Sub Auditory Communication and Facial EMG

Sanjay Kumar, Dinesh Kant Kumar and Melaku Alemu

School of Electrical and computer Engineering RMIT university Melbourne

Abstract. Availability of speech related information in the facial EMG is discussed. The primary objective of this preliminary work is to investigate the use of facial EMG as a voiceless communication medium. Subjects were asked to utter the five English vowels with no acoustic output (sub-auditory). Three independent EMG signals were acquired from three facial muscles as sub-auditory EMG activations. In order to classify and recognize each vowel based on EMG, RMS of the recorded signals were estimated and used as parametric inputs to a neural network.

1 Introduction

Electromyography (EMG) is the recording of the electrical activity of muscles. The electrical signal is a result of the combination of action potentials during contracting muscle fibers. EMG can be performed using invasive or non-invasive electrodes. Surface Electromyogram (SEMG), the electrical potential recorded non-invasively from the surface of the skin, can be used to identify the overall strength of contraction of muscles. Root mean square (RMS) of SEMG is a good indicator of that strength. Besides clinical applications, SEMG has been used as a control signal in prosthetic devices dating back in the 1970 [1].

Speech has been modelled by the source and filter. The filter of sound is a result of the mouth cavity and lips and results in giving the spectral content to the sound. Vowels are sounds that are relatively stationary while consonants are produced by dynamic variation of the filter characteristics. The shape of the lips and mouth cavity is controlled by the contraction of the corresponding muscles.

Based on the above, it is stated that speech produced by any person is dependent on the activity of the facial muscles controlling the shape of the mouth and lips. Extracting speech related information from facial muscle activities has a number of applications. People with speech related disabilities such as vocal cord damage can be offered the possibility of communicating with others via machines with facial EMG being used as a control signal. EMG, since it is immune from ambient noise, can also provide an alternative communication system in a noisy environment. Even so the application of EMG in speech research is not new, few work is reported in the literature where the above-mentioned applications are effectively realised.

Morse et al [1] have reported the use of EMG recorded from the neck and temple to analyse feasibility of using neural networks to recognize speech. Their parametric input to the neural network was the power spectral density of the EMG activated and

Kumar S., Kant Kumar D. and Alemu M. (2005).
Sub Auditory Communication and Facial EMG.
In *Proceedings of the 1st International Workshop on Biosignal Processing and Classification*, pages 142-149
DOI: 10.5220/0001193301420149
Copyright © SciTePress

recorded while subjects quasi-randomly spoke words. They report a very low overall accuracy of approximately 60% for the recognition of the signal [1]. A.D.C. Chan et al, report the use of facial EMG with linear discriminate analysis to recognize 10 separate numbers with a recognition accuracy of over 90% [2]. However H. Manabe et al [3], have observed language dependent nature of the Chan et al's work as a drawback and suggested the use of phonemes based recognition method [3]. Sugie et al [4] report the use of EMG for identifying the phonemes during the subject speaking five Japanese vowels but report a low accuracy of 60%. Other researchers such as C Jorgensen et al [5] have demonstrated possible application of EMG signal recorded from the Larynx and sublingual areas from below the jaw in speech recognition particularly for silent or sub-auditory speech. Using neural networks with a combination of feature sets, they have shown the potential of sub-acoustic speech recognition based on EMG with up to 92% accuracy.

From the literature reported, there appears to be a discrepancy of the reliability of EMG of the facial muscles to identify speech. Thus, there is a need to determine if the use of EMG to identify muscle activity to produce simple sub-auditory sounds is reliable and reproducible which would then be the basis for a more complex study. With that aim, this paper reports our work conducted to identify certain common sounds using surface EMG under controlled conditions.

2 Background

2.1 English Vowels

English vowels are speech gestures that represent stationary filter characteristics with no nasal involvement [11]. Based on this, it is argued that the mouth and lips shape would remain stationary during the pronunciation of the vowels and hence the muscle contraction during the utterance of the vowels would remain stationary. Utterance of consonants would result in temporal variation of shape and thus changing muscle contraction for the duration of the utterance. For this reason, this research has considered five English vowels. This is also important since English vowels are building blocks in modern speech. By including temporal variation, this can then be extended to consonants.

