VOL. 12, NO. 1, JANUARY 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ELECTROMYOGRAPHIC ANALYSIS FOR SILENT SPEECH DETECTION

Andrés Ussa Caycedo[1], Dario Amaya Hurtado[2] and Olga Lucía Ramos Sandoval[3]
[1]Mecatrónica, Universidad Militar Nueva Granada, Colombia
[2]Ingeniería Mecánica, Universidad Militar Nueva Granada, Colombia
[3]Teleinformática, Universidad Militar Nueva Granada, Colombia
E-Mail: u1801205@unimilitar.edu.co

ABSTRACT

Being speech the most natural way of communicating among humans, it should be possible to use it without complications in any aspect of a human live. Unfortunately this is not possible in certain situations like inappropriate environments or speaking disable people. Unvoiced speech recognition is capable of solving these issues through the acquisition of biological signals directly related to speech. Consequently, an analysis of silent speech recognition systems using electromyographic signals is presented. Applications of this technology in medicine, human interfaces, voiced and unvoiced recognition are showed. A description of hardware and software used in EMG-based projects is realized, along with an introduction to multiple techniques used for feature extraction and classification of myograhpic signals. The results obtained by the different projects are analyzed and the main difficulties still present in this kind of systems are commented.

Keywords: surface electromyography, sEMG, unvoiced, silent, speech recognition.

1. INTRODUCTION

Automatic speech recognition (ASR) is widely found in multiple industries and applications nowadays. It is used in automatic translation, telematics, transcription, home automation, mobile phone applications, among others. It has found a satisfactory level of reliability, accuracy and performance [1]. Although ASR has reached a high level of acceptance among its users, it still has meaningful problems while being used at noisy environments such as construction sites and factories. It can't be used when high privacy is required and it is unfeasible at very specific applications like underwater or spacewalk operations [2, 3].

Unvoiced or silent speech recognition has gained popular approval in the later years for being a suited alternative or improvement to current ASR. It consists of the measurement of vital signals of the face, neck and the back of the ear for its processing and analysis, in order to find a relation between its behavior and the human speech. One of the relevant approaches towards a silent speech recognition system is surface electromyography (sEMG), which intends to acquire a direct interpretation of the nervous system when receiving signals during speech [4]. These signals occur when reading or speaking to oneself without the need of lip or mouth movement. Surface EMG achieves to sense the muscles myoelectric activity through direct skin contact and then collect the information through a recording system, this means it has a non-invasive nature and the data can be stored for later processing and analysis [5]. For these reasons sEMG is used in clinical diagnostics, bioengineering research, man-machine interfaces and control of moving objects.

An analysis of sEMG as a good alternative for a silent speech recognition system is presented in this paper. Applications of sEMG in different fields are described and a comparison between each other is made to see the differences in hardware, software and methods of analysis. A deeper analysis is made in recent silent speech system projects, aiming to show the current state of the technology, the current drawbacks and the techniques applied for acquisition, processing and analysis.

This paper is organized as follows. All the application fields are presented in Section 2. This section includes 4 subsections which are: medicine, human interfaces, voiced speech recognition and unvoiced speech recognition. Finally, the conclusions are exposed in Section 3.

2. APPLICATION FIELDS

A detailed description of projects regarding surface electromyography in various fields is presented in this chapter. Each field has had important findings and developments on their own. The most remarkable studies are analyzed, showing the relevance and contribution to sEMG-based research.

2.1. Medicine

The main application of sEMG since its appearance is medicine. It has spread through several specific medical fields creating sub branches like electroencephalography (brain), electrocardiography (heart), electronystagmography (eyes), etc. Some examples of sEMG applied to medical uses will be displayed in the following paragraphs.

