Received: February 28, 2023. Revised: April 6, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Application of Fractal Analysis based Feature Extractor for Channel Reduction of Silent Speech Interface Using Facial Electromyography

Asif Abdullah[1]* **Omkar S Powar[2]** **Krishnan Chemmangat[1]**

[1]*Department of Electrical and Electronics, National Institute of Technology Karnataka, India*
[2]*Department of Biomedical Engineering, Manipal Institute of Technology, MAHE, Karnataka, India*
* Corresponding author's Email: asifabdullah92@gmail.com

Abstract: Surface electromyography (sEMG) based silent speech interface (SSI) is an actively investigated topic among the broad area of human computer interaction studies which is currently dominated by acoustic sound based speech recognition research. This research is an attempt to help people who have an impaired vocal system if they are having no issues with their facial muscle functions. The basic idea is to reduce the total number of sEMG electrodes that has to be affixed on the face thereby reducing the invasiveness of the silent speech recognition module. This is achieved by incorporating a new detrended fluctuation analysis (DFA) based feature along with the already existing features associated with electromyographic signals. DFA is used for the first time in literature in the area of surface electromyography based silent speech recognition. The main idea is to incorporate the DFA feature along with the state-of-the-art features to improve the performance of a sEMG based SSI model so that an efficient channel reduced model can be realised. Different channel combinations were tried to analyse the impact of each channel in word recognition accuracy and the optimal channel combination was identified. As a result of this research work, a reduced channel setup with 5 electrodes was proposed in place of the conventional 7 channel data acquisition setup. This was achieved while maintaining an accuracy of 83.88 % and 92.92 % using the decision tree (DT) model and K-nearest neighbours (KNN) model respectively.

Keywords: Surface electromyography, Silent speech interface, Human computer interaction, Detrended fluctuation analysis, Channel reduction, Pattern recognition, Decision trees, K-nearest neighbours.

1. Introduction

There are a number of discoveries like fire, wheel, sharp tools etc. which had a great impact in the advancement of Homo sapiens. While the above mentioned inventions are widely acknowledged as the foundation stones for human progress, there are some discoveries that are underrated or often go unnoticed. The form of communication using sound and its evolution into different sophisticated languages is such an invention. The fascinating story of languages has always motivated people to do research in many areas such as philology and linguistics. In the modern era, the invention of computers and the rapid development in technology has opened up hopes for researchers to delve into the area of Human Computer Interaction (HCI) [1]. The information contained in human speech has all the desirable characteristics that can be promising for an HCI researcher. There are various potential directions like gender recognition [2], speech detection [3], and emotion recognition [4-6] and so on for an HCI researcher to pursue. There are also several methods in HCI research and most of them pertain to acoustic speech recognition [7], whereas this work is focused on recognition of silent speech. Hence surface electromyographic (sEMG) signals that get activated during muscle movements are used in this work.

The methods developed in this work can contribute positively to HCI research in general, but the main focus of the work is to aid the patients who have an impaired vocal system. There is a surgical procedure called laryngectomy [8] in which the vocal chords of a person are removed due to some medical conditions. There are other medical conditions like dysarthria that cause speech impairment and some speech recognition research has been done for that [9]. But in this paper we are focussing mainly on laryngectomy patients with a hope that the methods developed can further be modified in the future to suit other speech impairment cases. The facial muscle movements of laryngectomy patients will be similar as that of a healthy person, but it will not produce any audible sound due to the lack of vocal chords. The objective is to recognise the words uttered by the user which can be later used to identify a specific sentence from a list of sentences. The paper focuses on reducing the number of channels required for sEMG data extraction by exploiting the improvement in word recognition accuracy obtained using an additional time-frequency domain feature along with the existing time domain feature vector. The feature that is being introduced is based on Detrended Fluctuation Analysis (DFA) and it is being introduced for the first time in the area of sEMG based silent speech recognition (SSR). The objective is to find out the importance of each channel in successfully recognising the words. This is done by evaluating the performance of different channel combinations. Two different classification techniques are employed to ensure the reliability of the speech recognition models devised in this paper.

The signals that are influenced by stochastic characteristics are not easy to analyse. DFA is a popular scaling analysis method used in such signals. The first implementation of DFA was done by Peng et al. to show a crossover phenomenon occurring in a physiological non stationary time series [10]. The scaling exponent obtained from DFA is a useful feature in studies that use electroencephalogram (EEG) and electrocardiogram (ECG) signals. EMG signals are non linear and hence Phinyomark et al. [11] used DFA to classify various upper limb movements. The classification accuracy obtained was better than other techniques that use non linear signals. Non stationary nature is an inherent characteristic of EMG signals and hence Phinyomark et al. [12] used the scaling exponents derived from DFA as the prominent feature for identifying upper limb movements. Multi fractal DFA was used by Garc´ıa-Espinosa et al. [13] for the analysis of EMG signals in order to detect and treat tempero mandibular disorder in people. In the research domain of uterine electromyography (uEMG), DFA was used to estimate generalised hurst exponents (GHE) which can be utilised to classify signals [14]. In the same domain, DFA was also utilised for the forecast of preterm birth [15]. In this work, DFA is being applied for maintaining the accuracy of silent speech recognition using sEMG while applying channel reduction. A reduced channel system can offer several advantages like, ease of use for the patients, less hardware requirement, less computation, and faster response.

Two computationally less expensive classifiers namely K-nearest neighbours (KNN) and decision trees (DT) are used in this work. The channel reduction technique was first applied to the KNN model and the results observed from the model were verified using DT model. The main target of this paper is to evaluate the influence of DFA feature in the improvement of word identification accuracy of the SSI model and to investigate the possibility of channel reduction without considerable loss of accuracy. The major contributions of this paper are enumerated below.

