rangle \quad (5)$$

Transforming (5) into spectral domain results in the following.

$$\hat{u}_k^{n+1}(w) = \frac{f(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \quad (6)$$

$$w_k^{n+1} = \frac{\int_0^\infty w|\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \quad (7)$$

The VMFs are processed by dividing them into smaller sections and applying a technique called windowing. This is done using a window that overlaps by 10% for a duration of 250 milliseconds.

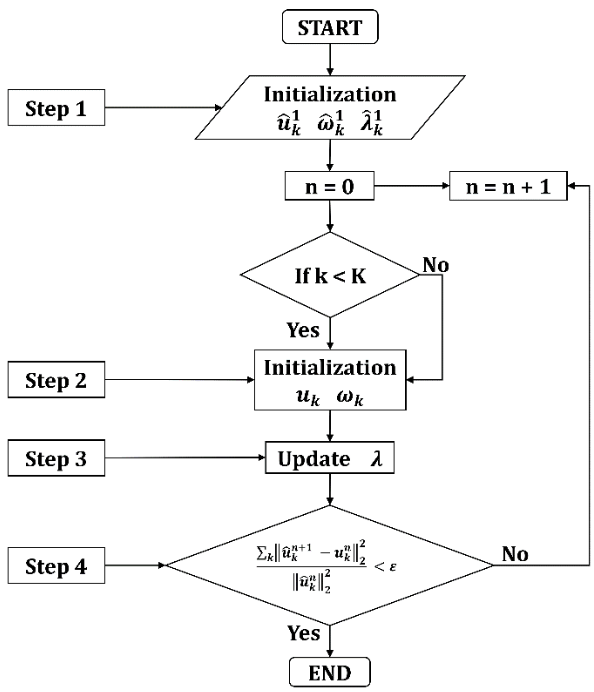

Figure 2: Flowchart showing the extraction of VMFs using VMD

The whole process of extracting VMFs is shown in fig. 2.

Time-frequency denoising techniques are based on the ideas of denoising in the wavelet domain. Because the decomposed VMFs look like signals that are either changing in amplitude or frequency, the average value of each VMF is zero. So, using wavelet-like thresholding directly, either SOFT or HARD, will not give correct results.
HARD thresholding makes any coefficients that are smaller than a specific value (T) zero. In HARD thresholding, the coefficients that are higher than T stay the same. After applying HARD thresholding, the signal that has been reconstructed shows breaks or gaps. To solve this problem, we use a technique called SOFT thresholding. Soft thresholding uses the actual values of the coefficients. We set any values smaller than T to zero. The other values that are not zero are moved closer to zero. This makes the reconstructed signal have a bias. To make things fair, the SCAD penalty makes some numbers zero, reduces a few numbers to zero, and keeps the larger numbers the same. In math, we can use the numbers 13, 14, and 15 to represent SOFT, HARD, and SCAD.

$$f(u_k) = \begin{cases} sgn(u_k)(|u_k| - T), |u_k| > T \\ 0\ , |u_k| \le T \end{cases} \tag{13}$$

$$f(u_k) = \begin{cases} u_k, |u_k| > T \\ 0, |u_k| \le T \end{cases} \tag{14}$$

$$f(u_k) = \begin{cases} sgn(y)\max(0, |u_k| - T)\ , |u_k| \le 2T \\ \dfrac{(z-1)u_k - zT(sgn(u_k))}{z-2}\ , 2T < |u_k| \le zT \\ u_k, |u_k| > zT \end{cases} \tag{15}$$

u_k represents the Virtual Machine File (VMF) that we got from the VMD process described earlier. The Bayesian argument suggests that it is best for the value of z to be 3.
The value of T is set higher than maximum component in the noisy signal [32]. The value for T is chosen by:

$$T = \sigma\sqrt{2\log_e N} \tag{16}$$

The equations for soft, hard and SCAD can be given as:

$$\tilde{p}_j^i = \begin{cases} p_j^{(i)}, \left|q_j^{(i)}\right| > T_i \\ 0, \left|q_j^{(i)}\right| > T_i \end{cases} \tag{18}$$

$$\tilde{p}_j^i = \begin{cases} p_j^{(i)} \dfrac{\left|q_j^{(i)}\right| - T_i}{\left|q_j^{(i)}\right|}, \left|q_j^{(i)}\right| > T_i \\ 0, \left|q_j^{(i)}\right| \le T_i \end{cases} \tag{19}$$

$$\tilde{p}_j^i = \begin{cases} p_j^{(i)} \dfrac{\max\left(0, |q_j^{(i)}| - T_i\right)}{|q_j^{(i)}|}, |q_j^{(i)}| > T_i \\ p_j^{(i)} \dfrac{(z-1)\left(0, |q_j^{(i)}| - zT_i\right)}{|q_j^{(i)}|}, 2T_i < |q_j^{(i)}| \le zT_i \\ 1, |q_j^{(i)}| > zT_i \end{cases} \tag{20}$$

The overall procedure for this objective is shown in fig. 3.

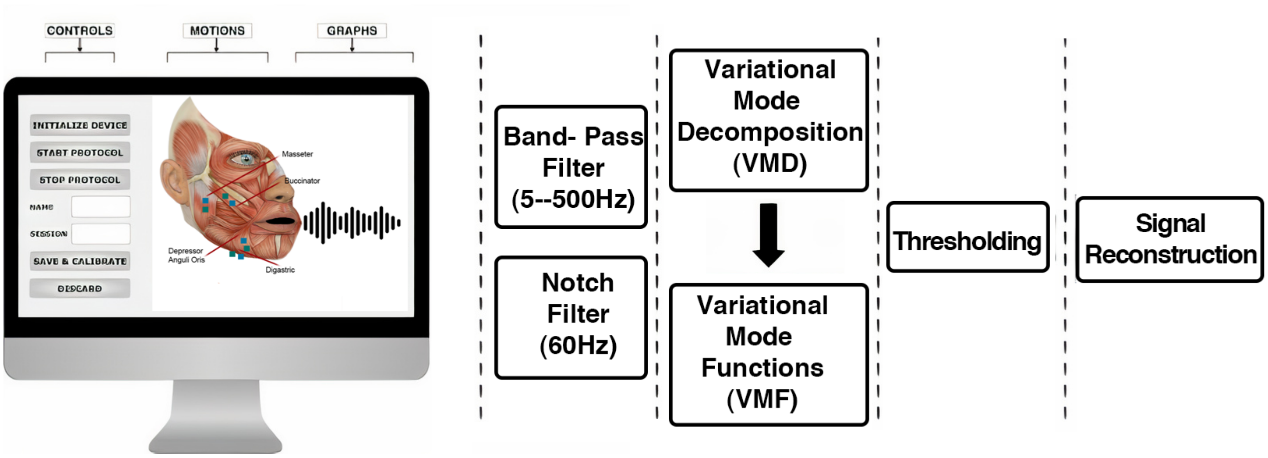

Figure 3: Graphical abstract of denoising process

Chapter 6: Methodology: Feature Extraction

After filtering and preparing the gathered data, the next thing we did was to gather important details from the signals. This collected information is called features. These features will be used to sort and separate the data into different groups based on the words used. The method to extract features was created using a complex signal processing technique called VMD and singular vector decomposition SVD.