Reference [6] used sEMG as a way to detect vocal hyperfunction, a common condition associated with many voice disorders. They measured myoelectric signals of the neck of 18 individuals with healthy normal voice and 18 individuals with vocal nodules, using two Delsys 3.1 double differential electrodes placed superficial to fibers of the thyrohyoid, omohyoid, and sternohyoid muscles. These electrodes were used to increase spatial selectivity and to minimize electrical cross-talk. Those signals were recorded using 2 two-channel Bagnoli systems (DelSys Inc.), which performed a preamplification and filtering set to a gain of 1000 and a band-pass filter with roll-off frequencies of 20 and 450 Hz. The analysis was based on the calculation of the mean

coherence in the beta band. The results showed that mean coherence was similar in both speech situations, but it was significantly reduced with vocal hyperfunction patients, showing that this measure could be a good indicator of this pathology.

Reference [7] also developed an EMG-controlled electrolarynx, which is a common rehabilitative aid for people who have undergone total laryngectomy. An electrolarynx generally lacks of pitch control and require the exclusive use of one hand. The electromyographic signals were meant to control the onset, offset and pitch of such a device. Seven differential sEMG electrodes (Delsys DE2.1) placed on the face and neck were used in this study (Figure-1). A test was conducted in which 8 individuals who had undergone total laryngectomy performed a speech using an average electrolarynx and the EMG-controlled electrolarynx. A NuVois electrolarynx was used, a digital signal processing board (Motorola DSP56311EVM) for signal processing and a computer running MATLAB for recording and post-processing. The activation and termination thresholds were set individually for each sensor. Each of the electrode's position was evaluated by the participant and the investigator during the test to dictate which electrodes were getting a more sensitive response. The performance was defined as the average percentage of achieved voicing (percentage of fully voiced words of those attempted) and of the percentage of achieved pauses. All participants showed high serial speech performance when using sEMG from electrode recording locations #4 (ventral neck midline), #5 (submental midline), and #6 (corner of the mouth). Finally, all participants were able to produce running and serial speech hands-free with the EMG-EL controlled by sEMG.

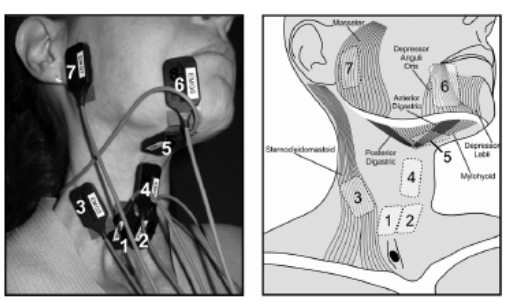


Figure-1. Electrodes location for EMG-controlled electrolarynx [7].

In 2006, [8] presented an analysis of finger braille using EMG. Finger braille is one of the communications methods for the deaf blind, and at the time it seemed like the quickest and most accurate. Prosody is the paralinguistic information that has functions to transmit a sentence structure, prominence and emotion. The aim of this study was to measure the myoelectric activity of the fingers (index, middle and ring) during finger braille in order to analyze prosody in this way of communication. The subject of the test was a male finger braille interpreter. The test consisted of typing a sentence written on a paper. The measurement was performed using compact active electrodes by the Laplacian EMG method. The EMG signals were amplified, sampled at a rate of 2 kHz, low-pass filtered at 5.3 kHz and high-pass filtered at 1 kHz. In order to represent the data recollected in significant values, a full-wave rectification and integration was applied to the signal. The obtained values were called "averaged rectified values" (AVR) and they differed according to the combination of fingers used. Results showed typing strength increases at the beginning of a phrase and a prominent phrase, and also opened the possibility of applying prosody in an interpreter system for the deaf blind.

An approach of signal pattern classification for surface electromyography applications was presented by [9]. The aim of this study was to exploit higher order statistics (HOS) in sEMG feature extraction in a two channel sEMG signal motion classification problem, in this case upper-limb prosthesis control. It intended to classify four upper-limb primitive motions. The signals were obtained from a 24 years man by two pairs of Ag/AgCl electrodes placed in his biceps brachii and triceps brachii. The signal was amplified and band-pass filtered. Sampling rate was set at 1000 using a homemade 12 bit A/D converter board. The use of higher order statistics for feature extraction was meant to get a high classification rate and to improve robustness to noise and interference compared to other techniques like integral of absolute value (IAV). A multilayer perceptron was used to train the network for pattern classification. The average rates of correct classification were 81.41% and 90.70% for IAV and for HOS+IAV methods, respectively. The results showed similar classification rates to other feature extractions techniques, although at the expense of computational complexity.