(1) The use of DFA feature for improving the word recognition accuracy of sEMG based silent speech model is done for the first time in literature.

(2) Word identification models described in contemporary research generally use independent utterances of words, whereas this research used word data extracted from sentences. Thus it is more challenging to recognise words as compared to independent utterances of words.

(3) The application of channel reduction in order to achieve better ease of use, reduced hardware and software complexity, and faster performance.

The paper is structured in the following order. Section 2 elaborates the background information associated with sEMG based SSI, features extracted, channel reduction technique, and the classifiers employed. Section 3 provides information regarding the data set used and the methodology of the research including detailed descriptions of the feature extraction, channel reduction, and classification steps. The results gathered from the work are presented and discussed in section 4. Then in section 5 the conclusion of the paper is presented along with some ideas for future research in the area.

2. Background

2.1 Silent speech recognition (SSR)

Silent speech recognition (SSR) refers to the identification of silently uttered speech that doesn’t involve the presence of an acoustic sound. An SSR model can be developed using many techniques, among which surface electromyography (sEMG) based SSR, is used in this work. The muscle activation in our body is capable of producing an electric activity, which is detected and recorded for further analysis, in this technique [16]. The muscular activity is captured using sEMG electrodes by positioning them on the skin of the muscles under investigation. The signal captured by the electrodes are filtered and then normalized before the extraction of appropriate features. A machine learning algorithm can then use the features to identify the pattern of the signals. In sEMG based SSR, the electrodes are positioned on relevant places of the user's face, in order to capture the resulting signal.

The initial step in capturing sEMG based signal is identifying the facial muscles involved in the production of speech. The potential of facial muscles is not just limited to acoustic communication; rather they are capable of performing reflexes to stimulating events and can produce various gestures to express emotions. The accuracy and reliability of data extraction depends on the selection of optimal locations on the face where electrodes can be placed. It has to be done such that the occurrence of cross talk between the electrodes due to overlapping muscles are minimised. Another important aspect to be noted here is the minimisation of the number of electrodes. Less number of electrodes on the face can relieve the uneasiness of the subject, and can reduce the overall complexity of the speech recognition model. Not only the hardware complexity is reduced, but also the computational expense is lowered since the total data involved is reduced. The topographic information of the facial muscles that are linked with speech is given in [17]. In literature the use of seven channels is widely described to achieve satisfactory values of word recognition accuracy. A novel method of high density placement of electrodes was developed by Zhu et al. recently and it improved the reliability of the speech recognition system that employs sEMG [18]. In future this can aid in the design of custom electrode placement for each subject thereby increasing the acceptability of the speech recognition model.

Fig. 1 shows the actual flow of a silent speech recognition model. The acquired EMG signals from face are pre processed and then appropriate features are extracted from the signal. These features are then used by the classification algorithm to recognize the pattern of each of the words under consideration. The performance of the model is evaluated using the accuracy obtained during the testing of the model. The word recognition accuracy is the deciding factor for the investigation of better features so that the performance of the model can be further improved.

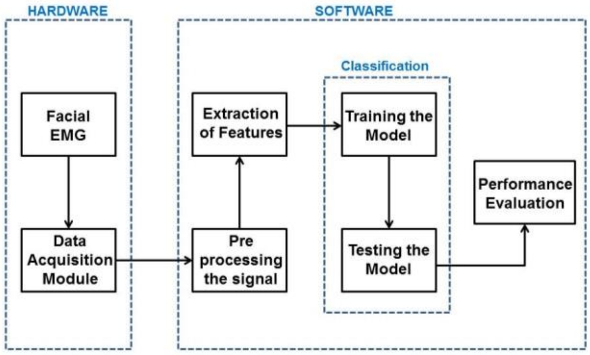


Figure. 1 SSR block diagram

This research work uses the data already acquired and readily available with the research team. The methods elaborated in this work starts from the data pre processing stage to the performance evaluation stage. The hyper parameters associated with the classifier are optimized and if the performance is still poor, then further investigations are done to obtain better features.

2.2 Time domain features

Research in the area of acoustic speech recognition is closely linked with frequency domain features. There are many features which have demonstrated their superiority when it comes to speech recognition based on audible sound. Mel frequency cepstral coefficients (MFCC) [19-21] is an example of such a feature. However they are seen to fail in the case of identifying sEMG based speech. The computational expense of frequency domain features is also more than that of time domain features. This was the motivating factor to investigate more about the time domain features that are useful in this research area. The various time domain features used in this work are given below.

2.2.1. Normalized sEMG ($\bar{x}$)

If the raw sEMG signal is denoted by *E(n)*,

$$x(n) = \frac{E(n)-\min(E)}{\max(E)-\min(E)} \tag{1}$$

$\bar{x}$ = frame-based time domain mean of *x(n)*

2.2.2. Nine-point double averaged sEMG ($\bar{w}$)

Nine-point double averaged signal *w(n)* is given by,

$$w(n) = \frac{1}{9} \sum_{k=-4}^{4} v(n+k) \tag{2}$$

where,

$$v(n) = \frac{1}{9} \sum_{k=-4}^{4} x(n+k) \qquad (3)$$

$\bar{w}$ = frame-based time domain mean of *w(n)*

2.2.3. Rectified nine-point double averaged sEMG ($\bar{r}$)

The rectified high frequency component of the signal is defined as,

$$r(n) = x(n) - w(n) \qquad (4)$$

$\bar{r}$ = frame-based time domain mean of *r(n)*

2.2.4. Power of nine-point double averaged sEMG (P_w)

The power of the signal *w(n)* is given by,

$$P_w = \frac{1}{9} \sum_{k=-4}^{4} w(n)^2 \qquad (5)$$

2.2.5. Power of rectified nine-point double averaged sEMG (P_r)

The power of the signal *r(n)* is given by,

$$P_r = \frac{1}{9} \sum_{k=-4}^{4} r(n)^2 \qquad (6)$$

2.2.6. Zero crossing rate (z_x)

z_x = frame-based zero crossing rate of *x(n)*

2.3 Detrended fluctuation analysis

DFA is a feature that effectively makes use of the characteristics of both time domain and time-frequency domain. In DFA method, the sEMG signal is first integrated to transform it into a Random Walk. It is then split into rectangular windows of same size, without any overlap. For each window, a least square fit is calculated to illustrate the semi local trend of that window. The last step is to calculate the root mean square (RMS) fluctuation of each of the windows to obtain the detrended time series.