6.1. Feature Extraction

VMD is mainly based on the idea of the Wiener filter [21]. The model estimation for VMD is variable because the center frequency is consistently changed. After each step, the way the data in time is calculated by using the inverse Fourier transform. VMD breaks down the original signal into separate parts called Variational Mode Functions (VMF) as given by equation (1).

$$s = \sum_{n=1}^{M} \mu^{n} \qquad (1)$$

μ^{n} is the component of the signal and is given as:

$$\mu^{n}(t) = b^{n}(t) \cdot \cos(\theta^{n}(t)) \qquad (2)$$

Hilbert transform is used to attain the marginal spectrum of the signal given by:

$$fs = [\delta(t) + j\pi t] * (t) \qquad (3)$$

A squared norm of the signal is obtained giving (4).

$$min_{\{u_k\},\{w_k\}} \left| \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \qquad (4)$$

Where u_k is the sub signals obtained from f. The constrained problem in (4) is converted by Langrangian multiplication operator λ and quadratic penalty factor α as given in (5).

$$L(\{u_k\},\{w_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_k(t) \right] e^{ijw_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(r) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k \rangle \qquad (5)$$

Transforming (5) into spectral domain results in the following.

$$\hat{u}_k^{n+1}(w) = \frac{f(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \qquad (6)$$

$$w_k^{n+1} = \frac{\int_0^\infty w|\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \quad (7)$$

The VMFs are processed by dividing them into smaller sections and applying a technique called windowing. This is done using a window that overlaps by 10% for a duration of 250 milliseconds. Each sub signal has many values, so a way to simplify the classifier needed to be put in place. SVD is used to calculate single values. The values that come out of it are thought of as unique features and make up the SVD-VMD feature vector. Fig. 3 gives an overview of the process.

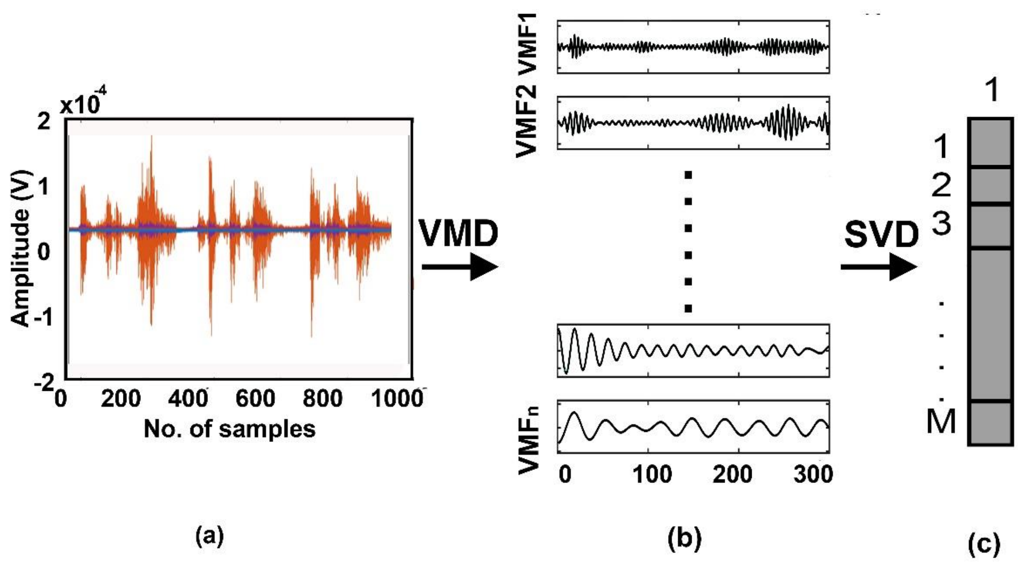

Figure 4: Overview of feature extraction using SVD-VMD

6.2. Classification

The effectiveness of the SVD-VMD classifier was checked for four non-parametric classifiers. Although Linear Discriminant Analysis (LDA) is considered the gold standard because of its minimal execution time for pattern recognition but LDA performs worse than Random Forest (RF) due to which it was not included in the study [33]. Moreover, [34] determined that Support Vector Machine (SVM) surpasses LDA, Artificial Neural Network (ANN) and K-Nearest Neighbor (KNN). Even though [34] concluded SVM performs better than KNN, [35] utilized KNN for controlling a dexterous artificial hand and achieved better results. [36] employed a type of Decision Tree (DT) to classify wrist movements but failed to devise a comparison with other classifiers. Thus, the classifiers selected include SVM, KNN, DT and RF.

To check the efficacy of the proposed algorithm it was tested on three types of datasets. One being the speech data collected for this study. But the initial results were obtained on a dataset collected for four upper limb motions (hand open, hand close, wrist extension and wrist flexion). The third dataset that was utilized to test the efficacy of the proposed algorithm was Ninapro DB2 dataset.
Data were randomly split into training, cross validation, and testing set. The dataset was evenly balanced for each motion thus the metric used for testing the efficacy of the feature sets was percentage accuracy.

Chapter 7: Methodology: Speech Activity Detection

Speech signals were recorded from four people who used a vocabulary of 97 words. The data included signals from both muscle movement (EMG) and muscle motion (IMU). I used the speech data recorded from these people to create a system that can detect when they are speaking by using IMU data. The algorithm uses signals from the IMU (accelerometer and gyroscope data) to find the areas where there is the most movement to detect speech regions. The algorithm is based on following stages:

7.1. Stage I

Local statistics of each component of IMU signal are calculated using an optimum window land overlap length. These statistics are then thresholded using the mean value of the statistics vector. Thresholding results in a vector that gives the activations occurring in the signal in the form of zeros and ones. The resulting binary values are labelled as onsets and offsets. The onsets and offsets are joined to get a continuous speech activity region.