Similar studies in the clinical field have been executed in the past years, some others are [10], [11] and [12], proving the importance of EMG technology in this field.

2.2. Human interfaces

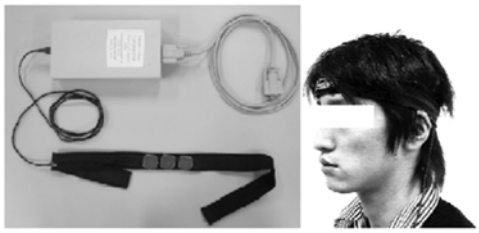
EMG technology has been very useful as an important method of developing human-machine interfaces (HMI). Through this, HMI has found a way to replace classic ways of communication, like mice and touch, and present a novel way of control by "basic thinking". Electromyographic signals are used as a representation of a person intentions, these signals are often acquired using EEG, i.e. using electrodes on the scalp. This method suffers from poor signal-to-noise ratios and usually requires multiple electrodes [13]; hence sEMG is being used as an alternative to correct these problems. The following studies are related to this topic.

Reference [14] performed the design and testing of a human machine interface using EMG to collect signals from a covert location. They acquired the signals bilaterally from the Auricularis Posterior muscle in five participants. These were trained to modulate their muscle activity to move a cursor in two dimensions producing different target vowel sounds. The sEMG recordings were

pre-amplified and filtered using a Delsys™ Bagnoli system set to a gain of 1000, with a band-pass filter with roll-off frequencies of 20 Hz and 450 Hz. sEMG signals were then digitized at 16-bit resolution using a Fast Track Pro USB sampling at 44,100 Hz. A DFT and a summing of the frequency's squared magnitude were used to calculate the power of each input channel. A one-pole IIR filter was used to de-noise the muscle activation. The ability of the participants to correctly position the cursor on the target location within 15 seconds was established as the percentage of success. Results indicate that individuals were able to learn sEMG control of vowel synthesis using auditory feedback alone with an average of 67% accuracy and that this skill could also generalize to new vowel targets.

A nonverbal interface for hands-free control of an electric chair was developed by [15]. This study aimed to recognize an operator's gestures to control linear and turning motions, velocity and steering angle of an electric wheelchair (Figure-2a). The gestures were closing the jaw, wrinkling the forehead, and looking towards left and right. The input data was acquired from EEG, EMG and EOG signals. These were collected through three dry electrodes embedded in a headband (Figure-2b), which was connected to an amplifier of gain 50000 and a bandpass filter of 0.5 to 45 Hz. The signals were separated using a frequency domain analysis algorithm. In order to test the correct operation of the system, an operator gave simple commands to the wheelchair using the gesture interface. The wheelchair moved according to the operator gestures. To show the feasibility of wheelchair operation using the gesture interface, they performed an experiment of indoor navigation in which it traveled 150 m.

(a)



(b)

Figure-2. (a). Electric wheelchair controlled by EMG signals. (b). Headband with embedded electrodes.

An important research regarding this topic was presented by [5]. Their approach consisted of a modified web browser interface controlled through subvocal electromyogram. The signals were measured on the side of the throat near the larynx and under the chin (Figure-3), places associated with the muscle activity of the vocal tract and tongue. Five subjects joined the test in which they had to sub acoustically pronounce six words and ten digits (0-9). The EMG signals were collected from each subject using two pairs of self-adhesive Ag-AgCl electrodes. Each electrode pair was connected to a commercial Neuroscan signal recorder, which recorded the EMG responses sampled at 2000 Hz, with a 60 Hz notch filter (to remove line interference), a 500 Hz low pass filter and a 30 Hz high pass filter. Each pronounced word was segmented and transformed into feature vectors, testing each one of these transformations: STFT, DWT, DTWT, moving averages and LPC. The previous feature vectors were used to train a neural network; multiple techniques for classification were applied, showing the best results with Scaled conjugate gradient nets and Support vector machines. A classification rate of 50% was achieved, removing the alveolar phonemes from the test, which caused some pronunciation problems. This study showed that task environment control using sub acoustic speech is possible. Further improvements of this same author will be shown in the following sections.

