

Received 12 January 2024, accepted 17 January 2024, date of publication 22 January 2024, date of current version 30 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3356597

RESEARCH ARTICLE

A Fusion of EMG and IMU for an Augmentative Speech Detection and Recognition System

UZMA SHAFIQ¹, ASIM WARIS¹, (Member, IEEE), JAVAID IQBAL¹, (Member, IEEE),
AND SYED OMER GILANI², (Senior Member, IEEE)

¹School of Mechanical and Manufacturing Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan

²Department of Electrical, Computer and Biomedical Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Corresponding author: Asim Waris (asim.waris@smme.nust.edu.pk)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Ethical Committee of National University of Sciences and Technology under Application No. ref#NUST/SMME-BME/REC/000424/23012022, and performed in line with the Declaration of Helsinki.

ABSTRACT Subvocal voice recognition is important in rehabilitation because it allows people with speech issues to communicate in an alternate manner. Capturing and interpreting subtle muscular movements during subvocalization can improve the rehabilitation process, allowing patients to have more autonomy and a higher quality of life. The scarcity of research on subvocal voice recognition, as well as the use of IMU (accelerometer and gyroscope) signals for speech activity detection, are significant challenges. The study focuses on the amalgamation of spectrotemporal feature extraction techniques for classification and IMU data for speech activity detection. The study carries out the classification of an isolated word vocabulary sets of 70 words corresponding to 6 subjects and 96 isolated words for subjects. A feature extraction algorithm was used utilizing Variational Mode Decomposition. The results demonstrated maximum accuracy rates of 98.6% for the 70-word set and 92% for the 96-word set. Automatic speech activity detection (ADSAS) was developed using only IMU data and was independent of the EMG data. The comparison was carried out by using a previously proposed EMG activity detection method based on Teager Kaiser Operator and morphological operations. The IMU based detection technique achieved a lowest error of 0.09 compared to 0.21 for the EMG based detection technique. The results proved that IMU based methods perform better than EMG based methods for detection of speech activity. Statistical analysis was carried out using paired t-test and resulted in a significant difference between the two techniques (p -value <0.05).

INDEX TERMS Subvocal speech recognition, activation detection, variational mode decomposition, IMU, EMG.

I. INTRODUCTION

Automatic Speech Recognition (ASR) System enables people to use their voices to speak with a computer interface in a way that resembles typical human conversation. ASRs use artificial intelligence (AI) or machine learning to convert spoken language into readable text. Over the past ten years, the discipline has experienced exponential growth, with ASR systems being widely used for numerous everyday applications [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Ghulam Muhammad¹.

Although ASR systems are being widely used due to their high accuracy but there are certain limitations as well [3]. ASRs fail to address disordered speech and due to any speech impairment or disease [4], [5]. A person with some kind of speech related disability, such as verbal dysarthria, apraxia or a person who has undergone laryngectomy, will be unable to use an ASR system. The World Health Organization (WHO) reported that 40% of the newly reported cases of oral cancer globally are from South Asia [6]. This leads to impairment of vocal cords and laryngectomy in some cases thus resulting in loss of communication. Elderly people, due to their issues with speech pace and articulation do not have high success with ASR systems [7]. Apart from that, the use of

ASR systems in quiet settings such as libraries is disturbing. Furthermore, if a secret message or any secret password such as in identification systems need to be delivered, ASR systems are not the suitable choice. The performance of an ASR system is highly affected by the presence of background noise due to harsh acoustic background.

The limitations urge us to develop a new system that does not use acoustic signals. A silent speech recognition system, based on whispered or Alaryngeal speech, independent of the acoustic signals, can perform well in noisy environments. As such a system does not rely on acoustic data, the system can be a viable method for rehabilitation of individuals with speech impairments. Due to an alternate signals source, the system maintains the secrecy of the message as well as addressing issues of speech pace in elderly and people with speech disorders.

Electromyography (EMG) signals can be obtained from muscles both invasively and non-invasively and are termed as surface EMG (sEMG) and intramuscular EMG (iEMG). Both sEMG and iEMG signals differ in characteristics and are widely used in multiple applications [8], [9]. Articulators are present on the face as well as the neck resulting in generation of electrical signals due to the contraction of muscles [10]. Silent speech recognition is one of the emerging applications of EMG signals. Silent speech means unvoiced speech, in which no audible signal is obtained.

A lot of research has been carried out for classification of isolated words, numbers, and small phrases. Ratnovsky et. al classified three and five-word subsets with a maximum accuracy of 88.8% and 74.6%, respectively [11]. Lee et. al classified 60 isolated words with a maximum accuracy of 87.07% [12]. Kumar et. al classified the 5 vowels with an accuracy of 92% [13]. References [11], [12], [13] carried out classification of only isolated words. Janke et. al and Vojtech et. al classified continuous speech but failed to report any activity detection method [14], [15].

Furthermore, the classification was carried out using either only temporal information or Mel frequency cepstral analysis. Mel frequency cepstral analysis has been widely used in acoustic speech recognition. Although it performs well for phoneme-based classification of subvocal speech [16], yet this method does not perform well for higher number of classes in case of isolated words classification. Thus, the study utilizes a new feature extraction technique based on Variational mode decomposition (VMD). Dragomiretskiy and Zosso proposed a new decomposition technique for signals carried out in both temporal and spectral domain [17]. Dragomiretskiy and Zosso also showed the robust nature of VMD for a tri harmonic signal effected by noise. VMD has been widely used previously for extracting spectral and temporal features using VMD for neuromuscular disorder detection [18], physical actions classification [19] and in detection of different fatigue conditions [20].

Meltzner et. al carried out a study for continuous speech data and achieved an accuracy of 86% for 4 sensors set [16].

One of the major issues in continuous speech is the detection of speech activity. Meltzner developed a speech activity detection system that detected continuous speech using statistical information of EMG data from each channel. This decision from each channel was then fed into a finite state machine to give a global decision regarding speech activity [16].

Although Meltzner's method performed well for detection of speech activity, it does not provide any measure for the start and end of a new word. Moreover, it is only suitable if the classification carried out is phoneme based. For a system that is generalizable and suitable for both phonemes based as well as isolated word classification extracted from continuous speech a more robust system needs to be developed. Meltzner's method used third order statistic which utilized the time domain information only which leads to a computationally less expensive architecture but is prone to poor performance in case of high levels of noise. The EMG activity of facial muscles is lower as compared to bigger muscles of the body such as muscles in upper and lower limb thus resulting in signals with lower amplitude. The signal is more prone to noise as well. Thus, a more robust activity detection method is required that addresses all these issues.

The Teager-Kaiser Energy (TKE) operator was used in speech signals for non-linear components of acoustic data [21], [22]. Later, its significance in onset and offset detection of EMG signals was discovered [23], [24]. Reference [25] proved that TKE improves EMG onset detection. Yang et Al. used Teager Kaiser Energy Operator (TKEO) combined with morphological operations to detect EMG activity in weak pathological signals [26]. Although the proposed method performs well for simulated EMG signals with lower number of motor units, the study fails to evaluate its performance on real EMG signals. Reference [27] used the method for real time onset detection of EMG signals. Reference [28] used the method to identify periodic bursts in EMG. Previous literature shows multiple uses of TKEO activation detection method used for comparison [29].

Although the use of TKEO method resolves the issue of activity detection being carried out using only temporal information, a few more issues persist. The EMG based speech activity detection assumes that facial muscles go back to rest position during the resting state and EMG amplitude falls below a threshold value. However, this is not true for all cases, such as failing to bring the muscles to rest after articulation. Furthermore, the subconscious movement of jaw muscles can also lead to generation of EMG signals which can result in false activity detection.

To resolve these issues, a new system needs to be developed which is independent of EMG and accurately detects speech activity. Inertial Motor Unit signals have been widely used for activity recognition for a variety of purposes thus it can also be used for developing an alternate method for speech activity detection [30], [31], [32].