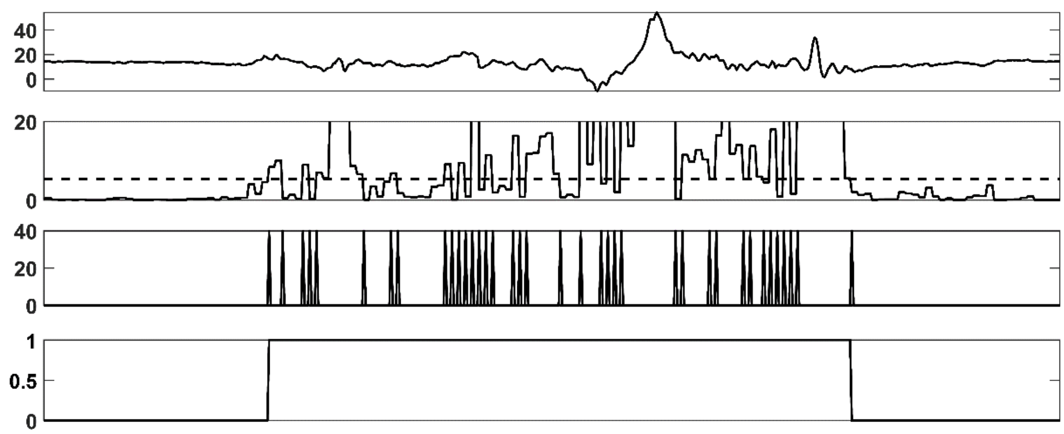

Figure 5: (a) Input IMU signal. (b) variance vector and threshold value (c) activations (d) Active region binary vector

This step is demonstrated in figure 4.

7.2. Stage II

Since each IMU signal has 6 subsignals (3 accelerometer along each axis and 3 gyroscopes along each axis) and 4 channels were utilized thus we had a total of 24 signals for each utterance. After the computation of activation region in each signal separately the next step was to compute an activation region representing the speech activity along accelerometer and gyroscope separately. Thus, at the end of this step we had 2 signals corresponding to each channel. This results in a total of 8 vectors showing the speech activity region. This step utilizes the basic concepts of sets. Union and intersection of different regions are calculated resulting in a region of activation. The mutual activation region in any two axes contributes to the region of activation. This step also takes care of small fluctuations which can be detected as speech. This is done using morphological open operation used in image processing. The steps have been depicted in figure 5.

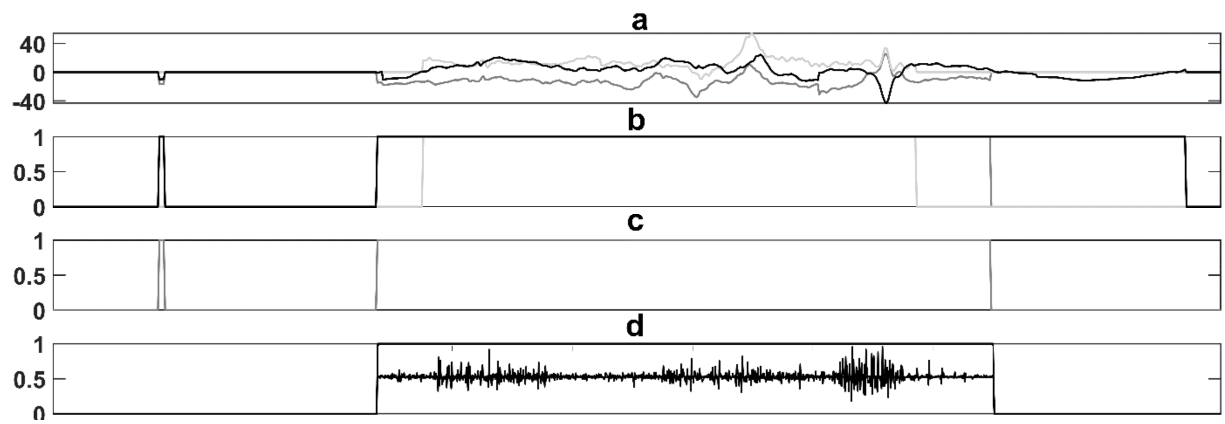

Figure 6: (a) IMU signals signals (c) activation regions separately. (d) common activation. (e) filtered activation and EMG signal

7.3. Stage III

As IMU signals contain data from both accelerometers and gyroscopes for each of the directions: x, y, and z.

In the previous step, we combined data from the accelerometer and gyroscope on all three axes to get one overall measurement of activation corresponding to each. So, it gives us two signals for each channel and a total of 8 signals. In this step, we calculate one activation area by using the two signals corresponding to accelerometer and gyroscope for each channel. This step considers the intersection region of the two signals as the overall activation. This step results in one signal for each channel and 4 signals corresponding to the four channels.

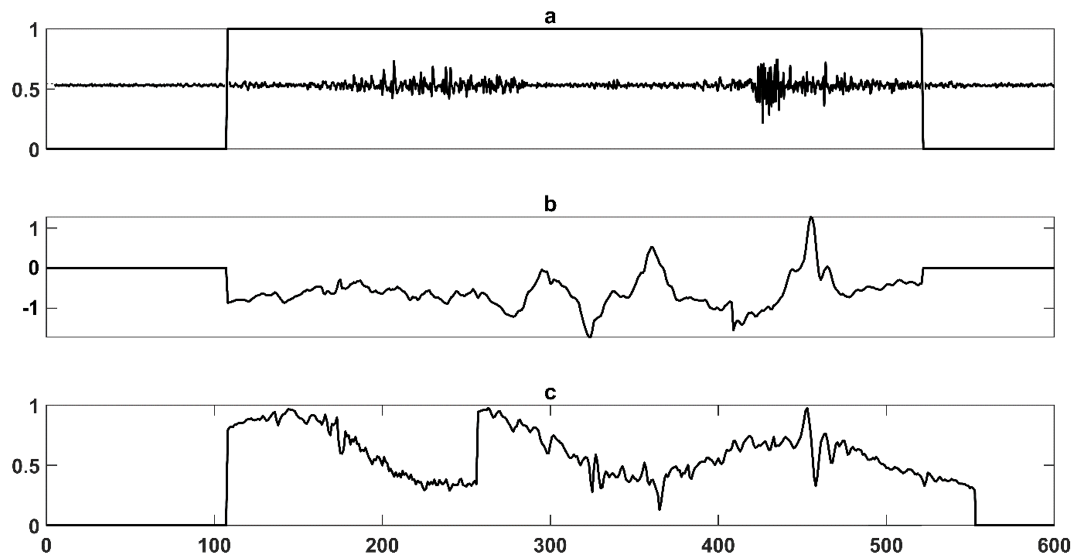

Figure 7: (a) Activation region (b) filtered gyro data (c) filtered acc data

7.4. Stage IV

The last stage has 4 activation regions as input and one activation region needs to be computed. These four activation signals are the activation of each channel. The basic principle for this step is that if any three channels have a common region that will be conclude as the speech activity region.