Figure-3. Electrode placement for web browser interface controlled through subvocal electromyogram.

2.3. Voiced speech recognition

Research of automatic speech recognition systems has been approached from multiple points of view. It is a well-developed technology now common in several aspects of people's life. Not only it is present in trending technology and commercial products [16], but it can also be found in healthcare, military, aerospace and robotics applications [17]. Besides the voice analysis technique for ASR, there has been EMG approaches, some of them will be presented now.

The development of a methodology to synthesize speech was presented by [18], surface electromyography signals are recorded from the cheek and chin. Two surface electrodes were used as input, an additional electrode was attached to the forehead as a reference point and a microphone was used to record speech. Both, acoustic and

myographic signals were recorded using a National Instruments Inc. PCI6024E PCI data acquisition card at a sampling rate of 8000 Hz. The training set consisted of data samples of seven phonemes. The speech signals were blocked into frames and for each one of these; a speech feature vector was formed, which included linear prediction coefficients, pitch and energy. A two-layer feed-forward back propagation neural network, which takes an sEMG feature vector as input and produced one of eight possible outputs (including silence and seven phonemes), was trained. The trained neural network was then used to classify the vectors and to produce sequences of speech feature indices. This method can be applied not only to phonemes, but also to words or even sentences. A 70.3% of correctly synthesized word was achieved. This way the feasibility of as sEMG based speech recognition was demonstrated. Further studies were performed in [19], where they evaluated the effects of varying the sEMG frame size, establishing a 112.5 ms frame size as the proper value to provide a good balance between time and frequency resolution. This research was also conducted in the [20] thesis project.

Some successful projects were made by [21], [22] and [23], where all of them had recognition accuracy higher to 88%. Similar sensors, amplifiers and recorders where used. Measure places were also similar aiming to recording signals from the face. Their aim was the design and implementation of a voiced speech recognition system. They differed from each other in the classification method. Showing the vast possibilities of feature extraction and identification.

Reference [24] has had an important role in sEMG based automatic speech recognition systems. Since 2005 they have presented important advances in this topic. Their approach is to design a complete EMG speech recognition system, in which a phone-based EMG speech with articulatory features and their relationship with signals of different channels are analyzed. Then a novel EMG feature extraction method is described. And lastly, the integration of feature extraction methods and articulatory feature classifies are added to the system. A test was made with a participant who read 10 times of a set of 38 phonetically-balanced sentences, which were used for training, and 10 times of 12 sentences from news articles, which were used for testing. The EMG signals were recorded with six pairs of Ag/Ag-Cl surface electrodes attached to the face and neck. All the electrode pairs were connected to the EMG recorder (Becker), where EMG responses were differentially amplified, filtered by a 300 Hz low-pass and a 1Hz high-pass filter and sampled at 600 Hz. Articulatory features (AF) are measured from phones, this features are expected to be more robust because this movements are less affected by speech signals and noise. The training of AF classifiers is done on middle frames of the phones only. The performance metrics used in this paper are F-score (for AF) and word error rate (for speech recognition). Performance was measured with different configurations of the system. Some results suggested that some EMG signals were complementary to each other, in which pairs performed better than both their single channels did. In summary, the EMG articulatory feature performance improved from 0.467 to 0.686 and the overall speech recognition word error rate improved from 86.8% to 29.9%, compared to previous studies conducted by the same team. This proved the advantages of better noise robustness and better applicability compared to traditional acoustic speech recognition. A complementary study [25] was also made, in which they evaluated a method to estimate the fundamental frequency from sEMG data, i.e. EMG-TO-F_0 conversion. This was done in order to find alternatives for voice synthesis, using a Gaussian mixture model (GMM)-based voice conversion technique. Further development [26] improved the previously exposed methods using smaller sub-word units, rather than full word models, as prerequisites for a large vocabulary speech recognition. This showed that sub-word units based EMG speech recognizer was possible, and that it even could recognize unseen words with a recognition rate of 62.4%. However, this method showed to need more training data than the formerly used. Finally, a thesis project [27] regarding these researches was realized concerning the problem of variability in recorded data through different acquisition sessions. Formerly known adaptation methods used in acoustic speech recognition were evaluated to establish its feasibility in unvoiced recognition systems, and a multi-session EMG-based recognition system is proposed based on a corpus of EMG data described in this work.