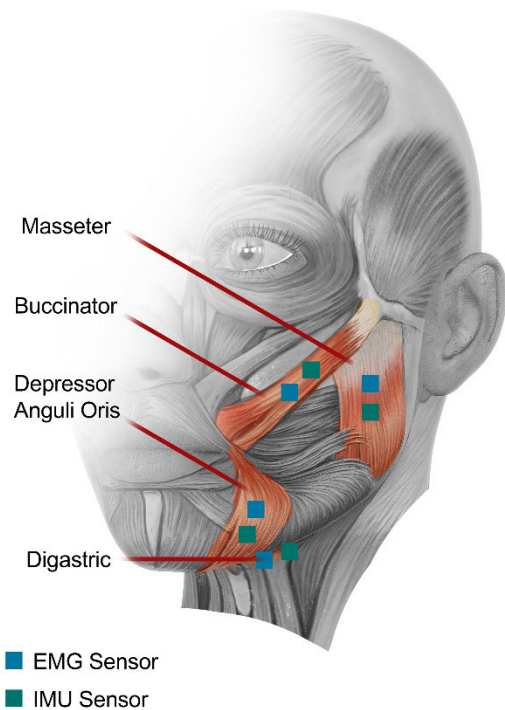


FIGURE 1. Electrode placement on muscles of interest. (Blue: EMG, Green: IMU).

This study serves two purposes. The first is to classify isolated words with an increased vocabulary set and lesser number of channels. The second is to develop an algorithm for Automatic Detection of Speech Activity System (ADSAS). The ADSAS utilized IMU signals obtained simultaneously from the facial muscles along with EMG signals. The paper is structured as follows: Section II provides information about the Data corpus, the subjects, feature extraction method and classifier for isolated words vocabulary and the ADSAS method. The results for both objectives (classification of isolated words and detection of speech activity) are reported in section III. Section IV discusses the findings in detail and highlights their significance. Section V concludes the study.

II. METHODOLOGY

A. DATA AND EXPERIMENTAL PROCEDURE

Informed consent from all the subjects was obtained before the start of the experiment. The data collection was started on 26th December 2022 and ended on 20th January 2023. All the subjects were adults and provided their written consent. The data collection was carried out as per the relevant guidelines and regulations. Ethical approval for this study was obtained under Approval No.: ref#NUST/SMME-BME/REC/000424/23012022 from local ethical committee of National University of Sciences and Technology.

Data were collected from 10 healthy subjects (25.4 ± 2.3). Before carrying out the experiment, a signed informed consent was obtained from all the subjects. The subjects selected were healthy and did not show any injury to the muscles

of interest. The selection criteria for participants were maintained. The criteria entailed that the subject must have healthy facial muscles and the subjects do not have any condition that effects the motion of the facial muscles such as facial palsy, facial paralysis, paresis (partial paralysis) and did not go under any kind of surgery. Moreover, it was also made sure that the participant did not undergo under any aesthetic procedures such as injectable filters, facial implant, and facial rejuvenation.

As the facial muscles are small and the activation amplitude is lower as compared to the muscles of the limbs, it was made sure that the selected muscles are the largest muscles on the face that are activated during speech activity. Four muscles were selected considering this criterion and the optimum muscles selected by [16]. These muscles include Buccinator, Masseter, Depressor anguli oris and Digastric Muscle and the placement of electrodes is shown in Fig 1.

Four bipolar Delsys trigno electrodes were placed on the above-mentioned muscles. The electrodes were used to record EMG data from six subjects and EMG and IMU data simultaneously for the remaining four. The IMU data gave the muscle's orientation in 3D space. The IMU data is based on three sensors: accelerometers, gyroscopes, and magnetometers. The accelerometer detects the changes in velocity of muscle movement. The gyroscopes measure angular velocity, which is the rate of change of an object's orientation over time. They provide information about the rotation or angular motion around multiple axes. The magnetometers are used to measure the force and direction of the magnetic field around the sensor. They provide information about the orientation of the IMU relative to the Earth's magnetic field. By combining the data from these sensors, an IMU estimates the object's orientation in 3D space, velocity, and position.

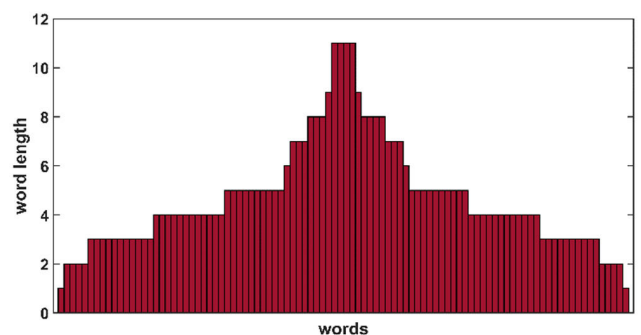


FIGURE 2. Data corpus words length.

The words in the data corpus were selected on following criteria: 1. The English alphabets should be uniformly distributed. 2. The words should be a good representation of variation of length in English words as shown in Fig 2.

Since the data collection for subvocal speech requires electrode placement on the face, it can be uncomfortable for the subject to record the data in one sitting. But recording the data over the span of different days can result in variations in the electrode placement and the activation levels depending

TABLE 1. Isolated vocabulary dataset (Sn represents nth subject).

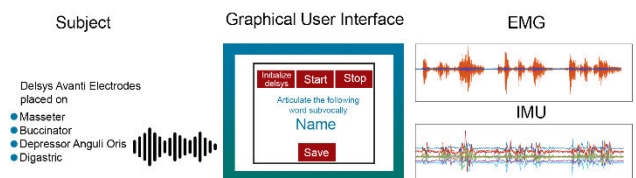
Digits S1-S10	Directions S1-S10	Commands S1-S10	Conjunctions S1-S10	Pronouns S1-S10	Random S1-S10	Additional S1-S10	Questions S7-S10	Helping verbs1 S7-S10	Miscellaneous S7-S10
Zero	Left	Start	Accordingly	I	Alpha	Name	What	Is	Have
One	Right	End	Actually	He	Bravo	Behavior	How	Am	May
Two	Forward	Synchronize	After	She	Charlie	Rewarded	Why	Are	Might
Three	Backward	Help	Further	It	Delta	Destination	When	Be	At
Four	Straight	Wait	Nevertheless	You	Angle	Please	Where	Been	A
Five	Reverse	Good	However	Your	Feet	Position	Who	Could	An
six	North	Bad	But	They	Meter	Move	Which	Do	the
Seven	South	Yes	Thus	We	Sideways	Push	Whom	Did	
Eight	East	No	Therefore	Him	Front	Pull	Whose	Does	
nine	west	turn	meanwhile	Her	tilt	tired	Can	has	

on the energy of the subject. Thus, an optimum size for vocabulary was selected that is large enough to get novel and substantial results but is also not too straining for the subject during data collection. A vocabulary set of 70 words was selected for 6 out of 10 subjects. These words were divided into 7 groups with ten words each. For the remaining 4 subjects the vocabulary set was increased, and 96 isolated words were included in the dataset. Continuous speech data was also collected from these subjects. Continuous speech data comprised of 60 sentences from each subject based on the vocabulary from the 96 words dataset. The details of the vocabulary are given in Table 1.

Four electrodes were placed on the muscles of interest (Buccinator, Masseter, Depressor anguli oris and Digastric). The electrodes were placed using an adhesive double-sided tape. To secure the electrode and skin contact surgical tape was used externally above the electrode. Since the experiment lasted for about 4-6 hours per subject, the double tapes and the external surgical tape were replaced after every one hour (after articulation of two groups). The subject was seated in a comfortable chair and a practice session was carried out. The practice session entailed articulating each word subvocally that was visible on the screen.

The vocabulary was prompted on the screen, one word at a time. 5 seconds were given for articulation of each word. For six subjects only EMG data were collected but for the remaining four EMG and IMU data were collected from the four muscles simultaneously. The data were collected at a sampling rate of 2000 Hz for EMG and 148Hz for IMU. Data were collected using MATLAB 2020a. A separate graphical user interface (GUI) was designed for data collection. This GUI prompted each word on the screen to be articulated subvocally. The words appearing on the screen were prompted randomly to avoid any bias. The data for each group of words was collected separately and twenty iterations for each

group were recorded. 5 second rest was allowed between each articulation to avoid any muscle fatigue. Thus, recording of each group for twenty iterations took approximately 30 minutes. A 5–10-minute break was given between recordings of each group. During the break the subject was allowed to sip water or any room temperature drink of choice. For subjects 1-6 a vocabulary set of 70 words the total time taken was 4 hours. (3.5 hours for 7 groups and 30 minutes for breaks in between the seven groups). The vocabulary set increased for subjects 7-10 thus the data and time also increased. The time for subjects 7-10 for a vocabulary of 96 words amounted to approximately 6 hours. Thus, the total data collection time for 10 subjects was 48 hours. The data collection procedure was spread over 10 days. Each day the data was collected from one subject only.