2.2 Speech Production and Facial EMG

The process of human speech is complex with the involvement of number of muscles. All facial muscles that are involved in pursing the lips, lifting the corners of the mouth and opening the jaw are activated during speech. A number of these muscles are not close to the surface making non-invasive EMG recording impossible. Further, to record EMG from each of these muscles would be extremely clumsy making it extremely uncomfortable for the human participants. Thus it is necessary to identify the most suitable muscles that can be used to identify the different vowels. Three facial muscles were identified that are more active when subjects attempt to pronounce the five vowels. For this aim, three facial muscles were selected. These include Mentalis, Depressor Anguli Oris and Massetter. The Mentalis originates from the mandible and inserts into the skin of the chin to elevate and protrude lower lip, pull chin skin into a pout. The Depressor anguli oris originates from the mandible and inserts skin at angle of mouth pulls corner of mouth downward. Masseter originates from maxilla and zygomatic arch and inserts to ramus of mandible to elevate and protrude, assists in side-to-side movements of mandible.

2.3 EMG Feature Extraction and Classification

Muscle contraction is a result of electrical stimulation received from the nerves to individual muscle fibers. This results in electrical activity that can be recorded by electrodes kept in the close proximity of the muscles. This recording is called EMG. The signal is a summation of number of motor unit action potentials that are spatially and temporally separated. The signal is complex and non-stationary, it is bi-phasic and cannot be represented by a simple mathematical function.

The force produced by contraction of muscles depends on the number of active muscle fibers and the rate of activation of these fibers. Zero- crossing and spectral analysis provide an indication of the rate of activation of the muscle fibers and the density of muscle fibers that are being activated. The amplitude of EMG is an indicator for the size of active motor units and the integrated EMG and the RMS–EMG are indicators of rate (density) of activation as well as the number of active motor units and the size of these motor units. RMS of EMG highlights the ‘strength of the signal’ and thus the strength of contraction of the muscle.

For applications where the machine can identify the function generated by the muscle based on SEMG, require automated analysis and classification of SEMG. For automated classification of SEMG related to movement, it is essential to develop the system that can extract appropriate features of SEMG with respect to the movement and have a mechanism for relating these features to the movement generating the signal. Numbers of researchers have used different techniques for the purpose including statistical analysis of the signal properties and auto-regression analysis Graupe et al. [13] with *85%* success rate. But this system was highly dependent on the subject and recording and required high degree of manual intervention.

Hudgins et al [10] reported the first major work of SEMG classification using Artificial Neural Networks (ANN). The ANN was used to introduce the flexibility and self-learning ability to the classification technique. The accuracy of the classification technique was ranging from *80%* to *90%*. The authors have also used the magnitude of SEMG and neural network based classification to classify pre-defined hand movements using three channels of SEMG [12].

Based on the above, this paper reports the use of multiple channels SEMG of the facial muscles. RMS of the signal is computed and backpropagation neural network has been used to classify SEMG with the shape of the mouth and lips so formed with the aim to identify the vowel in the sub-auditory speech.

3 Methodology

3.1 EMG Recording and Processing

Three male subjects participated in the investigation. The AMLAB workstation was used for EMG recording. The experiment used a 3-channel EMG configuration according to recommended recording guidelines [7]. Ag/AgCl electrodes (AMBU blue sensors from MEDICOTEST Denmark) were mounted on three selected facial muscles (*Mentalis, Depressor Anguli Oris and Massetter*) on the right side of the face. Inter electrode distance was arranged to be 1cm. Before the recording commences, EMG target sites were cleaned with alcohol wet swabs. Inter-electrode impendence was checked using a multimeter.

A pre-amplifier (with a Gain of 1000) was placed for each EMG channels. To minimise movement artifacts and aliasing, a band-pass filter (with low corner (-3dB) 8Hz and with high corner (-3dB) frequency of 79Hz) was implemented. A notch filter, to remove a 50Hz line noise, was also included. The EMG signal was amplified and sampled with a rate of 250Hz.

Three facial EMG simultaneously were recorded and observed while subjects utter the five English vowels (/a/, /e/, /i/, /o/, /u/) for three times with no acoustic out put (sub-auditory). Enough resting time was given in between the three activations. Overall fifteen data sessions were performed for each subject. To observe any changes in muscle activity, the recorded raw EMG signal was further processed.

3.2 EMG Recording and Processing

After the recording process was completed, the raw EMG was transferred to Matlab for further analysis. The RMS (Root Mean Square) value of each signal was estimated by applying equation (1):

$$\text{RMS}=\frac{1}{s}\left[\sum_{s}^{s} f^{2}(s)\right]^{1/2}$$

Where ‘s’ is the window length and f (s) is data within the window.