2.4. Unvoiced speech recognition

As depicted previously, unvoiced speech recognition solves some drawbacks present in common ASR, which disable its implementation in certain situations. Some of these are: usage in crowded and noisy places, under-water applications, spacewalk operations, military purposes, need of privacy, and communication interface for speech disabled, among others. Some interesting and important researches were developed in the past years; these are presented in the following paragraphs.

Reference [3] has made multiple studies regarding this subject. Their main objective was to identify speech using the facial muscle activity without the audio signals. Experiments were conducted to identify and classify speech from facial EMG. In it, three male subjects wore Ag/AgCl electrodes (AMBU Blue) on their face in order to record EMG signals from the face (Figure 4). A four channel, portable, continuous recording MEGAWIN equipment (from MEGA Electronics) was used for this purpose. The subjects were asked to speak the 5 English vowels. In order to identify the temporal location of muscle activity, moving root mean square (MRMS) of the recorded signal was computed and thresholded against 1 sigma of the signal. An artificial neural network (ANN) classifier with back propagation learning algorithm was trained using normalized RMS values of each vowel from an individual subject. Some clusters were identified using average linkage method in MATLAB; the first cluster was formed by vowels /a/ and /e/, the second by vowel /i/ and the third by /o/ and /u/. The ANN was able to identify non-

linear separations in clusters. Testing with the test data demonstrated an overall average accuracy of 80% of correct classification when testing and training data belong to the same subject, this way a correct identification between same cluster vowels was obtained. Average accuracy tested with test data belonging to different subjects was lower, and identification between same cluster vowels was not possible. Another paper was published [28] regarding the employment of the previous described system in subjects who spoke English and German. The same acquisition, conditioning, processing and classification were used. Results showed high recognition rates on the evaluated subjects, meaning the system is suitable for multiple languages, and ease to train and use the system. Another two approaches were developed concerning the addition of hand gestures recognition [29] and extraction of visual features of the mouth [30].

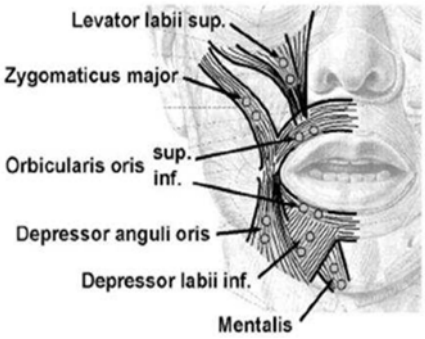


Figure-4. Muscles in which myograhic signals were recorded.