**FIGURE 3.** Experimental setup for data recording.

The experimental setup is depicted in fig 3. The collected EMG data were preprocessed using a band pass filter with cut-off frequencies of 10Hz to 450Hz and a notch filter of 60Hz.

B. FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUE

The feature extraction method utilizes an advanced signal processing technique called Variational Mode decomposition (VMD) proposed by Dragomiretsky and Zosso [17].

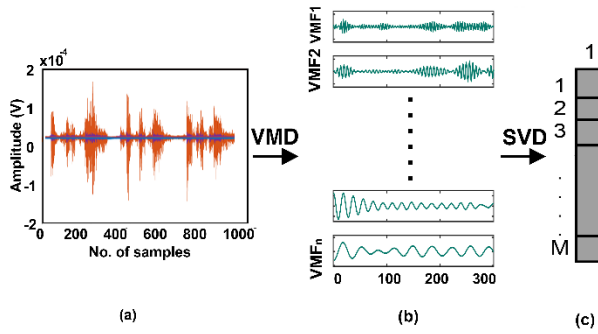


FIGURE 4. a) Raw EMG signal, (b) Decomposed variational mode functions after applying variational mode decomposition (c) dimensionality reduction using singular value decomposition.

VMD decomposes the signal into its subsequent sub signals called Variational mode functions (VMF). Each VMF comprises of several data points thus, a scheme of feature reduction is applied to make the classifier simpler. VMFs are used to obtain the eigen values. The eigen values are further used to achieve a singular value corresponding to each VMF. These singular values are used as features to carry out the classification of isolated words vocabulary. Since 12 VMFs are obtained for each signal and 4 channels are utilized thus a feature vector of $(12 \times 4) \times n$ is obtained where n is the number of samples. The process is depicted in Fig 4.

As the classification problem is like pattern recognition and LDA is considered as the gold standard, yet it was proved that Random Forest (RF) performs better than LDA [33]. Reference [34] utilized KNN for controlling a dexterous artificial hand and achieved better results than Support Vector Machines (SVM). Thus, we utilized KNN and RF for classification of the isolated word vocabulary.

C. AUTOMATIC SPEECH ACTIVITY DETECTION SYSTEM (ADSAS)

Speech signals collected from four subjects with a vocabulary set of 96 words comprised of data from both EMG as well as IMU. The continuous subvocal speech data collected from these subjects was used to develop an algorithm for detection of speech activity using IMU data. This algorithm is named Automatic Speech Activity Detection System (ADSAS). ADSAS utilizes the IMU signals (accelerometer and gyroscope data) and detects the regions with maximum activity for detecting speech regions. The algorithm is based on following stages:

Stage 1: Activation Detection:

The following steps are involved in the detection of activations at all the samples of input signal.

1. Calculate variance vector using a window of 30ms.
2. Calculate the threshold value for all the points. The threshold value is calculated by taking 80% of the mean value of the variance vector.

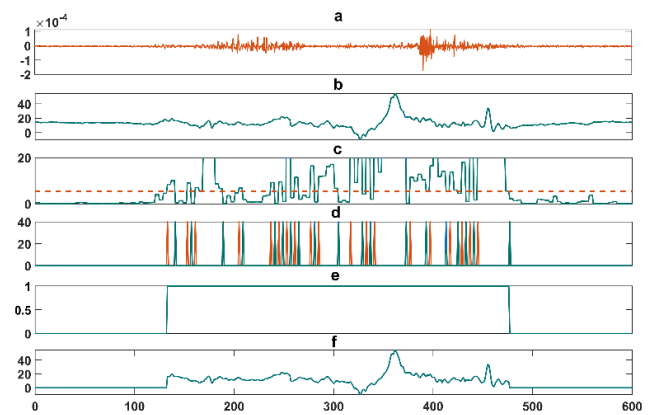


FIGURE 5. a) speech EMG signal (b) Corresponding Input IMU signal. (c) variance vector(blue) and threshold value (red), the figure is zoomed in to depict the thresholding step clearly. (d) Onsets (red) and offsets (blue). (e) Active region binary vector.

3. Carry out thresholding. A binary vector with zeros and ones is obtained as a result. Any values less than the threshold are substituted by zeros and greater by ones.
4. This binary vector gives the activations of the input signal at all the samples.
5. The resultant activation vector is then used to detect the points of onset and offset of the signal.
6. The points of onsets and offsets are then used to join the shorter event for detecting speech region.

This step is carried out for all the six signals (accelerometer (x, y, z) and gyroscope (x, y, z)). As four channels are used thus this step is carried out for 24 signals. As evident in Fig 5.

Stage 2: Computing Region of Activation:

After finding the activation region across all the axes (x, y, z) the next step was to compute the combine activation region of all the three axes. This step included following steps:

1. The activation regions from all three axes are obtained from the previous step. The common activation region between any two axes contributes to the overall activation. The intersection of all three combinations of axes is computed and the union of resulting intersection of the three combinations is considered as the activation for the corresponding channel. The condition to be implemented is given by (1).

$$ROA = (x \cup y) \cap (x \cup z) \cap (y \cup z) \quad (1)$$

where ROA is the region of activation, x, y, and z are IMU signals along x, y, and z axes.

2. The next step is to remove any small fluctuations in data that resemble the activation region but are unwanted fluctuations occurring due to the unnecessary jaw movement. As evident in figure 5d, smaller fluctuations are present in the computed ROA. These fluctuations are not a part of the speech signal and can occur due to subconscious jaw movement by the subject. This issue has been resolved by morphological

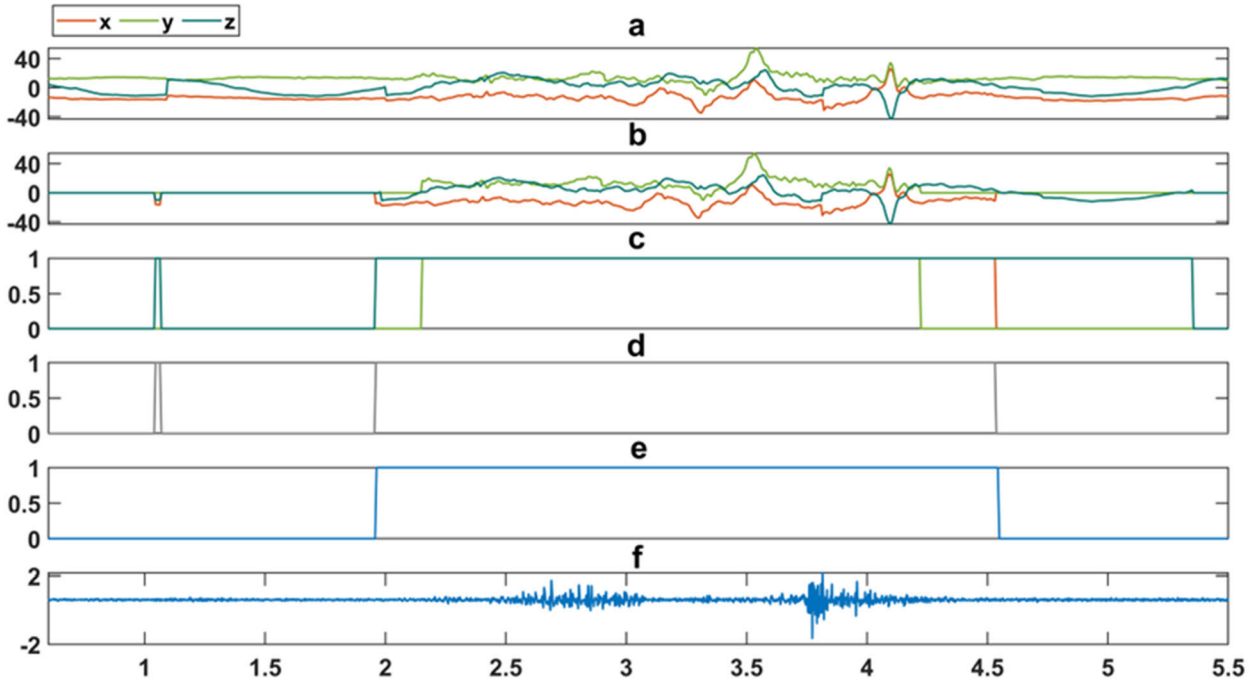


FIGURE 6. (a) Raw IMU signal for all three axes. (b) Filtered IMU signals for all three axes. (c) Detect activation region for all three axes. (d) Computed ROA in the channel with fluctuations. (e) ROA with fluctuations removed after opening. (f) Raw EMG.

erosion technique. It enables us to get rid of any region smaller than the optimum window for speech signal. Mathematically it can be written as (2).

$$ROA \ominus S = \cap_{a \in S} ROA_{-a} \quad (2)$$

where ROA is the region of activation, S is the structuring element and where ROA_{-a} denotes the translation of ROA by -a.