4 Recognition

Recognition of EMG based speech features may be achieved by applying a supervised artificial neural network. The artificial neural network is efficient regardless of data quality. Neural networks can learn from examples and once trained, are extremely fast making them suitable for real time applications [8-9]. The classification by ANN does not require any statistical assumptions of the data. ANNs learns to recognize the characteristic features of the data to classify the data efficiently and accurately.

SCITEPRESS Science and Technology Publications

Back Propagation (BPN) type artificial neural network (ANN) was used for the purpose. The advantage of choosing feed forward (FF) and BPN learning algorithm architecture is to overcome the drawback of the standard ANN architecture. Augmenting the input by hidden context units, which give feedback to the hidden layer, thus giving the network an ability of extracting features of the data from the training events is one advantage. The size of the hidden layer and other parameters of the network were chosen iteratively after experimentation with the back-propagation algorithm. There is an inherent trade off to be made. More hidden units results in more time required for each iteration of training; fewer hidden units results in faster update rate. For this study, two hidden layer structure was found sufficiently suitable for good performance but not prohibitive in terms of training time. Sigmoid has been used as the threshold function and gradient desent and adaptive learning with momentum as training algorithm. A learning rate of 0.02 and the default momentum rate was found to be suitable for training the network. The training stopped when the network converged and with error less than the target error. The weights and biases of the network were saved and used for testing the network. The data was divided into subsets of training, validation, and test subsets data. One fourth of the data was used for the validation set, one-fourth for the test set, and one half for the training set.

The RMS values of the three channels of EMG captured during the subject pronunciation of the vowels were the inputs to the ANN. The target of the ANN was the corresponding vowels. Fig. 1 depicts the ANN architecture. After training, the system was tested and the accuracy of correct identification by the network was tabulated (Table 1).

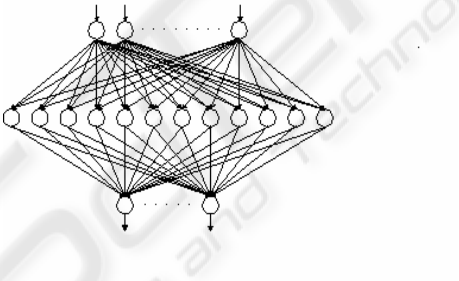

Fig. 1. Neural Network Architecture

5 Results and Discussion

Table 1. Recognition Accuracy (%)

	/a/	*/e/*	*/i/*	*/o/*	*/u/*	*Average*
Subject 1	*97*	*94*	*98*	*93*	*85*	*93.4*
Subject 2	*91*	*86*	*90*	*85*	*93*	*89*
Subject 3	*88*	*89*	*86*	*97*	*95*	*91*

Table 1 shows the experimental results. The results of the testing show that with the system described can classify the five vowels with an accuracy of up to 91%. The higher classification accuracy is due to better discriminating ability of neural network architecture and RMS of EMG as the features. At the present stage, the method has been tested successfully with only three subjects. In order to evaluate the intra and inter variability of the method, a study on a larger experimental population is required. Fig.3-4 depicts the statistical bar diagrams of the three sub-auditory RMS of EMG data. However, due to the small data bank, it is difficult to determine and conclude the significance of the same.

6 Conclusion

This paper describes a study to recognise human sub-auditory speech signal based on the EMG data extracted from the three articulatory facial muscles coupled with neural networks. Test results show recognition accuracy of 91 %. The system is accurate when compared to other attempts for EMG based sub-auditory speech recognition. These preliminary results suggest that the study is suitable to develop a real-time EMG based voiceless communication system.

7 Further Work

Authors are working with statistically larger population of experimental subjects.

