APPLICABILITY OF FACIAL EMG IN HCI AND VOICELESS COMMUNICATION

Sanjay Kumar, Dinesh Kant Kumar, Melaku Alemu
School of Electrical and Computer Engineering
RMIT University GPO BOX 2476 Melbourne VIC 3001

Keywords: Facial EMG, HCI, Voiceless Speech.

Abstract: This paper discusses the speech related information in the facial EMG for applications such as human computer interface. The primary objective of this work is to investigate the use of facial EMG as a voiceless communication medium or to drive computer based equipment by people who are unable to speak. Subjects were asked to pronounce the five English vowels with no acoustic output (voiceless). Three independent EMG signals were acquired from three facial muscles as 'voiceless' EMG activations. In order to classify and recognize each vowels based on EMG, RMS of the recorded signals were estimated and used as parametric/feature inputs to a neural network.

1 INTRODUCTION

Electromyogram (EMG) is the recording of the electrical activity of muscles. It is a result of the combination of action potentials during contracting muscle fibers. EMG can be recorded using invasive or non-invasive electrodes. Surface EMG (SEMG) is non-invasive recording from the surface and is used to identify the overall strength of contraction of muscles and root mean square (RMS) of SEMG is a good indicator of the strength Besides its clinical applications, EMG has been used as a control signal in prosthetic devices dating back in the 1970 (Morse et al., 1991).

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 muscle activity of the facial muscles controlling the shape of the mouth and lips. However, very little work is reported in literature where this relationship has been investigated for speech recognition or other related applications. Morse et al (Morse et al., 1991), are one such group who report 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 recorded while subjects quasi-randomly spoke words. They report a very low overall accuracy of approximately 60% for the recognition of the signal (Morse et al., 1991). 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% (Chan et al., 2001). However H. Manabe et al (Manabe, 2003), have observed language dependent nature of the Chan et al's work as a drawback and suggested the use of phonemes based recognition method (Manabe, 2003). Sugie et al (Sugie et al., 1985) 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 (Jorgensen et al.) 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

Kumar S., Kant Kumar D. and Alemu M. (2005).
APPLICABILITY OF FACIAL EMG IN HCI AND VOICELESS COMMUNICATION.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Robotics and Automation*, pages 379-382
DOI: 10.5220/0001155003790382

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 simple 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

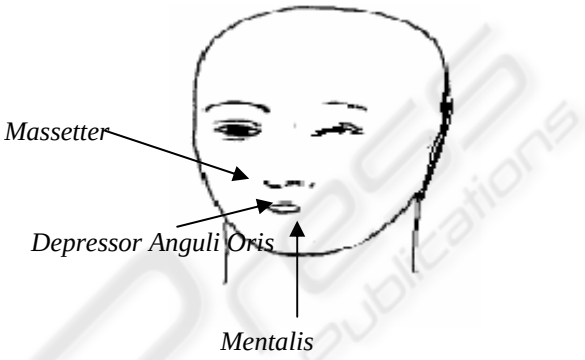
English vowels are speech gestures that represent stationary filter characteristics with no nasal involvement. Based on this, it is argued that the mouth and lips shape would remain stationary during the pronunciation of the vowels and hence the muscle contraction during the utterance of the vowels would 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 because English vowels are an important building stone in modern speech. By including temporal variation, this can then be extended to consonants.

2.2 Speech Production and Facial Muscles

Various facial muscles involve during speech production in pursing the lips, lifting the corners of the mouth, opening the jaw etc. In this trial study, only three facial muscles were selected (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 while 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.

It is impractical to consider all the facial muscles and record their electrical activity. To determine the best choice of muscles, authors are aware that the role of each individual muscle has to be identified and examined objectively. As a short cut, preliminary experiments were undertaken and it was observed that the above-mentioned three facial muscles relatively more active when subjects attempt to pronounce the five vowels. It has been also noticed that Massetter muscle tends to be electrically less active compare with Depressor Anguli Oris and the Mentalis muscle.

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 (Fridlund, 1986). 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. A sample schematic of the recording is shown in figure 2. 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 spoke (‘voicelessly’) the five English vowels (/a/, /e/, /i/, /o/, /u/) for three times. 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.

After the recording process was completed, the raw EMG was transferred to Matlab for further analysis. Using averaging filter, thresholding was done to remove the noise. The RMS (Root Mean Square) values of each signal was estimated with '*s*' the window length being 1.5 s. This window size was selected as it represented the maximum size of the envelope for the vowels spoken by the subjects.

4 TESTING

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 (Freeman and Skapura, 1991) (Haung, 2001). 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.

Back Propagation (BPN) type Artificial Neural Network has been designed and implemented. 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 were 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 stable learning of the network. The training stopped when the network converged and the network error is 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. Three RMS values of EMG captured during the subject pronounce the vowels were defined as inputs to the ANN. The output of the ANN was one of the five vowels.

5 RESULTS AND DISCUSSION

Table 1: Accuracy of recognition of vowel from EMG

	/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.

6 CONCLUSIONS

This paper describes a study to recognise human 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 speech recognition systems. These preliminary results suggest that the study is suitable to develop a real-time EMG based speech recognition system. This would have number of applications such as for voice control of machines and toys in noisy environment and for people who do not have the gift of speech. It would also find other applications such as for noise reduction for telephonic conversations in noisy environments.

7 FURTHER WORK

Authors are currently working with a larger population of subjects to determine the inter and

intra subject variability. Authors are also conducting experiments for consonants and other sounds and observing the temporal variation of the data.

REFERENCES

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.

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.

"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 Hiroyuki Manabe NTT DoCoMo MultimediaLaboratories†manabe@mml.yrp.nttdocomo.co.jp

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.

"Sub Auditory Speech Recognition Based on EMG Signals" Chuck Jorgensen, Diana D Lee & Shane Agabon.

Akira Hiraiwa NTT DoCoMo Multimedia Laboratories hiraiwa@mml.yrp.nttdocomo.co.jp Toshiaki Sugimura NTT DoCoMo Multimedia Laboratories sugi@mml.yrp.nttdocomo.co.jp

A J Fridlund, J.T.C., *Guidelines for human electrographic research.* Psycholphysiology, 1986. 23: p. 567-589.

A. Freeman and M. Skapura, Neural Networks: Algorithms, Applications, and Programming Techniques, Addison-Wesley, Mass., 1991.

Haung, K.-Y., "Neural networks for robust recognition of seismic patterns,". IEEE Transactions on Geoscience and Remote sensing 2001