3. After removing the unwanted fluctuations from the data, we obtain the activation region for the corresponding channel.

The steps have been depicted in Fig 6.

Stage 3. Computing Complete Activation Region (CAR)

As mentioned previously the IMU signals consists of both accelerometer and gyroscope data along each of the three axes, x, y, and z. Thus, a total of six signals are present in the IMU signal for each channel.

Obtaining a single vector for activation region from all three axes of accelerometer and gyroscope has been carried out in the previous step. Consequently, it gives us two activation signals, one corresponding to each (accelerometer and gyroscope). In this step we compute a single activation region using the two ROAs obtained in stage II. This stage is governed by condition in (3).

$$CAR = ROA_{acc} \cap ROA_{gyro} \quad (3)$$

As evident from Fig 7 the intersection lasts till 520 where activity region for both ROA_{acc} and ROA_{gyro} gyro is

common. Although ROA_{acc} extends beyond the common region but it is not included in the Combined activation region (CAR).

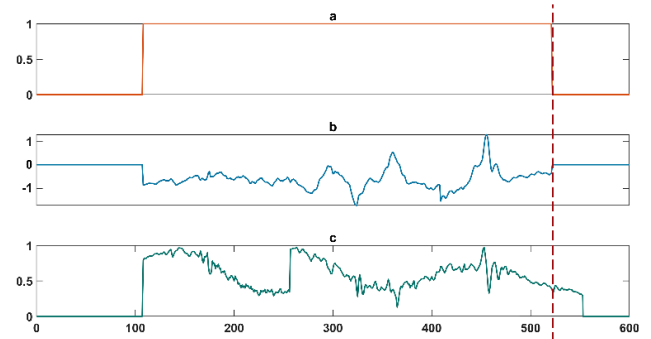


FIGURE 7. (a) Complete Activation Region (CAR) (b) filtered gyroscope data (ROA_{gyro}) (c) filtered accelerometer data (ROA_{acc}).

Stage 4: Computing Global Activation Region (GAR) for 4 Channels:

The final stage involves computing the global activation region (GAR) from all 4 channels. This step is governed by the condition, if three channels are active simultaneously and have $CAR > 0$ then that region is considered as GAR. The number of combinations for 4 channels out of which 3 are active simultaneously can be computed by (4).

$$C_k(n) = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (4)$$

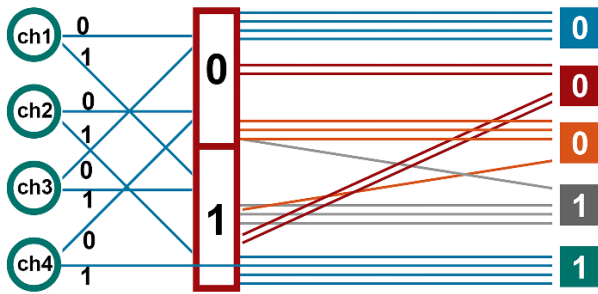


FIGURE 8. Channel states and output.

Thus for 4 channels in total and 3 channels active simultaneously the number of combinations can be 4. Mathematically the condition to be fulfilled can be written as (5). signal and any values outside of GAR are substituted by zeros. The different states and the resulting output corresponding to the input from all 4 channels is depicted in Fig 8. Fig 9 shows the resultant GAR. The computed GAR is used to filter the EMG.

$$\text{GAR} = \begin{cases} 1 & \text{if } \text{CAR}_{\text{chm}}, \text{CAR}_{\text{chn}}, \text{CAR}_{\text{cho}} > 0 \\ 0 & \text{if } \text{CAR}_{\text{chm}}, \text{CAR}_{\text{chn}}, \text{CAR}_{\text{cho}} = 0 \end{cases} \quad (5)$$

where $n, m, o \in \{1, 2, 3, 4\}$ (without duplication)

The different states and the resulting output corresponding to the input from all 4 channels is depicted in Fig 8. Fig 9 shows the resultant GAR. The computed GAR is used to filter the EMG signal and any values outside of GAR are substituted by zeros.

Performance Evaluation:

The selected evaluation metrics for the evaluation of the performance of the implemented method for classification were percentage accuracy as all the classes were balanced in the isolated data corpus. For IMU based speech activation algorithm we used Activation Error Rate (AER). AER is calculated as given in (6).

$$\text{AER} = \frac{\text{Detected Activation}}{\text{Original Activation}} \quad (6)$$

III. RESULTS

Two different types of vocabulary sets were collected from 10 subjects. Six subjects provided the data for a vocabulary set of 70 words. The remaining 4 subjects provided the data for a vocabulary set of 97 words as well as continuous data based on the vocabulary of these 97 words. The aim was to classify the individual words and develop a robust algorithm for this purpose. Moreover, it was also intended to develop a system that detects the speech activity region using the IMU signals rather than EMG signals. The efficacy of this devised algorithm was tested against an EMG based speech activity detection system proposed by [26].

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

A. CLASSIFICATION OF ISOLATED WORD VOCABULARY

Subject-wise classification of 10 healthy subjects using the developed feature extraction algorithm SVD-VMD was carried out. The features were extracted using 12 VMFs for each signal. Table 2 depicts the classification accuracy for each subject using the SVD-VMD technique along with RF and KNN classifier.

As evident from Table 2 the devised feature extraction method along with RF and ANN classifiers performs quite well and results in the highest accuracy of 98.6% for RF and ANN classifiers corresponding to a vocabulary of 70 words. Whereas an accuracy of 92% for RF and 95.2% for ANN is obtained corresponding to a vocabulary set of 96 words. The data is also illustrated in Fig 10.

The accuracy for subjects 3-6 was computed corresponding to a data corpus size of 96 words whereas for the remaining subjects the data corpus size was kept at 70. Subject 5 provides the highest accuracy for 96-word vocabulary corresponding to all three classifiers. Even though the accuracy corresponding to KNN is significantly less than the accuracy obtained corresponding to RF and ANN classifier, it is higher than other subjects with 96 words vocabulary. A similar trend can be seen for subject 10 corresponding to a vocabulary of 70 words.

TABLE 2. Classification accuracies for isolated words.

Subject	Vocabulary (No. of words)	Accuracy (%) KNN	Accuracy (%) RF	Accuracy (%) ANN
1	70	84.5	97.4	98.5
2	70	83.2	96.9	97.1
3	96	73.2	85.0	85.3
4	96	72.0	91.9	94.0
5	96	80.1	92.0	95.2
6	96	68.7	91.2	86.8
7	70	88.0	97.7	98.8
8	70	86.6	98.2	96.8
9	70	82.3	98.1	98.6
10	70	88.9	98.6	98.6

To test the efficacy of the proposed classification algorithm we varied the number of words in vocabulary set as well as the number of channels. The classification accuracy was computed as shown in Table 3 and Fig 11. The depicted accuracies are average accuracies for 10 subjects ($n=10$) and error bars depict the standard deviation. It is evident that the proposed method performs well for classification of 10 words corresponding to all the varied channels. It can be observed that the proposed method provides an accuracy of 85.9% for single channel classification of 10 words. This accuracy decreases with the increase in number of words in the vocabulary set.