The different activations for each channel and the resulting final output giving the overall activation are shown in figure.

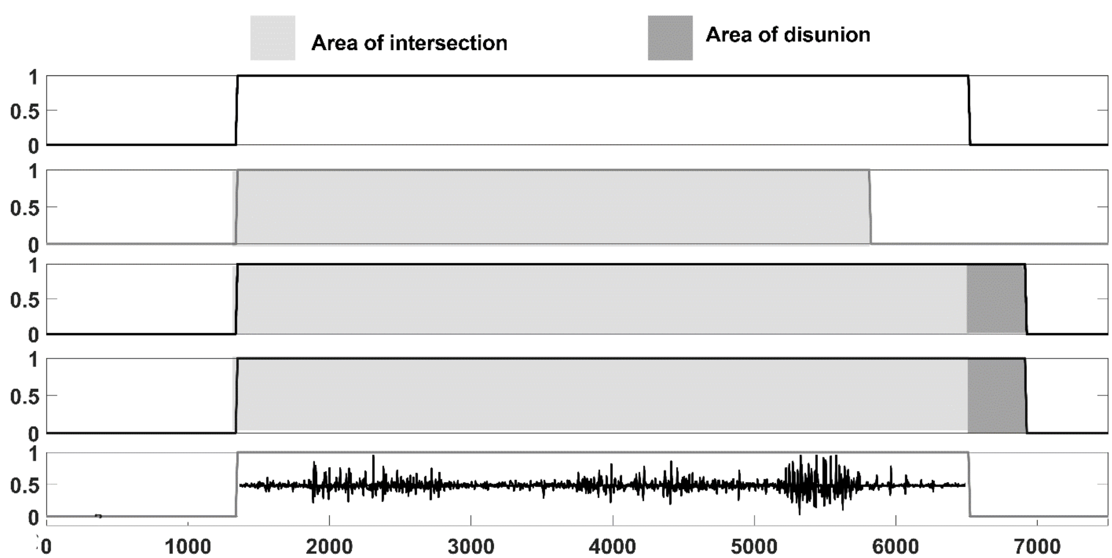

Figure 8: Computation of activation of 4 channels.

Chapter 8: Results

The results will be described separately for each objective under the following headings:

8.1. Denoising of data

In many different situations, such as diagnosing diseases, recognizing motions and gestures, and interacting with computers, EMG signals often get mixed with unwanted noise. This can make it harder to accurately analyze the signals. To solve this problem, we propose a method called variational mode decomposition (VMD) to reduce noise. By using VMD, muscle signals are separated into different types of movements. Then, we choose the right levels to get rid of unwanted sounds and make the signals clean again.

We tested how well two denoising techniques, called interval thresholding (IT) and iterative interval thresholding (IIT), worked. We used a measure called Signal-to-Noise Ratio (SNR) to evaluate their performance, and we verified the results using a statistical test called the Friedman test. Surprisingly, IIT was consistently better than IT at removing noise from EMG signals, no matter how much noise there was (with a P-value of less than 0. 05) Furthermore, SOFT performed better than HARD and SCAD among the thresholding operators.

To get the best noise reduction without changing the original signal, the best method was found to be combining IIT thresholding with VMD-based SOFT thresholding. As a result, our new method shows great potential for use in disease diagnosis, recognizing patterns, and classifying movement. This has already been published in a journal [32].

The results are shown in table 1 [32].

Table 1: Mean Signal-to-noise Ratio Corresponding to surface EMG.

Noise in db	**0**	**0**	**5**	**5**	**10**	**10**	**15**	**15**
Performance parameter	**StNR**	**RSE**	**StNR**	**RSE**	**StNR**	**RSE**	**StNR**	**RSE**
Wavelet	0.61	1.20	1.31	0.91	1.7	0.80	2.02	0.80

Denoising								
Non-Local Means	1.32	1.00	2.00	0.70	2.21	0.61	2.22	0.61
Weiner Filter	-1.32	1.52	1.4	0.90	2.41	0.61	4.10	0.51
EMD +IT+ SOFT	3.21	0.61	5.93	0.51	8.92	0.31	11.0	0.30
VMD +IT+ SOFT	3.21	0.61	8.51	0.31	14.00	0.21	19.4	0.11
EMD +IIT+ SOFT	3.43	0.61	6.22	0.41	9.22	0.30	11.3	0.31
VMD +IIT+ SOFT	**3.91**	**0.61**	**9.2**	**0.3**	**14.6**	**0.2**	**19.8**	**0.11**
VMD +IIT+ HARD	1.74	0.70	6.90	0.41	12.22	0.2	17.41	0.12
VMD +IIT+ SCAD	2.80	0.70	8.10	0.36	13.61	0.2	19.02	0.11

(RSE= root mean square error and StNR = signal to noise ratio)

8.2. Robust Classification of individual Word Vocabulary

In this study, a new way of getting features from the recorded signals was used, as explained before. To check how well the technique worked, two sets of data were used. Both sets of data were about movements made with the upper part of the body. The initial set of information included data from 10 people who were in good health. It covered four different types of movements and three categories of data. The second set of data used was the Ninapro DB2 dataset, which is freely accessible to the public. In simpler words, the researchers tested how accurately they could classify different combinations of motions and channels in the Ninapro dataset.

The results we found are shown in Table 2. The results in Table 2 show that the best accuracy in predicting movements was achieved using 12 specific channels that correspond to four different types of motions. Furthermore, it is clear that the accuracy of classifying information is still very good even when there are more channels involved.

However, it is important to understand that if you increase the number of actions but decrease the number of ways to track them, the accuracy of categorizing them may decrease.

This means that the accuracy of this classification was measured for 40 people. Each person did a different number of still movements. The results show that the new technique of extracting features could accurately classify different movements, especially when using the right combination of movements and channels. These findings are important for improving motion analysis and recognizing gestures in applications.