The DFA feature can thus be given as:

$$F(i) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_i(k)]^2} \qquad (7)$$

where *i* denotes the window number, *N* denotes the total signal length, and *y(k)* denotes the random walk conversion of the EMG signal *x(n)* and is given by:

$$y(k) = \sum_{n=1}^{k} \left[x(n) - \overline{x(n)}\right], \; k = 1, \dots, N \qquad (8)$$

where $\overline{x(n)}$ is the mean value of *x(n).*

2.4 Classification techniques

This research work is based on the utilisation of a new feature vector that consists of DFA in association with the state-of-the-art features employed in EMG based silent speech recognition. The objective is to exploit the improvement in accuracy achieved using this feature vector, so that an effective channel reduction strategy is facilitated. The importance of each channel in word recognition accuracy and the optimal channel combination needs to be validated using appropriate classification methods. In this work, K- nearest neighbors (KNN) and decision trees (DT) are the classifiers that are employed for this purpose. These two classifiers are selected due to their similarity on the grounds of computational cost and model complexity. Both of them have lower computational cost and model complexity when compared with other machine learning algorithms and deep learning methods.

K-nearest neighbors (KNN) is a non parametric method of classification in which a sample is included in a specific class based on plurality votes of the neighbouring samples. During training phase of KNN, the algorithm performs a local approximation of the classifying function, and all the computation is done at the time of testing. If the data used is normalized, then it can further enhance the performance of KNN. It is also important to note that the computational expense associated with KNN is very less. Therefore processing a large dataset is computationally affordable and consistency of the KNN algorithm improves with the use of more data. These characteristics favoured the use of KNN in this work. Powar et al. [22] and Cerci and Temeltas¸ [23] employed KNN for the classification of EMG signals produced from the motion of the muscles in the hand. Ma et al. used it to classify a set of ten Chinese words using surface electromyographic signals where the words were uttered silently [24]. Chatterjee et al. applied KNN for the segregation of EEG signals using multi-fractal DFA as the feature set [25].

Decision trees (DT) [26] use a simple mechanism to perform classification and hence the computation cost associated with it is less when compared to other classification methods. In a tree, the algorithm assigns various values as nodes of the tree using patterns identified from the data. Addition of further nodes and branches are done as per the required number of classes. A mathematical index is used to decide the most optimal split criterion. Decision Trees have some advantages that can be useful in silent speech recognition. In the data acquisition process of sEMG signals, there are issues of missing samples, non linear values, and the existence of more outliers. DT classifier is immune to such problems related to sEMG data. Missing data values and outliers cannot stop spitting of data to build the tree. This happens because the splitting is independent of absolute values, instead it occurs based on the amount of samples inside the specific split ranges. Also, data linearity is not always required. In the work done by Povey et al., DT is employed with deep learning methods for speech recognition [27].

2.5 Accuracy benchmark from literature

In 2017, Meltzner et al. [28] performed SSI on the data of eight people in which eight sensors were employed. The eight people who were part of this work had earlier undergone laryngectomy. MFCC features were used as a baseline method for feature extraction. Thus the dimensionality of the feature vector was very high and hence linear discriminant analysis (LDA) was applied for dimensionality reduction. They used hidden markov models (HMM) and Guassian mixture models (GMM) for identifying words. Deep learning techniques were not used due to the data hungry aspect and the subsequent requirement of better data acquisition methods. The model used in the work was phoneme based and they got an accuracy of 89.7% (word error rate of 10.3%). They also obtained 86.4% accuracy while using a model that employed only four sensors. The work was carried out on a vocabulary of 2500 words. In 2018 the accuracy was further improved to 91.1% in a reduced vocabulary consisting of 2200 words [29].

The reason for selecting this particular work for comparison is given below.

(1) They used only sEMG signals and no other devices or modalities were used along with it.
(2) They reported the best accuracy so far for a sEMG based SSI model with a large vocabulary.
(3) They didn’t use deep learning techniques for classification.

All these factors closely matches with the aspects of the work presented in this paper.

3. Methods

3.1 Data used

The dataset used in this research work is the EMG UKA corpus which is developed and maintained by the researchers of the interactive systems labs, University of Karlsruhe [30]. The data corpus is very much helpful in aiding the research activities related to speech analysis and recognition. In addition to sEMG data, the corpus also consists of acoustic speech of the subjects. The acquisition of data was performed in 3 different modes - silent, whispered, and audible. This helped the classifier to get some information regarding various force levels associated with human speech. The corpus contains data with a total duration of 7 hours and 40 minutes. It is further divided into 63 sessions comprising of a total of eight speakers. The silent mode consists of a data duration of 1 hour and 46 minutes, and the whispered mode consists of a data duration of 1 hour and 47 minutes. The remaining data is in the audible mode.

The sEMG data acquisition was performed using a 7 channel electrode arrangement (including reference channel) which was set up by Maier-Hein et al. [31]. The muscles chosen for electrode placement are given as follows: the platysma, zygomaticus major, depressor anguli oris, levator anguli oris, anterior belly of the digastric and the tongue. Four channels (1, 3, 4, 5) were unipolar, and the remaining two channels (2, 6) were bipolar. The electrode placement is given in Fig. 2.