A very interesting work made by [31], conducted the study of the recognition of isolated Spanish language syllables for a silent-speech interface based on EMG signals. The surface electrodes were placed on the facial muscles related to pronunciation and articulation of speech. For testing, a vowel was selected for each group, which was going to be combined with five different consonants. The eight bipolar EMG signals were acquired and digitized (using a gUSBamp amplifier) at a sampling frequency of 2400 Hz, power-line notch-filtered to remove the 50 Hz line interference, and band-pass filtered between 5 and 500 Hz. Three male Spanish speakers participated in the test. Multiple techniques were applied for feature vector calculation, including: Fast Fourier Transform, Root Mean Square, Average amplitude of the signal, Maximum amplitude, Kurtosis, Mel-frequency cepstral coefficients, Mean absolute value and Zero-crossing points. The classifier selected was the boosting algorithm AdaBoost.M1, using the J4.8 decision tree as the weak classifier. The training and classification processes were carried out using the software Weka. 10-fold cross validation showed a global mean recognition rate of 69%, proving a high performance and potential of the recognition system given the large number of classes involved in the problem.

Reference [32] aimed to apply the technological advances from EMG-to-text conversion to the direct EMG-to-speech mapping. They used six channels to capture signals from the face and neck muscles. For EMG recording, they used a computer-controlled 6-channel EMG data acquisition system (Varioport, Becker-Meditec). These recordings were made on five different male subjects in a quiet room, which performed the reading of 50 sentences. The feature extraction for source EMG signals is based on time-domain features. Mel Cepstral Distortions (MCD) was used for evaluation. The MCD is a scaled Euclidian distance between the spectral features of the target audible/whispered speech and the spectral features of the synthesized EMG speech in decibel. Overall results showed better performance from the EMG-to-Whisperer compared to the EMG-to-Audio conversion. They also worked on a sEMG based phone recognition analysis [33], its feasibility and weaknesses, along with a comparison to acoustic based recognition. They found that, even though it was possible to classify phones significantly above chance, it still was way behind the same experiment using an acoustic recognition system. They also pointed that the articulation position was well recognized, the manner of articulation was more difficult to recognize, and the voicing feature even harder. They proposed a rearrangement study of the electrodes position in order to improve these results. Another study [34] was directed to reduce the impact of between session variability because of electrodes repositioning, environmental temperature changes and skin tissue properties. This research showed that an average word accuracy of 97.3% achieved in one recording session, dropped to 76.2% after repositioning, and was restored to 87.1% accuracy after applying normalization and adaptation methods. These methods were based on data sharing between sessions, Variance Normalization and Maximum Likelihood adaptation. Finally, a thesis project [35] was linked to this team, which dealt with initialization methods for silent speech recognition based on sEMG signals, due to the impossibility of audio-generated time-alignments initialization. The study found that this could be achieved through "Cross-Modal Labeling", which involves computing time-alignments for the EMG recordings of silent speech and then training a full EMG recognizer for silent speech recordings.

A voiceless Arabic vowels recognition system using facial EMG was presented by [36]. Their aim was to build a reliable and accurate system for voiceless speech recognition of Arabic language vowels, which are known for its difficult recognition. 20 subjects were involved in the test. It consisted of the recording of EMG data while they said the Arabic vowels in a non-acoustically way. A three channel facial EMG system was used to collect data from the subjects. Three pairs of pre-gelled Ag/AgCl electrodes were used to record the surface EMG signal from three facial muscles that are mainly responsible for speech. Data was amplified and recorded using a signal conditioner and a data acquisition system from National Instruments-USA with a sampling rate of 4050 Hz. A Butterworth band pass IIR filter with a bandwidth of 20-

500 Hz was applied, along with a IIR notch filter used to eliminate noise caused by the 50 Hz power line interference. The data was smoothed by a Savitzky–Golay filter and a cubing operation of the EMG signal. Features were extracted from the segmented EMG signals in three different domains: temporal, spectral, and time–frequency using wavelet packets. Weka software was used for the classification and evaluation of the extracted features. Testing of the method was done based on a 10-fold cross validation. The accuracy of the segmentation procedure tested against the manual annotation made during the recording procedure was 94%, with wavelets packets features having the highest accuracy. This implementation proved that time–frequency information was required to correctly classify the vowels and that a random forest classifier and wavelet packets system could be proposed for that matter.