Table 2 % accuracy of proposed algorithm Ninapro

No. of Channels	**No. of Motions**			
	4	**8**	**13**	**17**
three	90.±0.4	85 ±0.3	80±0.2	77±0.3
five	95±0.6	92±0.4	89±0.3	89±0.2
eight	96±0.7	95±0.4	93±0.2	92±1.6
twelve	98±0.5	98±0.1	96±0.1	96±0.1

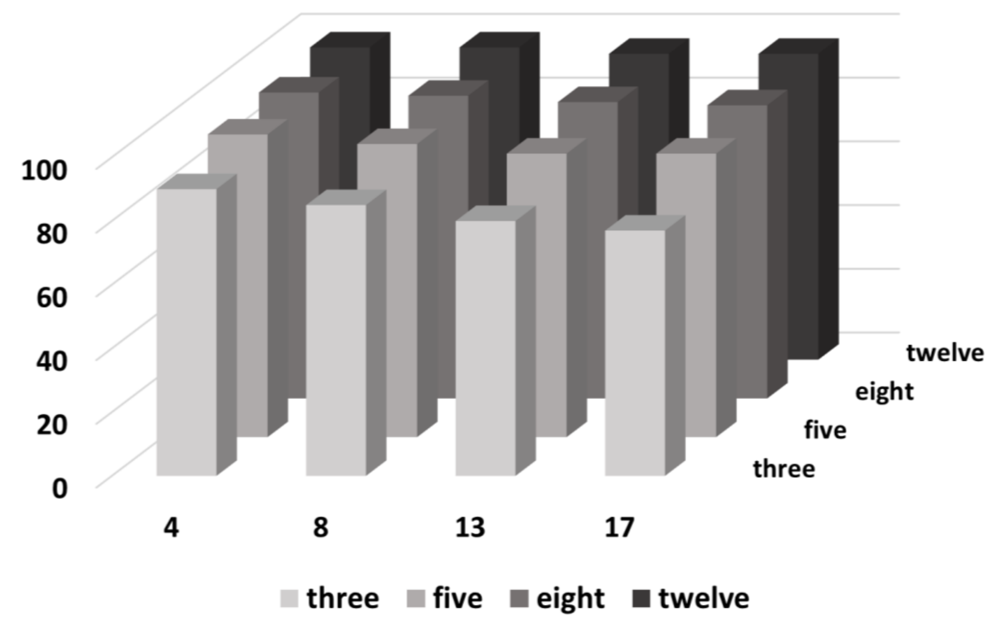

Figure 9: % accuracy of proposed algorithm Ninapro

In the past, we tested the algorithm we created with the Random Forest classifier to check if it works well in controlling myoelectric systems. The research looked at how well the algorithm worked using a set of data called Ninapro DB2. This data involved 40 people who were healthy and they did four different movements: opening their hand, closing their hand, extending their wrist, and flexing their wrist. A paper on this topic has been written and is being reviewed.

The research wanted to see how three different ways of extracting information from something compared to each other. They also looked at two combinations of these ways of extracting information and compared them as well. These new methods, called SVD-VMD and SVD-EMD, use different mathematical processes like VMD, SVD, and EMD to break down data into smaller parts. On the other hand, the third technique was simpler and involved getting important information from the time domain. The two combined sets of features were created by blending two advanced techniques with features related to time.

In Table 3, a comparison of five different sets of features is shown. Each set of features is tested with four different classifiers. This study wanted to find the best way to use certain techniques and tools to control myoelectric systems effectively. By analyzing how well each group of features worked with different classifiers, we learned important things that could help improve how myoelectric systems are controlled. These improvements could be useful in many different areas and applications.

Table 3: Comparison different feature extraction techniques

Feature vectors	Support Vector Machine (%)	K-Nearest Neighbor (%)	Decision Tree (%)	Random Forest (%)
Time-domain	81±1	82±0.4	87±0.7	92±0.4
EMD-based	75±1	83±1	86±0.8	90±0.4
VMD-based	87±1	97±0.3	93±0.5	**97±0.2**
Time-domain+ EMD-based	83±1	87±0.6	89±0.6	92±0.4
Time-domain+ VMD-based	90±1	92±0.7	92±0.5	92±0.5

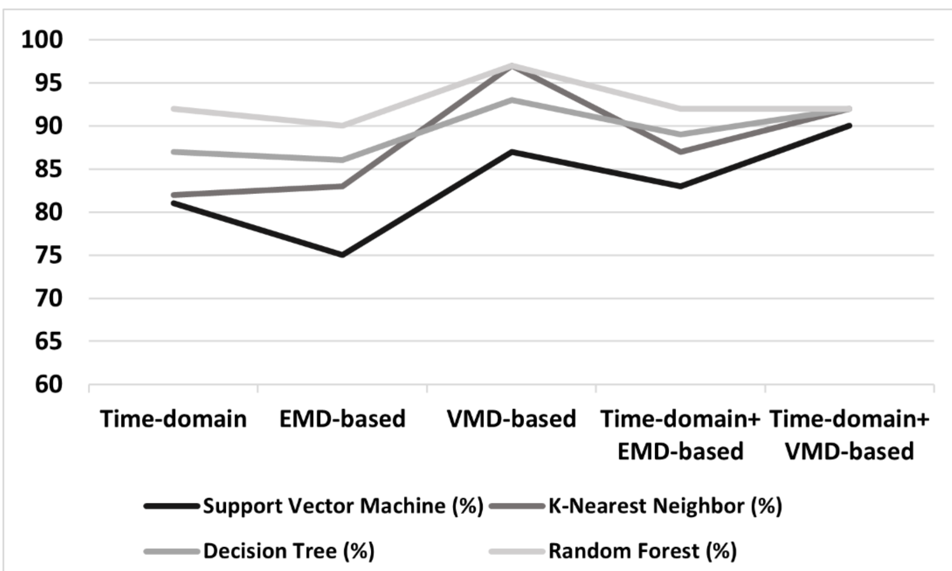

Figure 10: Comparison different feature extraction techniques

The new technique worked well for the upper limb's movements and was also used to categorize lists of individual words. We analyzed 10 healthy subjects using the SVD-VMD algorithm to categorize them based on different subjects. The characteristics were taken from each signal using 12 VMFs. Table 4 shows how accurately each subject was classified using a technique called SVD-VMD and a classifier called RF.

The feature extraction method we created, combined with RF classifiers, showed good performance. It achieved the highest accuracy of 98. 6% and 92% for vocabulary sets containing 70 and 96 words, respectively. The mean accuracies corresponding to two different types of data corpuses and classifiers is given in table 4.