The data available in the corpus is clearly marked and well documented. Separate markings for both words and phonemes are present. This enables researchers to carry out their investigations in both directions - word based speech recognition and phoneme based speech recognition. This work uses word recognition based method for the identification of speech. A total of 1100 words were used in this work. The words having sufficient number of utterances were selected and the choice of pre-positions, post-positions, articles etc. were minimised. The words chosen for the work are extracted from the total utterance of each of the sentences, with the help of signal markers. The word based speech identification discussed in the literature uses individual utterances of words and therefore errors associated with data extraction will be minimum. Obviously performance of the classifier becomes better. However in a practical sense there is a need for the classifier to address the errors associated with data extraction when words are taken from sentences and hence such a dataset was chosen for this work.

3.2 Channel reduction

The sEMG data used in this work consists of 7

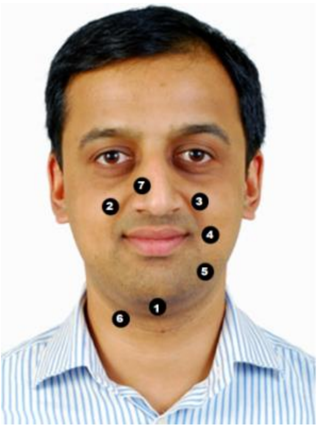


Figure. 2 Actual electrode locations

channels and the locations of the electrodes on the human face are given in Fig. 2. It can be seen that the electrodes occupy a considerable portion of the face which causes difficulty to the user. Thus minimising the number of channels leads to reduced hardware complexity. A reduction in the number of channels can also provide the advantage of decreasing the computational expense of the model.

The aim is to reduce the number of channels by at least two. Hence there would be a total of 5 electrodes (including the reference electrode) remaining. To achieve this, it is important to find out the impact of each of these channels on word recognition accuracy. So different channel combinations have to be tested for the optimal selection of the channels. It is also important to analyze the impact of the DFA based feature in the whole channel reduction process. Hence the comparison between the combinations that use the DFA feature and the ones that doesn't use it needs to be done.

3.3 Feature extraction

The extracted sEMG data cannot be used directly by the classifier since raw EMG data is not capable of producing necessary patterns that is useful for the classifier. Therefore suitable features need to be calculated form raw data and the input to the classifier would be these features. There are 6 time domain features and a time-frequency domain feature that is being considered in this work. The vector in which all the features are stacked together can be expressed as follows:

$$\mathbf{FV}7 = [\bar{x} , \bar{w} , \bar{r} , P_w , P_r , z_x , F] \qquad (9)$$

In FV7, the first 6 features constitute state-of-the-art feature vector which is commonly used in the area of sEMG based silent speech identification and the last one denotes the DFA feature. The feature values for all the 7 channels are stacked in the same way. Two combinations of this feature vector (with and without DFA) for two channel combinations (full channel and reduced channel) each are analysed in this work. Therefore there are a total of four combinations to be performed for each of the two classifiers.

3.4 Classification

The extracted features were used to constitute four unique combinations of the feature vector as mentioned in the above section. For this work, 50 trials of classification were implemented. 80% of the total available data was utilised for training and the testing was done on the remaining 20% data. To ensure the reliability of the algorithm, the selection of data points for each trial was done randomly. Both training and testing of the model was implemented in frame level. To recognize the uttered word, voting was performed on the classified frames once testing process was finished. After the voting process, the word recognition accuracy was calculated for each word in the testing samples.

Two distinct classifiers used in this work are K-nearest neighbors (KNN) and decision trees (DT). The number of nearest neighbours in a KNN algorithm is represented by the ’k’ value of the algorithm which was taken as ’3’ for this work. The distance measure employed for the algorithm was ‘euclidean’. Trial and error method was performed for the selection of the appropriate distance measure and optimal k value. The split criterion used by the DT classifier was ‘gini’s diversity index (gdi)’. It also used a ‘maximum number of splits’ of ‘80000’. Both these were found out using trial and error. Exhaustive grid search technique was employed to determine the optimal parameters of both classifiers.

4. Results and discussion

The work detailed in this paper consists of a resourceful vocabulary of 1100 words which was helpful in determining the impact of various features and channels in enhancing the word identification accuracy of a sEMG based model. The data fed to the classifier consists of 6 time domain features and 1 time-frequency domain feature for each of the 7 sEMG channels. Hence the total number of features used in this work counts to 49. For the extraction of features, a rectangular window of length 54 ms and an overlapping window shift of 1.6 ms were used. The accuracy of classification improved with the increase in window length and saturated after a particular value. Fig. 3 gives an idea about how the optimal values for window length and window shift

Received: February 28, 2023. Revised: April 6, 2023.

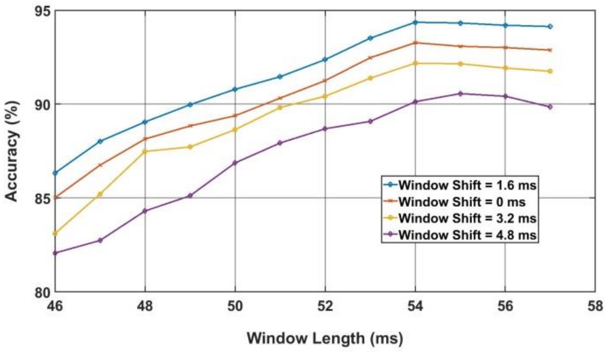


Figure. 3 Optimal window length and window shift

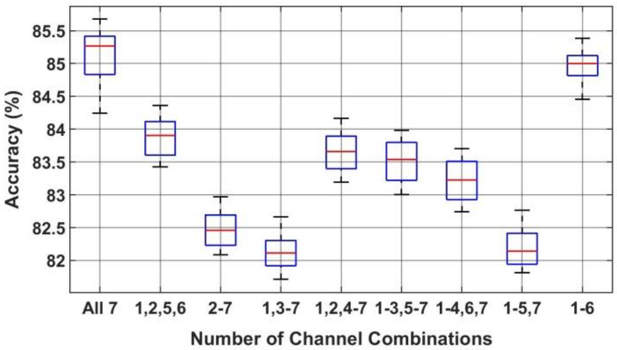


Figure. 4 Accuracy boxplot of channel combinations

were chosen. The influence of each feature as well as each channel was investigated in order to identify the most impactful features and channels. The model accuracy presented in this paper is the average value of 50 separate trials. In each of the trials all of the 1100 words were used and the training and testing data was ensured to be different for each trial.