References

1. M.S. Morse, Y.N. Gopalan, M. Wright: Speech recognition using myoelectric signals with neural network, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol.13, No.4, pp.1977-1878, 1991.
2. A.D.C. Chan, K.E., B. Hudgins, D.F. Lovely, Myo-electric signals to augment speech recognition. Medical & Biological Engineering & Computing, 2001. 39: p. 500-504.
3. Hiroyuki Manabe, "Unvoiced Speech Recognition using EMG - Mime Speech Recognition –" Short Talks: Specialized Section CHI 2003: NEW HORIZONS Short Talk: Brains, Eyes and Ears CHI 2003: NEW HORIZONS NTT DoCoMo MultimediaLaboratories†manabe@mml.yrp.nttdocomo.co.jp
4. N. Sugie, K. Tsunoda,: A speech prosthesis employing a speech synthesizer. IEEE Transaction on Biomedical Engineering, Vol.BME-32, No.7, pp.485- 490, 1985.
5. Chuck Jorgensen, Diana D Lee & Shane "Sub Auditory Speech Recognition Based on EMG Signals" Agabon. , Proc. of the IEEE conference, 2003
6. Akira Hiraiwa NTT DoCoMo Multimedia Laboratories hiraiwa@mml.yrp.nttdocomo.co.jp Toshiaki Sugimura NTT DoCoMo Multimedia Laboratories sugi@mml.yrp.nttdocomo.co.jp
7. A J Fridlund, J.T.C., *Guidelines for human electrographic research.* Psycholphysiology, 1986. 23: p. 567-589.
8. A. Freeman and M. Skapura, Neural Networks: Algorithms, Applications, and Programming Techniques, Addison-Wesley, Mass., 1991.

9. Haung, K.-Y., "Neural networks for robust recognition of seismic patterns,". IEEE Transactions on Geoscience and Remote sensing 2001
10. Hudgins B, Parker PA, Scott RN, A new strategy for multifunction myoelectric control. IEEE Trans Biomedical Engineering;40(1):82–94, 1993.
11. Paul A Lynn, Signal Processing of Speech Macmillan New Electronics S Publisher: Palgrave Macmillan 1993, ISBN: 0333519213
12. Nigma M, Kumar D,Nemuel Pah,,Proc of the Seventh Australian and New Zealand Intelligent Information System Conference Perth,West Australia,2001
13. Graupe D, Salahi J, Kohn K H Multifunction prosthesis and orthosis control via microcomputer identification of temporal patter differences in single- site myoelectric signals. J Biomed Eng 1982; 4:17–22.

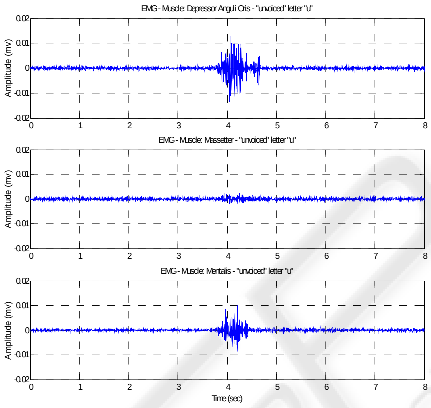

Fig. 2. EMG from three muscles of unvoiced data from subject for vowel “U”

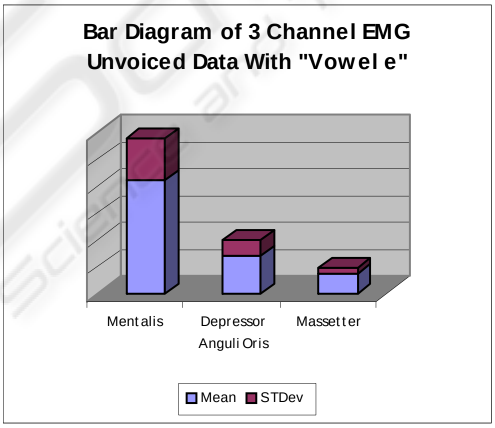

Fig. 3. Bar Diagram of 3 channels EMG data for Mean and Standard Deviation for “Vowel e” Unvoiced data

SciTePress Science and Technology Publications

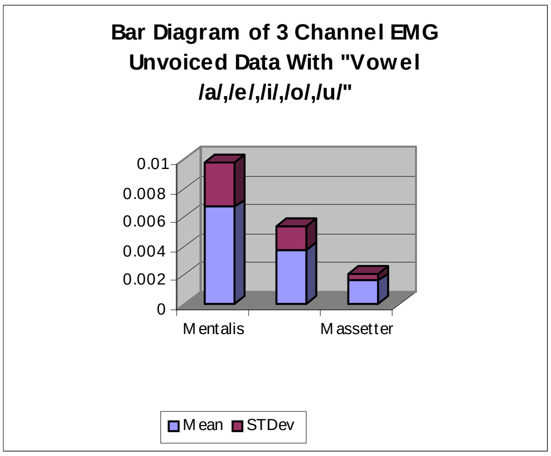

Fig. 4. Bar Diagram of 3 channel EMG data for Mean and Standard Deviation for “Vowel /a/,/e/,/i/,/o/,/u/” unvoiced data

SciTePress Science and Technology Publications