Table 4: % accuracy for isolated data corpus

Classifier	Vocabulary	Accuracy	Standard deviation
RF	70	97.82	0.61
RF	96	90.03	3.37
KNN	70	85.58	2.43
KNN	96	73.50	4.15

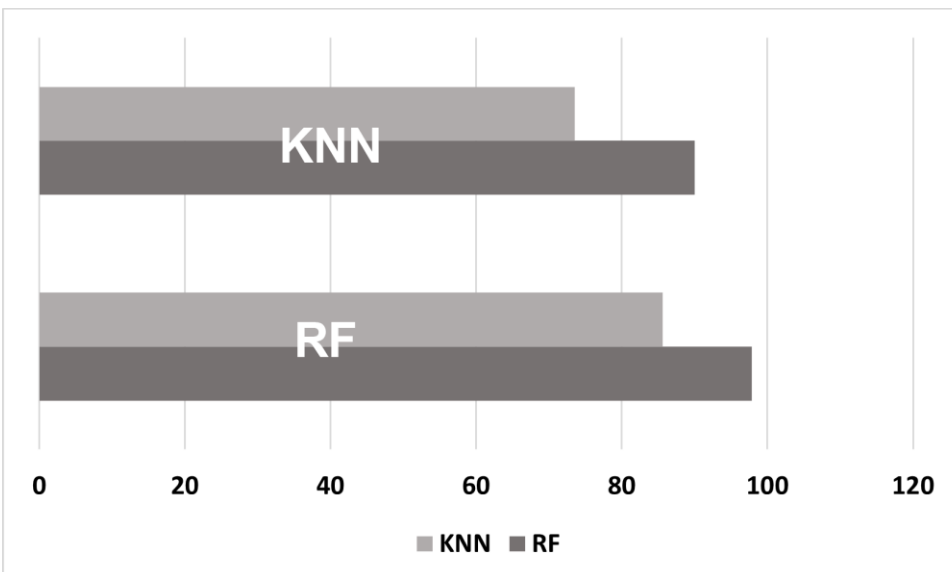

Figure 11: % accuracy for isolated data corpus

8.3. Efficacy of Speech Activity Detection

Table 5 shows the results obtained for IMU based speech activation detection system. The table shows the mean activation error rate corresponding to two detection systems. The first being the developed algorithm which is IMU based whereas the second is an EMG based activation detection system. It works on the principal of Teager Kaiser operator and morphological transforms.

The able shows the mean and standard deviation corresponding to both the methods. The results are the mean activation error rates obtained by using data from 4 different subjects corresponding to 96 words.

Table 5: Error rate for speech activity detection

Method	Mean	Standard Deviation
IMU	0.138	0.04
EMG	0.275	0.073

Chapter 9: Discussion

9.1. Denoising

The study focusses on three main objectives. All three will be discussed one by one. The first objective was to carry out the denoising of the EMG signals using advanced signal processing technique VMD combined with two thresholding techniques (IT, and IIT) and three thresholding operators (SOFT, HARD and SCAD). The results for this objective have been reported in chapter 8 section 8.1. It was evident that the proposed denoising technique (VMD-SIIT) outperformed the remaining techniques utilized for denoising purposes. VMD results in narrow band frequency spectra [37] making it easier to get rid of the noise thus aiding to the better performance of VMD for denoising purposes. VMD does not entail the issue of mode mixing because of its mathematical basis due to which it performs better than other methods such as EMD which has the issue of mode mixing [38].

IIT and IT were used along with VMD to carry out the denoising of the signals. It was evident in the results that IIT outperformed IT. It is because of the iterative assumptions of the IIT operator for the denoised signals [39]. Among the three types of thresholding operators we used, SIIT-VMD produced the highest signal-to-noise ratio (SNR) and the lowest root mean square error (RMSE) for the reconstructed signal. SIIT-VMD solves the problem of signal breaks that happen with HIIT-VMD because it uses a strict cutoff point [40].

The method suggested can be useful in extracting time-frequency features from electromyography data, which can provide us with more valuable information. These discoveries can be useful in different areas such as detecting diseases [41], observing muscle activity over a few days, analyzing limb movements [42], understanding movement intentions [43], interacting with machines [44], and detecting motion. In addition, the SIIT-VMD denoising method can be used for various purposes like evaluating spasticity, checking for signal consistency, and recognizing gestures.

9.2. Feature Extraction

Traditional features such as time domain signals, although computationally less expensive but depend only on the amplitude of the signals. Due to their dependance only on amplitude these features are not sufficient for robust classification. Time domain signals are more prone to distortion due to noise and thus causing faulty classification for the main objective. Thus, a new feature extraction technique needed to be developed for robust classification. The results of this proposed method have been described in chapter 8 section 8.2.

The use of VMD for extracting the features results in vital information from both time and spectral domain. As VMD works in both domains it covers all the information from both domains. VMD has narrow frequency bands [37] which makes it easier to identify signals from the noise resulting in better performance of VMD based features.

The use of SVD makes the features more stable and results in a more robust classification. After decomposition of the signal into its subsequent signals, SVD is applied to them resulting in eigen values. The square root of these eigen values give us singular values which are used as individual features. It is evident from the results that VMD based features outperform all the other feature vectors. The classifier that works best was random forest and KNN. This is since KNN computes the distance and forms a very complex boundary making it possible to classify high dimensionality data [45]. Similarly, RF, as an ensemble learning technique, harnesses the power of multiple decision trees, leading to effective performance in EMG-based classification tasks [46].

However, it is important to recognize some problems with VMD, such as its dependence on a set number of ways to break things down. You should think about VMD's difficulty in separating the base level of a signal and its limited ability to handle sudden changes in the signal, like sudden increases or decreases. Moreover, when using the SOFT operator in SIIT-VMD, it causes the signal to move by a set threshold value. This can create unnecessary bias in the resulting signal and make the computation more complicated because it requires going through a repetitive process [47]. In the future, we should do a study to see how well the proposed method can remove noise from concentric and eccentric isotonic motions.

9.3. Speech Activity Detection

The previously proposed speech activity detection proposed by [14] had various drawbacks. These drawbacks included that it assumes that there are times when the muscles used for speaking are not very active, which can indicate times when someone is not speaking. But we discovered that there are times when speech sounds are produced without any obvious decrease in muscle activity. This happens when people don't relax their muscles after speaking, leading to increased muscle activity in the background. Including sEMG activity biofeedback while collecting tokens could help people relax their speech muscles. This can improve EMG-based Speech Activity Detection (SAD) [14].