Reference [37] work has come to a point where a portable, practical silent speech recognition system, called MUTE (Mouthed-speech Understanding and Transcription Engine), was attained. This study was focused on speaker dependent, flexible and unlimited vocabulary continuous recognition tasks, for practical military communication applications. A customized version of the Trigno™ wireless sensors (Delsys Inc.) was used on the subject's face (Figure 5a). The sensors are placed in relation to a template pattern that uses the corners of the mouth and eye as reference points (Figure 5b). The signal segmentation was applied using customized speech activity detection (SAD). Mel-Frequency Cepstral Coefficients and muscle co-activation levels were used as feature vectors. In order to get an unlimited vocabulary speech recognition system, sub-word models were developed. A three-state hidden Markov model was trained base on discriminative feature dimension reduction and minimum phone error of a special operations data set that was collected from three subjects. For training and testing data are from the same domain, their sEMG speech recognition achieved a useful level of recognition accuracy. Under un-trained testing domain, the sEMG recognition accuracy was usable only for a subset of speakers. This showed that sEMG silent speech communication for unconstrained task and user remains a challenging problem. The level of progression on this unvoiced speech recognition system was achieved through several previous studies like [38], in which electrodes' positioning was tested on eleven different neck and face places, and a comparison between vocalized, mouthed and mentally rehearsed speaking modes was realized. Where vocalized speaking got 92.1% recognition accuracy, mouthed 86.7% and mentally rehearsed speaking was enabled to be recognized due to lack of sufficient sEMG activity. Another study was [39], in which they described signal acquisition and processing strategies employed to address challenges during the development of a silent speech recognition system. Some other researches regarding sEMG based silent speech recognition systems were conducted by [40], [41], [42] and [2].

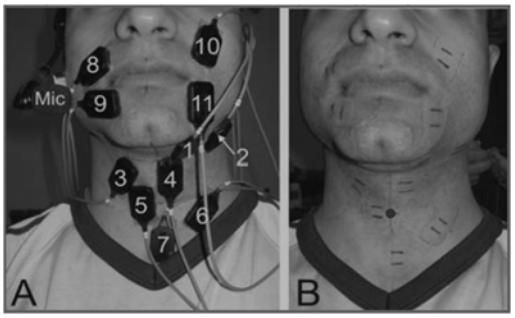


Figure-5. (a). Electrodes in their measurement position. (b). Template pattern used for standard electrode placement.

3. CONCLUSIONS

The studies described in this paper show the progress and current state of sEMG based voiced and unvoiced speech recognition systems. A common EMG signals acquisition configuration was found for most of the researches, where preamplification, filtering and recording remained mostly constant. A range of 2-7 electrodes were generally used. A large number of feature extraction and parameterization techniques were applied for later feature classification. This was mainly performed by artificial neural networks and hidden Markov models. Surface electromyographic based voiced speech recognition showed the highest accuracy, followed by mouthed speech recognition, and unvoiced speech recognition presented the least accuracy, which implies that further research and improvement is required. The majority of tests showed constant performance results when a specific individual was evaluated. The systems still have difficulties on untrained or general purpose subjects. The main differences among documents principally remained on the challenges of the languages used, most specifically on their phonemes.

A lot of projects promise breakthrough developments in this field, where hardware and software improvements are demanded. Some problems to be solved in future research maybe complete session independent speech recognition, multilingual corpus data, less training dependent systems and improvement in phoneme only recognition, in order to achieve higher word recognition accuracy.

REFERENCES

[1] Koester H. 2004. Usage, performance and satisfaction outcomes for experienced users of automatic speech recognition, JRRD, 739-754.

[2] Jorgensen C., Lee, D.D. and Agabont S. 2003. Sub auditory speech recognition based on EMG signals. IJCNN, 3128-3133 Vol. 4.

[3] Arjunan S.P. and Kumar, D.K. Yau W.C. and Weghorn H. 2006. Unspoken Vowel Recognition

Using Facial Electromyogram, EMBS. pp. 2191-2194.