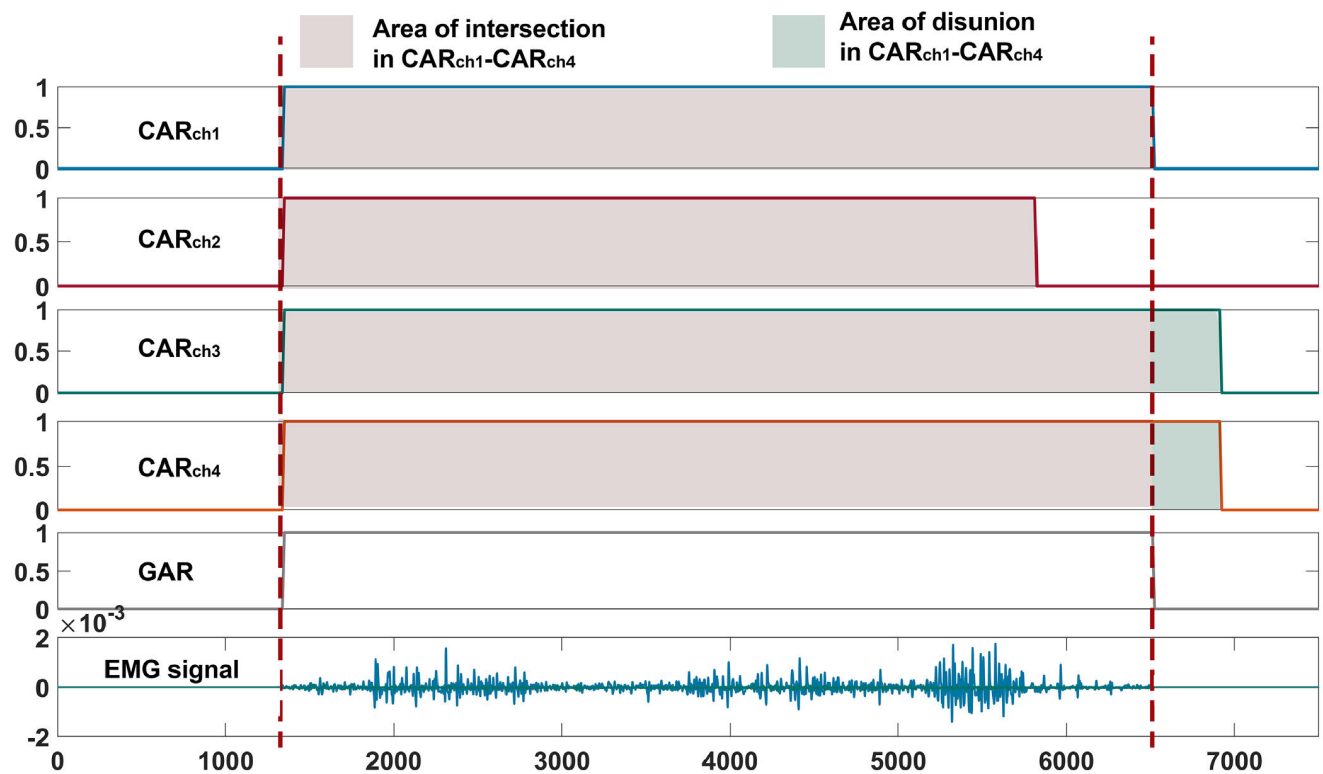


FIGURE 9. Global activation region (GAR) computed from CAR of all 4 channels.

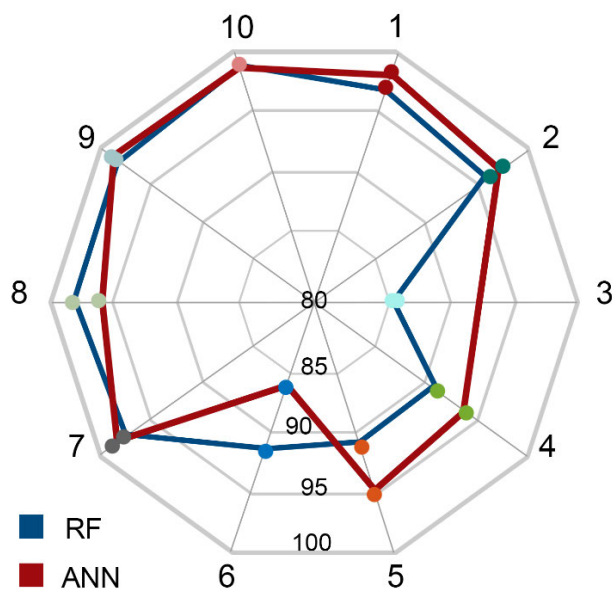


FIGURE 10. Isolated word classification accuracy of 10 subjects.

B. EFFICACY OF AUTOMATIC SPEECH ACTIVITY DETECTION SYSTEM (ADSAS)

To test the efficacy of the proposed classification algorithm we varied the number of words in vocabulary set as well

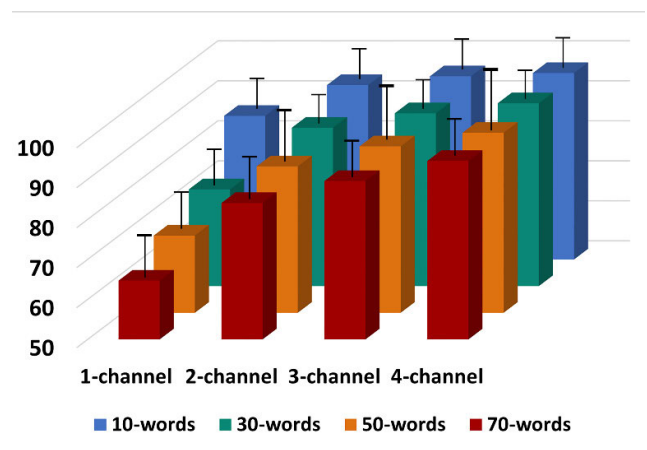


FIGURE 11. % Average accuracies corresponding to varied number of channels and words in data corpus.

as the number of channels. The classification accuracy was computed as shown in Table 3 and Fig 11. It is evident that the proposed method performs well for classification of 10 words corresponding to all the varied channels. It can be observed that the proposed method provides an accuracy of 85.9% for single channel classification of 10 words. This accuracy decreases with the increase in number of words in the vocabulary set.

TABLE 3. Average classification accuracies corresponding to varied number of channels and words in data corpus.

	1-channel	2-channel	3-channel	4-channel
10-words	85.9±8.7	93.6±7.7	95.8±6.4	96.6±7.0
30-words	74.2±8.3	89.6±5.6	93.2±4.9	95.7±4.9
50-words	69.3±11	86.6±6.4	91.6±5.3	95±5.6
70-words	64.7±10.5	84.1±7.0	89.6±5.6	94.6±6.4

Four subjects (S3-S6) provided two types of datasets, one being an individual word recording for 96 words and the other comprising of continuous sentences derived from these 96 words. EMG and IMU data collected simultaneously for these continuous sentences was utilized to develop a system for recognition of speech activity called ADSAS. This system uses IMU signals (accelerometer as well as gyroscope) to detect the presence of speech activity from all four channels.

Table 4 and Fig 12 depict the results obtained using the proposed method. The results are shown in terms of error rate. The table illustrates a comparison of an EMG based method as well as the developed IMU based method for speech detection.

The EMG based method was proposed by [26]. The efficacy of this method had been used by [27], [28], [29] for various purposes. The quantitative comparison of two methods is shown in Table 4 and Fig 9. The qualitative comparison of the two methods is evident in Fig 13.

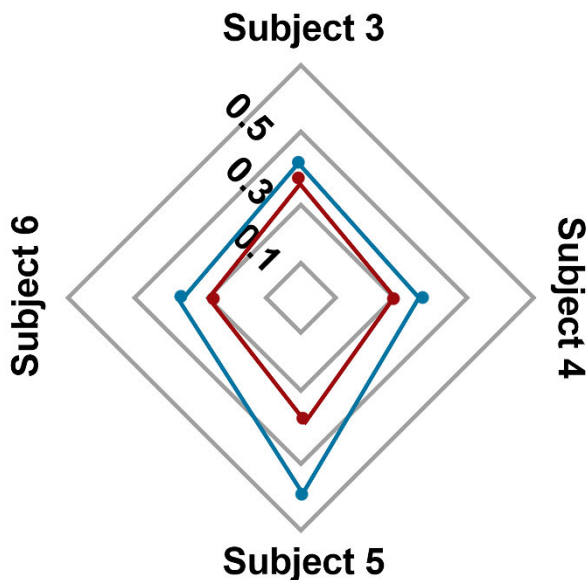


FIGURE 12. Activation error rate for speech activity detection.