Instead, using an Inertial Measurement Unit (IMU) could also be an option. The IMU can follow the movement of the jaw and/or lips. It can be combined with the EMG sensors or positioned in a strategic way to accurately track the movement of the jaw. Thus, this study proposes an IMU based speech activation detection technique.

The results of this objective are explained in chapter 8 section 8.3. The major drawback in [14] was the detection of unnecessary jaw movement which was not a part of speech activity and was considered as the speech region. This results in faulty speech recognition. IMU is independent of muscle activation and is only dependent on the muscle orientation and acceleration thus it performed well for the required objective [48].

Chapter 10: Conclusion

The study aimed to resolve three major drawbacks reported in previous literature. The first was to introduce a denoising technique using advanced signals processing, the second was to develop a robust feature extraction technique and the third was to develop a speech activity detection algorithm. The first and second objective was addressed using VMD. In the first objective the denoising was done using VMD with Soft operator and iterative interval thresholding. The second objective was carried out using VMD along with singular vector decomposition and its efficacy was proven by comparison with other state of the art feature extraction techniques. The third objective was addressed using IMU data instead of EMG signals. It was observed that IMU based detection outperforms EMG based techniques.

References

[1] De Luca, C. J. (1997). "The Use of Surface Electromyography in Biomechanics." Journal of Applied Biomechanics, 13(2), 135-163.

[2] Merletti, R., & Parker, P. A. (2004). "Electromyography: Physiology, Engineering, and Non-Invasive Applications." Wiley-IEEE Press.

[3] Hermens, H. J., Freriks, B., Disselhorst-Klug, C., & Rau, G. (2000). "Development of Recommendations for SEMG Sensors and Sensor Placement Procedures." Journal of Electromyography and Kinesiology, 10(5), 361-374.

[4] Farina, D., Merletti, R., & Enoka, R. M. (2014). "The Extraction of Neural Strategies from the Surface EMG." Journal of Applied Physiology, 117(11), 1215-1230.

[5] Muceli, S., Boye, A. T., d'Avella, A., & Farina, D. (2010). "Identifying Representative Synergy Matrices for Describing Muscular Activation Patterns During Multidirectional Reaching in the Horizontal Plane." Journal of Neurophysiology, 103(3), 1532-1542.

[6] Li, Y., Guo, Y., Zhao, Y., & Peng, H. (2020). "A Review of Subvocal Speech Recognition and Its Brain-Computer Interface Applications." Journal of Neural Engineering, 17(4), 041001. doi: 10.1088/1741-2552/ab9b3a

[7] Rogin, E., & Porrill, J. (2016). "Subvocal Speech Recognition with a Surface EMG Sensor." Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 230(6), 553-561. doi: 10.1177/0954411916643685

[8] Verma, S., & Guo, Y. (2018). "EMG-Based Subvocal Speech Recognition: A Systematic Review." IEEE Transactions on Biomedical Engineering, 65(2), 477-487. doi: 10.1109/TBME.2017.2695908

[9] Vojtech JM, Chan MD, Shiwani B, Roy SH, Heaton JT, Meltzner GS, Contessa P, De Luca G, Patel R, Kline JC. “Surface Electromyography-Based Recognition, Synthesis, and Perception of Prosodic Subvocal Speech.” J Speech Lang Hear Res. 2021 Jun 18;64(6S):2134-2153. doi: 10.1044/2021_JSLHR-20-00257. Epub 2021 May 12. PMID: 33979177; PMCID: PMC8740708.

[10] Ratnovsky, A., Malayev, S., Ratnovsky, S. et al. “EMG-based speech recognition using dimensionality reduction methods.” J Ambient Intell Human Comput 14, 597–607 (2023). https://doi.org/10.1007/s12652-021-03315-5

[11] Kumar, Suresh & Ali, Zahid & Kumar, Chandar & Abid, Ghulam & Shaikh, Shoaib & Memon, Vengas. (2020). Speech recognition based robotic mart. International Journal of Intelligent Robotics and Applications. 4. 10.1007/s41315-020-00144-1.

[12] Lee KS. “EMG-based speech recognition using hidden markov models with global control variables.” IEEE Trans Biomed Eng. 2008 Mar;55(3):930-40. doi: 10.1109/TBME.2008.915658. PMID: 18334384.

[13] M. Janke and L. Diener, “EMG-to-Speech: Direct Generation of Speech From Facial Electromyographic Signals,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2375-2385, Dec. 2017, doi: 10.1109/TASLP.2017.2738568.

[14] Meltzner GS, Heaton JT, Deng Y, De Luca G, Roy SH, Kline JC. “Silent Speech Recognition as an Alternative Communication Device for Persons with Laryngectomy. IEEE/ACM Trans Audio Speech Lang Process.” 2017 Dec;25(12):2386-2398. doi: 10.1109/TASLP.2017.2740000. Epub 2017 Nov 28. PMID: 29552581; PMCID: PMC5851476.

[15] J. Wang, L. Tang and J. E Bronlund, "Surface EMG Signal Amplification and Filtering", International Journal of Computer Applications, vol. 82, no. 1, pp. 15-22, 2013. Available: 10.5120/14079-2073 [Accessed 23 February 2022].

[16] “Variational mode decomposition,” Variational Mode Decomposition - an overview | ScienceDirect Topics. [Online].Available:www.sciencedirect.com/topics/engineering/variational-mode-decomposition [Accessed: 23-Feb-2022].

[17] G. Li, G. Tang, G. Luo and H. Wang, "Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition", Mechanical Systems and Signal Processing, vol. 120, pp. 83-97, 2019. Available: 10.1016/j.ymssp.2018.10.016 [Accessed 23 February 2022].

[18] K. Hong, L. Wang and S. Xu, "A Variational Mode Decomposition Approach for Degradation Assessment of Power Transformer Windings," in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 4, pp. 1221-1229, April 2019, doi: 10.1109/TIM.2018.2865048.

[19] Q. Wang, C. Yang, H. Wan, D. Deng and A. Nandi, 2021. Bearing fault diagnosis based on optimized variational mode decomposition and 1D convolutional neural networks. Measurement Science and Technology, 32(10), p.104007.

[20] S. Lahmiri and M. Boukadoum, 2014. Biomedical image denoising using variational mode decomposition. 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings.

[21] P. Singh and G. Pradhan, 2018. Variational mode decomposition based ECG denoising using non-local means and wavelet domain filtering. Australasian Physical & Engineering Sciences in Medicine, 41(4), pp.891-904.