The whole work was implemented in MATLAB and that includes processing the data, extraction of features, and classification. In a computation work it is vital to present the system architecture being used especially when the computation time is analysed for performing a quantitative comparison between different combinations. The technical specifications of the processor used in this research for feature extraction and classification are given below.

Processor : Intel(R) Core(TM) i7-4770 @3.40 GHz
RAM : 24.00 GB
System Type : 64-bit Operating System, x64-based processor

4.1 Investigation of the channel combinations

The search for the best performing channel combination started with the investigation of the impact of each of the channel in the word recognition accuracy. To achieve this, trials were performed for different combinations where one channel was excluded at a time. The resultant accuracy was an indication to the impact of the excluded channel on the word recognition accuracy. Fig. 4 shows the accuracy of the combinations under consideration. It can be seen that the channels 1, 2, 5, and 6 had a greater impact on word recognition accuracy as compared to other channels. So the channel combination 1, 2, 5, 6 was also tested and the resultant accuracy can be seen to perform well enough as compared to the all channel combination.

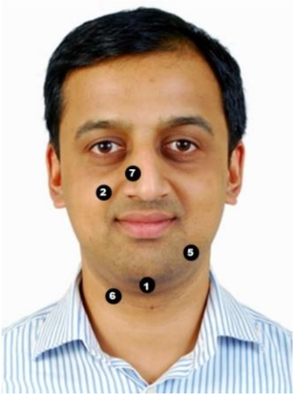


Figure. 5 Proposed electrode locations after channel reduction

Table 1. Comparison of accuracy for DT based model

Channels Taken	All		1.2.5.6	
Features Used	TDFV-DFA	TDFV Only	TDFV-DFA	TDFV Only
Accuracy (%)	85.1506	82.0778	83.8841	79.0240
Standard Deviation (%)	0.3801	0.2250	0.2865	0.0847
Training Time (μs)	1.14	1.09	1.08	1.05
Testing Time (ms)	29.39	28.71	27.23	26.59

The investigation of different channel combinations was first performed on DT based model and the results were validated using KNN model. Based on the results obtained from different channel combinations, the reduced channel locations can be visualized in Fig. 5.

4.2 Impact of DFA in channel reduction of DT based model

The application of DFA as an additional feature along with the existing time domain features had a crucial impact in improving the model performance which is visible in the accuracy plot given in Fig. 6. The mean values for all the 50 trials, the standard deviation, and computation times are presented in Table 1. The results presented here pertain to the DT

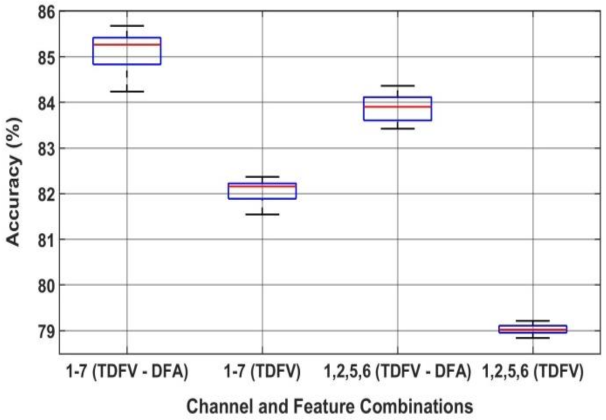
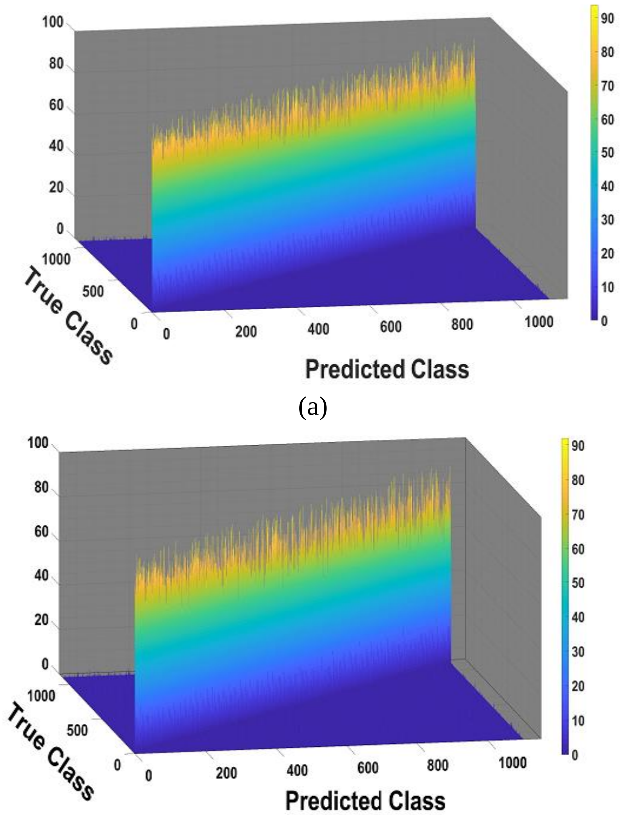


Figure. 6 Accuracy comparison using DT (TDFV vs TDFV-DFA)



(a)

(b)

Figure. 7 Confusion matrices for DT based model: (a) Channels 1,2,5,6 (TDFV-DFA) and (b) Channels 1,2,5,6 (TDFV Only)

based silent speech recognition model.