As evident from Table 4 and Fig 12 the AER for EMG based activity detection method is higher as compared to that of IMU based activity detection. Subject 4 provided the

lowest value of error 0.085 and subject 5 resulted in the highest value 0.172.

Fig 13 shows a comparison of the detected activity regions corresponding to EMG based and IMU based method.

TABLE 4. Activation error rate for continuous speech activity detection.

Subject	IMU based	EMG based
3	0.16	0.25
4	0.11	0.21
5	0.19	0.40
6	0.09	0.24

IV. DISCUSSION

The paper has two objectives, to provide a robust classification for a larger vocabulary set of isolated words and devise a new algorithm for automatic detection of speech from the IMU signals. Before this study the highest number of isolated words classification was carried out for 67 words. Moreover, the literature does not report a speech detection method using IMU signals. The proposed method automatically detects the speech activity region using statistics of the IMU signals combined with morphological operations. The proposed classification method provides better results than the previously reported methods for classification [11], [12], [13], [14]. Furthermore, the study also reports better results for detection of activation region detection using IMU as compared to EMG [26].

The feature extraction method for classifying the isolated words utilizes VMD. VMD decomposes each individual signal into its subsequent sub signals. A single value is obtained from the eigen values of the obtained sub signals. These single values serve as a single representation of each sub signal. These features were classified using the RF classifier as it performed the best as evident in Table 2. RF is an ensemble learning method which employs several decision trees thus performing efficiently with EMG based classification [35]. It can be observed from Table 2 that the highest accuracy obtained was 98.6% corresponding to a vocabulary set of 70 words. However, the highest accuracy obtained for a vocabulary set of 96 words was 92%. This significant decrease results because of two reasons, first being the increased number of classes and secondly due to data collection of the whole vocabulary set distributed over three sessions. Due to the increased vocabulary set for the four subjects the data was collected over three sessions whereas the data was collected in a single session. The electrode placement can vary a little due to human error in three different sessions. Furthermore, the consistency of muscle activation can also vary in different sessions. Although, with the increase in classes the accuracy has decreased yet it is

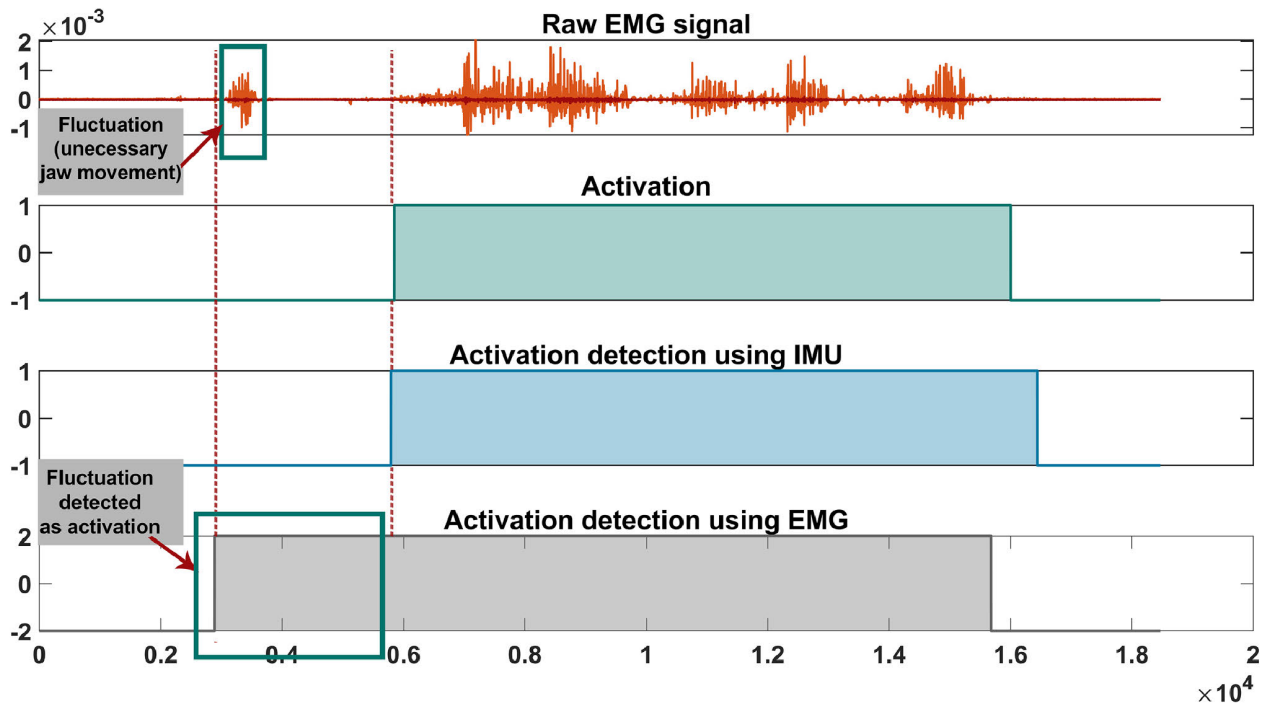


FIGURE 13. Comparison of EMG and IMU based activity detection method.

almost equal to Meltzner's accuracy reported for 67 isolated words using 11 channels whereas the proposed method only utilizes 4 channels [16].

VMD extracts multitone components and spectral bands more effectively and overcomes mode aliasing, resulting in much enhanced anti-noise performance and computing efficacy [17]. The use of a quadratic penalty in VMD to convert a constrained problem to an unconstrained one using a Lagrangian operator preserves the decomposed signal's original features while increasing reconstruction fidelity. As a result, the obtained feature vector becomes more robust resulting in a higher classification accuracy for greater vocabulary set. It has been previously reported in the literature that VMD outperforms MFCC for applications such as Parkinson disease detection thus the obtained results agree with previous literature [36].

The previous literature reports EMG based speech activity detection method but fails to provide a robust method for speech activity detection using IMU signals for subvocal speech detection as evident in Table 5. The EMG based methods assume that during the resting period the EMG amplitude falls below a certain threshold which is not always true. For instance, in articulation of certain words the EMG activation might not be significant. Similarly, if the subject does not bring the muscles back to rest during resting period the facial muscles will still produce EMG signals which can be falsely detected as speech signals. Moreover, the subconscious movement of facial muscles can also result in activation of the muscles. As evident in Table 4 ADSAS

outperforms the EMG based method for all the four subjects. Even though the ADSAS reports different AER for each subject, it is higher than EMG based method.

IMU signals detect the velocity and orientation of the muscle in all the three axes rather than activation of the muscles. The IMU signals are independent of muscle activation thus resolving the issue of subconscious muscle activation.

Reference [37] utilized IMU signals for speech activity detection for acoustic signals and proved that it is a cost effective as well as an easily applicable method. The qualitative analysis of the proposed method is demonstrated in Fig 13. It is demonstrated that the EMG based activity detection method assumes the fluctuation caused by unnecessary jaw movement as a part of the speech activity whereas the IMU based activity detection method disregards this fluctuation thus providing a lower AER. IBM SPSS statistics 20 was used to run paired t-test for the statistical analysis of the results. The test proved that a significant difference exists between the two methods ($p\text{-value} = 0.0398$).

Although VMD-based classification is an efficient method, it has its own drawbacks. The biggest drawback of using VMD-based feature extraction is its high dimensionality leading to a higher time complexity. In the future dimensionality reduction techniques need to be introduced to reduce the dimensions of the feature vector so that the computational time of the system can be reduced.

One of the biggest limitations of this method is that the data collection process can get tiresome due to larger vocabulary sets thus requiring the data collection to be distributed into

TABLE 5. Comparison of proposed method with previous literature.