[22] F. Xiao, D. Yang, X. Guo and Y. Wang, 2019. VMD-based denoising methods for surface electromyography signals. Journal of Neural Engineering, 16(5), p.056017.

[23] H. Ashraf, A. Waris, S. O. Gilani, M. U. Tariq and H. Alquhayz, 2021. Threshold Parameters Selection for Empirical Mode Decomposition-Based EMG Signal Denoising. Intelligent Automation & Soft Computing, 27(3), pp.799-815

[24] Ç. P. Dautov and M. S. Özerdem, "Wavelet transform and signal denoising using Wavelet method," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404418.

[25] A. Buades, B. Coll and J. . -M. Morel, "A non-local algorithm for image denoising," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 60-65 vol. 2, doi: 10.1109/CVPR.2005.38.

[26] F. Xiao, D. Yang, X. Guo and Y. Wang, 2019. VMD-based denoising methods for surface electromyography signals. Journal of Neural Engineering, 16(5), p.056017.

[27] Li, K., & Mourikis, A. I. (2013). "Improving the Accuracy of EKF-based Visual-Inertial Odometry." Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).

[28] Tsoupikova, D., Kaufman, A. E., & Jaskolka, M. (2007). "Head Tracking with the Intersense IS-900 for Visualization of Data in Virtual Reality." Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)

[29] Sabatini, A. M. (2011). "Quaternion-based Extended Kalman Filter for Determining Orientation by Inertial and Magnetic Sensing." IEEE Transactions on Biomedical Engineering

[30] Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., & Meijer, K. (2009). "Activity Identification Using Body-Mounted Sensors - A Review of Classification Techniques." Physiological Measurement.

[31] Frew, E. W., Johnson, E. N., & Rathbun, D. L. (2003). "IMU-enhanced GPS for Precise High Dynamic Kinematic Positioning." Proceedings of the Institute of Navigation (ION) GPS/GNSS.

[32] H. Ashraf, U. Shafiq, Q. Sajjad, A. Waris, O. Gilani, M. Boutaayamou, O. Brüls, "Variational mode decomposition for surface and intramuscular EMG signal denoising, Biomedical Signal Processing and Control", Volume 82, 2023, 104560, ISSN 1746-8094,

[33] H. Zhou and G. Alici, "A compact and cost-effective pattern recognition based myoelectric control system for robotic prosthetic hands," 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 2020, pp. 270-275, doi: 10.1109/AIM43001.2020.9158830.

[34] M. V. Arteaga, J. C. Castiblanco, I. F. Mondragon, J. D. Colorado and C. Alvarado-Rojas, "EMG-based adaptive trajectory generation for an exoskeleton model during hand rehabilitation exercises," 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), 2020, pp. 416-421, doi: 10.1109/BioRob49111.2020.9224328.

[35] Su, H., Li, Z., Li, G., & Yang, C. (2013). EMG-based neural network control of an upper-limb power-assist exoskeleton robot. Advances in Neural Networks – ISNN 2013, 204–211. https://doi.org/10.1007/978-3-642-39068-5_25

[36] C. Cipriani et al., "Online Myoelectric Control of a Dexterous Hand Prosthesis by Transradial Amputees," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 19, no. 3, pp. 260-270, June 2011, doi: 10.1109/TNSRE.2011.2108667

[37] U. Raghavendra et al., "Automated screening of congestive heart failure using variational mode decomposition and texture features extracted from ultrasound images", Neural Computing and Applications, vol. 28, no. 10, pp. 2869-2878, 2017. Available: 10.1007/s00521-017-2839-5 [Accessed 21 March 2022].

[38] H. Ge, G. Chen, H. Yu, H. Chen and F. An, "Theoretical Analysis of Empirical Mode Decomposition", Symmetry, vol. 10, no. 11, p. 623, 2018. Available: 10.3390/sym10110623 [Accessed 23 February 2022].

[39] S. Lahmiri and A. Shmuel, “Variational mode decomposition based approach for accurate classification of color fundus images with hemorrhages,” Optics & Laser Technology, vol. 96, pp. 243–248, 2017.

[40] U. Raghavendra et al., "Automated screening of congestive heart failure using variational mode decomposition and texture features extracted from ultrasound images", Neural Computing and Applications, vol. 28, no. 10, pp. 2869-2878, 2017. Available: 10.1007/s00521-017-2839-5 [Accessed 21 March 2022].

[41] S. Becker, S. von Werder, A. Lassek and C. Disselhorst-Klug, "Time-frequency coherence of categorized sEMG data during dynamic contractions of biceps, triceps, and brachioradialis as an approach for spasticity detection", Medical & Biological Engineering & Computing, vol. 57, no. 3, pp. 703-713, 2018. Available: 10.1007/s11517-018-1911-3 [Accessed 31 January 2022].

[42] A. Waris, M. Zia ur Rehman, I. Niazi, M. Jochumsen, K. Englehart, W. Jensen, H. Haavik and E. Kamavuako, 2020. A Multiday Evaluation of Real-Time Intramuscular EMG Usability with ANN. Sensors, 20(12), p.3385.

[43] P. Shull, S. Jiang, Y. Zhu and X. Zhu, "Hand Gesture Recognition and Finger Angle Estimation via Wrist-Worn Modified Barometric Pressure Sensing", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 4, pp. 724-732, 2019. Available: 10.1109/tnsre.2019.2905658.

[44] F. Xiao, Y. Wang, Y. Gao, Y. Zhu and J. Zhao, "Continuous estimation of joint angle from electromyography using multiple time-delayed features and random forests", Biomedical Signal Processing and Control, vol. 39, pp. 303-311, 2018. Available: 10.1016/j.bspc.2017.08.015 [Accessed 1 February 2022].

[45] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[46] Javaid, H., Tiwana, M., Alsanad, A., Iqbal, J., Riaz, M., Ahmad, S. and Almisned, F., 2021. Classification of Hand Movements Using MYO Armband on an Embedded Platform. Electronics, 10(11), p.1322

[47] H. Ashraf, A. Waris, S. O. Gilani, M. U. Tariq and H. Alquhayz, 2021. Threshold Parameters Selection for Empirical Mode Decomposition-Based EMG Signal Denoising. Intelligent Automation & Soft Computing, 27(3), pp.799-815.

[48] A. Matic, V. Osmani and O. Mayora, "Speech activity detection using accelerometer," 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 2012, pp. 2112-2115, doi: 10.1109/EMBC.2012.6346377.