The confusion matrices for both the channel reduced models - the one that uses only the time domain feature vector and the one that uses both time domain features and DFA feature - are plotted in Fig. 7.

Two important findings can be comprehended from the results. One is the impact of the DFA based feature. For both the full channel and the channel reduced combination, the presence of DFA has contributed significantly in maintaining the accuracy. The second important observation is the closely matching profile of the TDFV-DFA channel reduced combination with that of the TDFV-DFA all channel plot. This demonstrates the ability of DFA in maintaining accuracy even when a significant reduction in data occurs and hence the suitability of DFA feature in channel reduction is established.

4.3 Impact of DFA in channel reduction of KNN based model

The accuracy plot obtained by using KNN based model is presented in Fig. 8. The mean values of accuracy, standard deviation, and computation times are tabulated in Table 2. The confusion matrices for the KNN based model are plotted in Fig. 9.

The results of KNN model also follow the same pattern as in the case of the DT based model. Thus both the classifiers strongly suggest the ability of the DFA based feature in implementing channel reduction while maintaining higher word recognition accuracy.

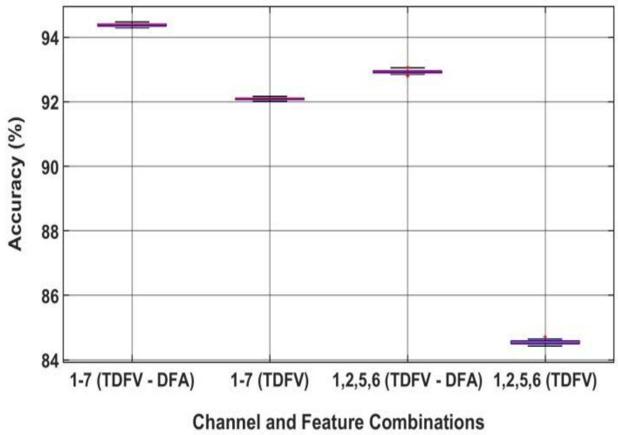
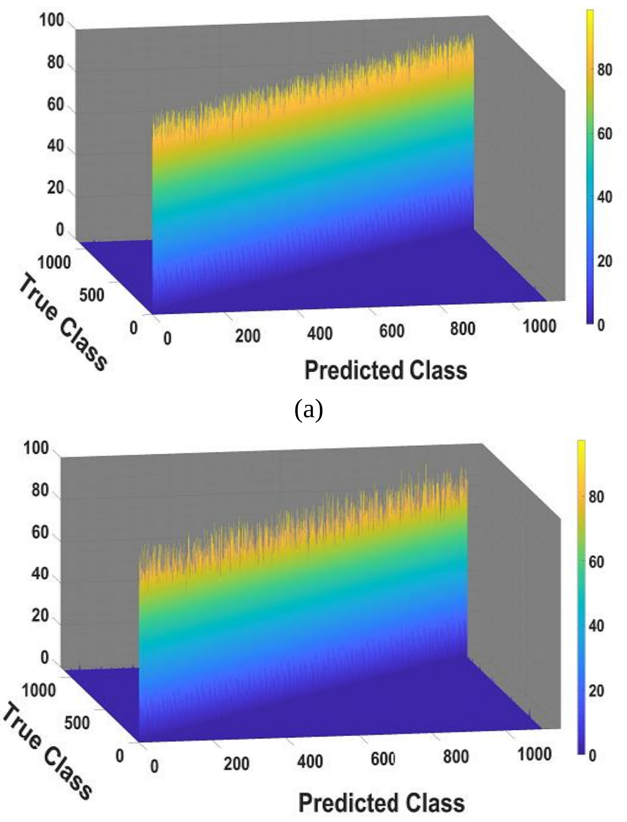


Figure. 8 Accuracy comparison using KNN (TDFV vs TDFV-DFA)

Table 2. Comparison of accuracy for KNN based model

Channels Taken	All		1.2.5.6	
Features Used	TDFV-DFA	TDFV Only	TDFV-DFA	TDFV Only
Accuracy (%)	94.3770	92.0953	92.9263	84.5468
Standard Deviation (%)	0.0432	0.0373	0.0511	0.0630
Training Time (µs)	8.71	8.36	8.42	8.27
Testing Time (ms)	11.4	11.5	10.7	10.3



(a)

(b)

Figure. 9 Confusion matrices for KNN based model (a) Channels 1,2,5,6 (TDFV-DFA) and (b) Channels 1,2,5,6 (TDFV Only)

Table 3. Comparison with accuracy benchmark

	This Work	**Meltzner et.al**
No: of words	1100	2500
Type of data	sEMG	sEMG
SSI Technique	Word-based	Word-phoneme
Classifier	KNN, DT	HMM-GMM
Accuracy **(Std. Dev.)**	94.4% (0.04%), 85.1% (0.38%)	89.7% (5.3%)
Complexity of Model	Low	High

4.4 Comparison with the accuracy benchmark

The importance of DFA feature in improving word recognition accuracy and its ability to aid in channel reduction has been presented so far. Both classifiers have demonstrated the superiority of this time-frequency domain feature in the research area of sEMG based SSI that is pioneered by time domain features. It can also be noted that the accuracy benchmarked in the literature was achieved by the KNN based model devised in this work. A word recognition accuracy of 94.37% obtained by the KNN model is comparable with the benchmark of 89.7% (that was later improved to 91.1% on a reduced vocabulary) obtained by Meltzner et al. [28, 29] where deep learning algorithms are used on a word-phoneme hybrid method. The work devised in this paper could achieve similar accuracy by the use of only EMG data and machine learning methods that are computationally less expensive. Table 3 presents a comparison between the SSI model developed in the benchmarked study and the one discussed in this research work.