Vocabulary	Method	Accuracy
5 isolated words	Kumar et. al [13]	92%
	Ratnovsky et. al [11]	74.6%
	Proposed method	97.44%
7 isolated words	Ratnovsky et. al [11]	49.6%
	Proposed method	96.8%
60 isolated words	Lee et. al [12]	88.13%
	Proposed method	95.05%
67 isolated words	Meltzner et. al [16]	92.3%
	Proposed method	94.6%
97 isolated words	Proposed method	92%

numerous sittings. This can result in comprising the accuracy of the results. Furthermore, when faced with variances in speaking styles, speaking situations, or user demographics, subvocal speech recognition algorithms frequently suffer with robustness and generalizability. For effective deployment, robustness and generalizability across varied user populations and real-world contexts are critical.

One of the most essential articulation muscle groups is the tongue. The positioning of the tongue is the only trait that distinguishes specific words. However, the location of EMG sensors on the skin's surface prevents immediate approach to the tongue. Sensor placed beneath the chin on the digastric muscle, may be recording some tongue activity. It is the only non-facial sensor that was discovered as the most effective sensor for speech recognition [16]. However, there is a substantial amount of non-lingual muscle and tissues between the sensor and the tongue, which limits the crucial information acquired. Other methods such as ultrasound [38] impulse radio ultra-wide band (IR-UWB) radar [39], and permanent magnet articulography (PMA) [40] been used previously that utilize different techniques to capture the information from tongue. These methods can be effective, but they will increase the computational and time complexity of the developed system.

Table 5 represents the comparison of proposed method with previous literature. As evident, this study provides the highest accuracy corresponding to 70 words. Even though [16] and [13] provide comparable accuracies of 92% however, the vocabulary size for the two studies is 67 words and 7 words respectively, which is lesser than the vocabulary size of this study. Furthermore, it can be observed none of the literature reports an IMU based activation detection method.

Since the scope of this study was to develop a robust method for classifying larger isolated words vocabulary and developing a system for detection of speech activity, the classification of continuous sentences was not carried out. In the future we plan to improve the ADSAS method enabling it to detect each word and then classify these extracted words for classifying the continuous vocabulary set. To deploy the speech recognition using EMG it is vital to have a larger vocabulary set for both isolated words and continuous sentences. In the future we also aim to incorporate a more extensive data set collected over multiple days and test the performance of the proposed method with an increased vocabulary set. Since the study aimed primarily on the development of a speech recognition and activity detection method and utilized healthy subjects only thus, we aim on testing the proposed method for impaired individuals as well in the future.

V. CONCLUSION

The study aimed to classify a wider vocabulary set of isolated words comprising 96 words at maximum and 70 words at minimum. The study carried out the classification of isolated vocabulary using feature extraction based on an advanced signal processing technique called Variational Mode Decomposition. This provided spectral as well as temporal information of the speech signals. The study provided higher classification accuracy than the previous studies. The second objective of the study was to develop an algorithm for automatic detection of speech activity using IMU signals rather than EMG signals. The proposed method computes the local statistics of IMU signals and combines the thresholding process with morphological process to detect the activation region. Results have shown that the proposed activity detection system performs better than the EMG based activity detection method.

ACKNOWLEDGMENT

The 3D rendered model for figure 1 was created using assets from freepik.com.

REFERENCES

- [1] S. Tomko, T. K. Harris, A. Toth, J. Sanders, A. Rudnick, and R. Rosenfeld, "Towards efficient human machine speech communication," *ACM Trans. Speech Lang. Process.*, vol. 2, no. 1, pp. 1–27, Feb. 2005, doi: [10.1145/1075389.1075391](https://doi.org/10.1145/1075389.1075391).
- [2] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, and G. Zweig, "Achieving human parity in conversational speech recognition," 2017, *arXiv:1610.05256*.
- [3] F. Chen and A. Kostov, "Optimization of dysarthric speech recognition," in *Proc. IEEE EMBS Conf.*, vol. 4, Oct. 1997, pp. 1436–1439.
- [4] P. Green, J. Carmichael, A. Hatz, P. Enderby, M. Hawley, and P. Mark, "Automatic speech recognition with sparse training data for dysarthric speakers," in *Proc. Eur. Conf. Speech Commun. Technol.*, 2003, pp. 1189–1192.
- [5] J. Noyes and C. Frankish, "Speech recognition technology for individuals with disabilities," *Augmentative Alternative Commun.*, vol. 8, no. 4, pp. 297–303, Jan. 1992.
- [6] J. Ferlay, F. Bray, P. Pisani, and D. M. Parkin, *Globocan 2002: Cancer Incidence, Mortality and Prevalence Worldwide, Version 2.0*. Accessed: Dec. 22, 2004. [Online]. Available: <http://www.depdbiarc.fr/globocan2002.htm>

- [7] F. Aman, M. Vacher, S. Rossato, and F. Portet, "Analysing the performance of automatic speech recognition for ageing voice: Does it correlate with dependency level?" in *Proc. 4th Workshop Speech Lang. Process. Assistive Technol.*, 2013, pp. 9–15.
- [8] D. Farina, D. F. Stegeman, and R. Merletti, "Biophysics of the generation of EMG signals," in *Surface Electromyography: Physiology, Engineering, and Applications*. Hoboken, NJ, USA: Wiley, 2016, pp. 1–24.
- [9] I. Campanini, C. Disselhorst-Klug, W. Z. Rymer, and R. Merletti, "Surface EMG in clinical assessment and neurorehabilitation: Barriers limiting its use," *Front Neurol.*, vol. 11, Sep. 2020, Art. no. 556522, doi: [10.3389/fneur.2020.00934](https://doi.org/10.3389/fneur.2020.00934).
- [10] S.-C. Jou, T. Schultz, M. Walliczek, F. Kraft, and A. Waibel, "Towards continuous speech recognition using surface electromyography," in *Proc. Interspeech*, Sep. 2006, pp. 17–21.
- [11] A. Ratnovsky, S. Malayev, S. Ratnovsky, S. Naftali, and N. Rabin, "EMG-based speech recognition using dimensionality reduction methods," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 1, pp. 597–607, Jan. 2023, doi: [10.1007/s12652-021-03315-5](https://doi.org/10.1007/s12652-021-03315-5).
- [12] K. S. Lee, "EMG-based speech recognition using hidden Markov models with global control variables," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 930–940, Mar. 2008, doi: [10.1109/TBME.2008.915658](https://doi.org/10.1109/TBME.2008.915658).
- [13] S. Kumar, Z. Ali, C. Kumar, G. Abid, S. A. Shaikh, and V. Memon, "Speech recognition based robotic mart," *Int. J. Intell. Robot. Appl.*, vol. 4, no. 3, pp. 342–353, Sep. 2020, doi: [10.1007/s41315-020-00144-1](https://doi.org/10.1007/s41315-020-00144-1).
- [14] M. Janke and L. Diener, "EMG-to-speech: Direct generation of speech from facial electromyographic signals," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 12, pp. 2375–2385, Dec. 2017, doi: [10.1109/TASLP.2017.2738568](https://doi.org/10.1109/TASLP.2017.2738568).
- [15] J. M. Vojtech, M. D. Chan, B. Shiwani, S. H. Roy, J. T. Heaton, G. S. Meltzner, P. Contessa, G. De Luca, R. Patel, and J. C. Kline, "Surface electromyography-based recognition, synthesis, and perception of prosodic subvocal speech," *J. Speech, Lang., Hearing Res.*, vol. 64, no. 6, pp. 2134–2153, Jun. 2021, doi: [10.1044/2021.JSLHR-20-00257](https://doi.org/10.1044/2021.JSLHR-20-00257).
- [16] G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, and J. C. Kline, "Silent speech recognition as an alternative communication device for persons with laryngectomy," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 12, pp. 2386–2398, Dec. 2017, doi: [10.1109/TASLP.2017.2740000](https://doi.org/10.1109/TASLP.2017.2740000).
- [17] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: [10.1109/TSP.2013.2288675](https://doi.org/10.1109/TSP.2013.2288675).
- [18] S. Nagineni, S. Taran, and V. Bajaj, "Features based on variational mode decomposition for identification of neuromuscular disorder using EMG signals," *Health Inf. Sci. Syst.*, vol. 6, no. 1, pp. 1–10, Dec. 2018, doi: [10.1007/s13755-018-0050-4](https://doi.org/10.1007/s13755-018-0050-4).
- [19] N. Sukumar, S. Taran, and V. Bajaj, "Physical actions classification of surface EMG signals using VMD," in *Proc. Int. Conf. Commun. Signal Process. (ICCCSP)*, Apr. 2018, pp. 0705–0709.
- [20] D. B. Krishnamani and R. Swaminathan, "Variational mode decomposition based differentiation of fatigue conditions in muscles using surface electromyography signals," *IET Signal Process.*, vol. 14, no. 10, pp. 745–753, Dec. 2020, doi: [10.1049/iet-spr.2020.0315](https://doi.org/10.1049/iet-spr.2020.0315).
- [21] H. M. Teager and S. M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Speech Production and Speech Modelling (NATO ASI Series)*, vol. 55, W. J. Hardcastle and A. Marchal, Eds. Dordrecht, The Netherlands: Springer, 1990. [Online]. Available: https://doi.org/10.1007/978-94-009-2037-8_10
- [22] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque, NM, USA, 1990, pp. 381–384.
- [23] X. Li, P. Zhou, and A. S. Aruin, "Teager–Kaiser energy operation of surface EMG improves muscle activity onset detection," *Ann. Biomed. Eng.*, vol. 35, no. 9, pp. 1532–1538, Aug. 2007.
- [24] S. Solnik, P. Devita, P. Rider, B. Long, and T. Hortobágyi, "Teager–Kaiser operator improves the accuracy of EMG onset detection independent of signal-to-noise ratio," *Acta Bioeng. Biomechanics/Wroclaw Univ. Technol.*, vol. 10, no. 2, p. 65, 2008.
- [25] S. Solnik, P. Rider, K. Steinweg, P. DeVita, and T. Hortobágyi, "Teager–Kaiser energy operator signal conditioning improves EMG onset detection," *Eur. J. Appl. Physiol.*, vol. 110, no. 3, pp. 489–498, Oct. 2010, doi: [10.1007/s00421-010-1521-8](https://doi.org/10.1007/s00421-010-1521-8).
- [26] D. Yang, H. Zhang, Y. Gu, and H. Liu, "Accurate EMG onset detection in pathological, weak and noisy myoelectric signals," *Biomed. Signal Process. Control*, vol. 33, pp. 306–315, Mar. 2017, doi: [10.1016/j.bspc.2016.12.014](https://doi.org/10.1016/j.bspc.2016.12.014).
- [27] G. Cho, W. Yang, D. Lee, D. You, H. Lee, S. Kim, S. Lee, and W. Nam, "Characterization of signal features for real-time sEMG onset detection," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104774, doi: [10.1016/j.bspc.2023.104774](https://doi.org/10.1016/j.bspc.2023.104774).
- [28] A. Khorrami Chokami, M. Gasparini, and R. Merletti, "Identification of periodic bursts in surface EMG: Applications to the erector spinae muscles of sitting violin players," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102369, doi: [10.1016/j.bspc.2020.102369](https://doi.org/10.1016/j.bspc.2020.102369).
- [29] S. Wang, S. Zhu, and Z. Shang, "Comparison of different algorithms based on TKEO for EMG change point detection," *Physiological Meas.*, vol. 43, no. 7, Jul. 2022, Art. no. 075001, doi: [10.1088/1361-6579/ac783f](https://doi.org/10.1088/1361-6579/ac783f).
- [30] I. E. Jaramillo, J. G. Jeong, P. R. Lopez, C.-H. Lee, D.-Y. Kang, T.-J. Ha, J.-H. Oh, H. Jung, J. H. Lee, W. H. Lee, and T.-S. Kim, "Real-time human activity recognition with IMU and encoder sensors in wearable exoskeleton robot via deep learning networks," *Sensors*, vol. 22, no. 24, p. 9690, Dec. 2022, doi: [10.3390/s22249690](https://doi.org/10.3390/s22249690).
- [31] M. Pesenti, G. Invernizzi, J. Mazzella, M. Boccione, A. Pedrocchi, and M. Gandolla, "IMU-based human activity recognition and payload classification for low-back exoskeletons," *Sci. Rep.*, vol. 13, no. 1, p. 1184, Jan. 2023, doi: [10.1038/s41598-023-28195-x](https://doi.org/10.1038/s41598-023-28195-x).
- [32] J. Hoffmann, S. Roldan-Vasco, K. Kruger, F. Niekkel, C. Hansen, W. Maetzel, J. R. Orozco-Arroyave, and G. Schmidt, "Pilot study: Magnetic motion analysis for swallowing detection using MEMS cantilever actuators," *Sensors*, vol. 23, no. 7, p. 3594, Mar. 2023, doi: [10.3390/s23073594](https://doi.org/10.3390/s23073594).
- [33] 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," in *Proc. 8th IEEE RAS/EMBS Int. Conf. for Biomed. Robot. Biomechanics (BioRob)*, Nov. 2020, pp. 416–421.
- [34] H. Su, Z. Li, G. Li, and C. Yang, "EMG-based neural network control of an upper-limb power-assist exoskeleton robot," in *Advances in Neural Networks—ISNN (Lecture Notes in Computer Science)*, vol. 7952, C. Guo, Z. G. Hou, and Z. Zeng, Eds. Berlin, Germany: Springer, 2013. [Online]. Available: https://doi.org/10.1007/978-3-642-39068-5_25
- [35] H. A. Javaid, M. I. Tiwana, A. Alsanad, J. Iqbal, M. T. Riaz, S. Ahmad, and F. A. Almisned, "Classification of hand movements using MYO armband on an embedded platform," *Electronics*, vol. 10, no. 11, p. 1322, May 2021.
- [36] B. Karan, K. Mahto, and S. S. Sahu, "Detection of Parkinson disease using variational mode decomposition of speech signal," in *Proc. Int. Conf. Commun. Signal Process. (ICCCSP)*, Apr. 2018, pp. 0508–0512.
- [37] A. Matic, V. Osmari, and O. Mayora, "Speech activity detection using accelerometer," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 2112–2115.
- [38] T. Hueber, E.-L. Benaroya, G. Chollet, B. Denby, G. Dreyfus, and M. Stone, "Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips," *Speech Commun.*, vol. 52, no. 4, pp. 288–300, Apr. 2010.
- [39] Y. Shin and J. Seo, "Towards contactless silent speech recognition based on detection of active and visible articulators using IR-UWB radar," *Sensors*, vol. 16, no. 11, p. 1812, Oct. 2016.
- [40] J. A. Gonzalez, L. A. Cheah, J. M. Gilbert, J. Bai, S. R. Eil, P. D. Green, and R. K. Moore, "A silent speech system based on permanent magnet articulography and direct synthesis," *Comput. Speech Lang.*, vol. 39, pp. 67–87, Sep. 2016, doi: [10.1016/j.csl.2016.02.002](https://doi.org/10.1016/j.csl.2016.02.002).



UZMA SHAFIQ received the bachelor's degree in electrical engineering and the master's degree in biomedical engineering from the National University of Sciences and Technology, in 2018 and 2023, respectively.

Her research interests include EMG signal processing, pattern recognition, deep learning, subvocal voice recognition, computational algebra, and myoelectric prosthetic control.



ASIM WARIS (Member, IEEE) received the B.Sc. degree in mechatronics engineering and the master's degree in biomedical engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2011 and 2013, respectively, and the Ph.D. degree from the Department of Health Science and Technology, Aalborg University, in 2018. He is currently an Assistant Professor with NUST. His research interests include EMG signal processing, pattern recognition, deep learning, and myoelectric prosthetic control.



SYED OMER GILANI (Senior Member, IEEE) received the M.Sc. degree in computer engineering in Sweden, in 2006, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, in 2013. He is currently an Assistant Professor with Abu Dhabi University. Between 2006 and 2008, he was with the Interactive Multimedia Laboratory, Singapore. His research interests include the use of power electronic in power systems, human-machine interaction, and networking.

• • •



JAVOID IQBAL (Member, IEEE) received the bachelor's and master's degrees in mechanical engineering from the College of Electrical and Mechanical Engineering (CE & ME), National University of Science and Technology (NUST), Islamabad, and the Ph.D. degree from The University of New South Wales, Australia. He is currently performing his duties as the Dean of the Faculty of Engineering, CE & ME, NUST. His research interests include industrial simulation, machine learning, mechanical systems, and mobile robots.