4.5 Discussion on the superiority of DFA

The success of DFA in the pattern recognition of EMG signals can be attributed to the capability of the feature to utilise the non stationary behaviour of EMG signals. This is also a reason for the reduced computational expense of the models presented in this paper. As a feature belonging to the time-frequency domain, DFA has some unique properties when compared to other features in the same research area. DFA provides more class separability in cases of low level muscle activation than other conventional features [12]. Low level muscle activation is a common characteristic in the case of silent speech recognition. Chatterjee et.al [25] reported the potential of DFA to categorize EEG signals into focal and non-focal classes. This finding gave an idea about the possible benefit of employing DFA in cases where the occurrence of cross talk between different channels is anticipated to be more than usual. It is well known that the facial muscular activity is highly prone to cross talk occurring from adjacent muscles. All these aspects provided the theoretical base for the DFA based investigations carried out in this research work and the results demonstrated the superior performance of the feature.

5. Conclusion and future scope

The use of DFA feature turned out to be helpful in enhancing the word recognition accuracy of surface electromyography based silent speech recognition. The DT based model achieved an accuracy of 85.1% and the accuracy of the KNN based model was found to be 94.4%. Thus the superior performance of the models paved way for the successful implementation of channel reduction while limiting the loss of accuracy to less than 2%

for both the models. The loss of accuracy was more than 3% and 8% for DT and KNN models respectively when DFA feature was not incorporated along with state-of-the-art features. This demonstrates the ability of DFA to perform better even when channel reduction is applied. The area of HCI research is pioneered by mostly time domain features. Some frequency domain features have been used in the area of speech recognition using acoustic signals, but in the case of silent speech recognition even these features also didn't perform well. As a time-frequency domain feature, DFA has created a path for researchers to pursue further investigations to find such features to be employed in the area of silent speech recognition. It can also be noted that the DFA feature could maintain satisfactory accuracy even in the absence of some channels. Thus other time-frequency domain features can be checked for their ability to implement channel reduction.

This research can be carried forward to check for methods that can be used to enhance the performance of the DFA based feature thereby trying to further reduce the number of channels. It can also be tried to implement systems such as an array of electrodes instead of distinct electrodes on the face. Thus the processes such as fixing electrodes on the face and maintaining the electrodes can be simplified. There are also high scopes for carrying out investigations on other time-frequency domain features to be used in the area of surface electromyography based silent speech recognition. It can be useful for the whole area of human computer interaction research.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization - author 1 and 2 ; methodology - author 1 and 3 ; software - author 1 and 2 ; validation - author 1, 2, and 3 ; formal analysis - author 1 ; investigation - author 2 ; resources - author 1 and 3 ; data curation - author 1 and 3 ; writing—original draft preparation - author 1 ; writing—review and editing - author 2 and 3 ; visualization – author 1 ; supervision - author 2 and 3 ; project administration - author 1 and 3 ; funding acquisition – author 3. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported by the ministry of Electronics and IT, Government of India, under Visvesvaraya PhD Scheme with Grant No: DIC/MUM/GA/10(37)D.

References

[1] J. A. Rubin and R. S. Crockett, "Whole body human computer interface", *U.S. Patent*, No. 9 652 037, May 16, 2017.

[2] A. A. Badr and A. K. A. Hassan, “CatBoost Machine Learning Based Feature Selection for Age and Gender Recognition in Short Speech Utterances”, *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 3, pp. 150-159, 2021, doi: 10.22266/ijies2021.0630.14.

[3] N. Pitropakis, K. Kokot, D. Gkatzia, R. Ludwiniak, A. Mylonas, and M. Kandias, “Monitoring Users’ Behavior: Anti-Immigration Speech Detection on Twitter”, *Machine Learning and Knowledge Extraction*, Vol. 2, No. 3, pp. 192-215, 2020.

[4] A. D. Wibawa, E. S. Pane, D. Risqiwati, and M. H. Purnomo, “Rules Extraction of Relevance Vector Machine for Predicting Negative Emotions from EEG Signals”, *International Journal of Intelligent Engineering and Systems,* Vol. 15, No. 1, pp. 42-54, 2022, doi: 10.22266/ijies2022.0228.05.

[5] B. H. Prasetio, E. R. Widasari, and F. Bachtiar, “A Study of Machine Learning Based Stressed Speech Recognition System”, *International Journal of Intelligent Engineering and Systems,* Vol. 15, No. 4, pp. 31-42, 2022, doi: 10.22266/ijies2022.0831.04.

[6] R. Reddy and A. U. Motagi, “Fake Review Detection and Emotion Recognition Based on Semantic Feature Selection with Bi-Directional Long Short Term Memory”, *International Journal of Intelligent Engineering and Systems,* Vol. 15, No. 5, pp. 473-482, 2022, doi: 10.22266/ijies2022.1031.41.

[7] T. Schultz and A. Waibel, “Language independent and language adaptive acoustic modeling for speech recognition”, *Speech Communication*, Vol. 35, No. 1-2, pp. 31–51, 2001.

[8] C. Tiple, T. Drugan, F. V. Dinescu, R. Muresan, M. Chirila, and M. Cosgarea, “The impact of vocal rehabilitation on quality of life and voice handicap in patients with total laryngectomy”, *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, Vol. 21, No. 127, pp. 1-8, 2016.

[9] H. Albaqshi and A. Sagheer, “Dysarthric speech recognition using convolutional recurrent neural networks”, *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 6, pp. 384-392, 2020, doi: 10.22266/ijies2020.1231.34.

[10] C.K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, “Quantification of scaling exponents and crossover phenomena in non stationary heartbeat time series”, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 5, No. 1, pp. 82–87, 1995.

[11] A. Phinyomark, P. Phukpattaranont, C. Limsakul, and M. Phothisonothai, “Electromyography (EMG) signal classification based on detrended fluctuation analysis”, *Fluctuation and Noise Letters*, Vol. 10, No. 03, pp. 281–301, 2011.

[12] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, “Fractal analysis features for weak and single channel upper limb EMG signals”, *Expert Systems with Applications*, Vol. 39, No. 12, pp. 11156–11163, 2012.

[13] L. A. G. Espinosa, A. M. Mart´ınez, F. P. Escamirosa, J. M. Gonza´lez, I. R. Castan˜eda, B. F. Ram´ırez, N. P. Guerrero, and F. A. Medina, “Multi fractal dfa analysis of masseter muscles semg signal in patients with tmd, pilot study”, *Biomedical Signal Processing and Control*, Vol. 68, p. 102732, 2021.

[14] N. Punitha and S. Ramakrishnan, “Analysis of uter ine emg signals in term and preterm conditions using generalised hurst exponent features”, *Electronics Letters*, Vol. 55, No. 12, pp. 681–683, 2019.

[15] P. S. Asmi, K. Subramaniam, and N. V. Iqbal, “Classification of fractal features of uterine emg signal for the prediction of preterm birth”, *Biomedical and Pharmacology Journal*, Vol. 11, No. 1, pp. 369–374, 2018.

[16] R. Merletti and D. Farina, *Surface Electromyography: Physiology, Engineering, and Applications*, John Wiley & Sons, New Jersy, 2016.

[17] W. C. Yau, S. P. Arjunan, and D. K. Kumar, “Classification of voiceless speech using facial muscle activity and vision based techniques”, In: *Proc. of TENCON IEEE Region 10 Conference,* Hyderabad, India, pp. 1–6, 2008.

[18] M. Zhu, H. Zhang, X. Wang, X. Wang, Z. Yang, C. Wang, O. W. Samuel, S. Chen, and G. Li, “Towards optimizing electrode configurations for silent speech recognition based on high density surface electromyography”, *Journal of Neural Engineering*, Vol. 18, No. 1, p. 016005, 2021.

[19] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, “Speech recognition using MFCC”, In: *Proc. of International Conference on Computer Graphics, Simulation and Modeling*, Pattaya, Thailand, pp. 135–138, 2012.

[20] R. Hidayat and A. Winursito, "A modified MFCC for improved wavelet based denoising on robust speech recognition", *International Journal of Intelligent Engineering and Systems* Vol. 14, No. 1, pp. 12-21, 2021, doi: 10.22266/ijies2021.0228.02.

[21] A. Nosan and S. Sitjongsataporn, “Enhanced Feature Extraction Based on Absolute Sort Delta Mean Algorithm and MFCC for Noise Robustness Speech Recognition”, *International Journal of Intelligent Engineering and Systems* Vol. 14, No. 4, pp. 422-436, 2021, doi: 10.22266/ijies2021.0831.37.

[22] O. S. Powar, K. Chemmangat, and S. Figarado, “A novel pre-processing procedure for enhanced feature extraction and characterization of electromyogram signals”, *Biomedical Signal Processing and Control*, Vol. 42, pp. 277–286, 2018.

[23] C. Cerci and H. Temeltas¸ “Feature extraction of emg signals, classification with ANN and KNN algorithms”, In: *Proc. of 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, pp. 1–4, 2018.

[24] S. Ma, D. Jin, M. Zhang, B. Zhang, Y. Wang, G. Li, and M. Yang, “Silent speech recognition based on surface electromyography”, In: *Proc. of Chinese Automation Congress (CAC)*, Hangzhou, China, pp. 4497–4501, 2019.

[25] S. Chatterjee, S. Pratiher, and R. Bose, “Multifractal detrended fluctuation analysis based novel feature extraction technique for automated detection of focal and non-focal electroencephalogram signals”, *IET Science, Measurement & Technology*, Vol. 11, No. 8, pp. 1014–1021, 2017.

[26] O. Z. Maimon and L. Rokach, *Data Mining with Decision Trees: Theory and Applications*, Vol. 69, World Scientific, Singapore, 2008.

[27] D. Povey, L. Burget, M. Agarwal, P. Akyazi, F. Kai, A. Ghoshal, O. Glembek, N. Goel, M. Karafia´t, A. Rastrow, R. C. Rose, P. Schwarz, and S. Thomas, “The subspace gaussian mixture model—a structured model for speech recognition”, *Computer Speech & Language*, Vol. 25, No. 2, pp. 404–439, 2011.

[28] G. S. Meltzner, J. T. Heaton, Y. Deng, G. D. Luca, S. H. Roy, and J. C. Kline, “Silent speech

recognition as an alternative communication device for persons with laryngectomy", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 25, No. 12, pp. 2386–2398, 2017.

[29] G. S. Meltzner, J. T. Heaton, Y. Deng, G. D. Luca, S. H. Roy, and J. C. Kline, "Development of semg sensors and algorithms for silent speech recognition", *Journal of Neural Engineering*, Vol. 15, No. 4, p. 046031, 2018.

[30] M. Wand, M. Janke, and T. Schultz, "The EMG-UKA corpus for electromyographic speech processing", In: *Proc. of Fifteenth Annual Conference of the International Speech Communication Association*, Singapore, pp. 1593-1597, 2014.

[31] L. M. Hein, F. Metze, T. Schultz, and A. Waibel, "Session independent non audible speech recognition using surface electromyography", In: *Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding*, Cancun, Mexico, pp. 331–336, 2005.