[4] Chan A.D.C., Englehart K., Hudgins B. and Lovely D.F. 2001. Hidden Markov model classification of myoelectric signals in speech, EMBS. 2: 1727-1730.

[5] Jorgensen C. and Binsted K. 2005. Web Browser Control Using EMG Based Sub Vocal Speech Recognition, HICSS, 294c.

[6] Stepp C.E., Hillman R.E. and Heaton J.T. 2010. Use of Neck Strap Muscle Intermuscular Coherence as an Indicator of Vocal Hyperfunction, TNSRE. pp. 329-335.

[7] Stepp C.E., Heaton J.T., Rolland R.G. and Hillman R.E. 2009. Neck and Face Surface Electromyography for Prosthetic Voice Control after Total Laryngectomy, TNSRE. pp. 146-155.

[8] Miyagi M., Nishida M., Horiuchi Y. and Ichikawa A. 2006. Analysis of Prosody in Finger Braille Using Electromyography, EMBS. pp. 4901-4904.

[9] Khadivi A., Nazarpour K. and Zadeh H.S. 2005. SEMG classification for upper-limb prosthesis control using higher order statistics, ICASSP, v/385-v/388 Vol. 5.

[10] Deng Y. et al. 2009. Disordered speech recognition using acoustic and sEMG signals, INTERSPEECH. pp. 644-647.

[11] Wang J.Z. and McKeown M.J. 2007. Relevance Network Modeling for Muscle Association Pattern in Reaching Movements, ICASSP, I-389-I-392.

[12] Zhang L., Xu Z., Li Y., Shang X. and Tian Y. 2012. sEMG activities detection by improved Sobel algorithm, ICICIP. pp. 476-481.

[13] Thulasidas M.X., Cuntai G. and Jiankang W. 2006. Robust classification of EEG signals for brain-computer interface, TNSRE. pp. 24-29.

[14] Larson E., Terry H.P. and Stepp C.E. 2012. Audio-visual feedback for electromyographic control of vowel synthesis, EMBC. pp. 3600 -3603.

[15] Hashimoto M., Takahashi K. and Shimada M. 2009. Wheelchair control using an EOG- and EMG-based gesture interface, AIM. pp. 1212-1217.

[16] Iso-Sipila J., Moberg M. and Viikki O. 2006. Multi-Lingual Speaker-Independent Voice User Interface for Mobile Devices, ICASSP, I-I.

[17] Rabiner L. and Juang B.H. 1993. Fundamentals of Speech Recognition, Prentice Hall.

[18] Lam Y.M., Mak M.W. and Leong P.H.-W. 2005. Speech synthesis from surface electromyogram signal, ISSPIT, 749-754.

[19] Lam Y.M., Leong P.H.-W. and Mak M.W. 2006. Frame-Based SEMG-to-Speech Conversion, MWSCAS. pp. 240-244.

[20] Lam, Y.M. Speech Synthesis from Surface Electromyogram Signals [PhD Thesis]. Hong Kong, China: Chinese University of Hong Kong, 2007.

[21] Kumar S., Kumar D.K., Alemu M. and Burry M. 2004. EMG based voice recognition, ISSNIP. pp. 593-597.

[22] Lee K.S. 2008. EMG-Based Speech Recognition Using Hidden Markov Models with Global Control Variables, TBME, 930-940.

[23] Betts B.J. and Jorgensen C. 2005. Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment.

[24] Szu-chen S.J. and Schultz T. 2008. EARS: Electromyographical automatic recognition of speech.

[25] Nakamura K., Janke M., Wand M. and Schultz T. 2011. Estimation of fundamental frequency from surface electromyographic data: EMG-to-F0, ICASSP. pp. 573-576.

[26] Walliczek M and Kraft F. 2006. Sub-word unit based non-audible speech recognition using surface electromyography.

[27] Schultz T., El Kara K. and Wand M. 2010. Session-Adaptive Speech Recognition Based On Surface Electromyography.

[28] Arjunan S.P., Weghorn, H., Kumar, D.K. and Yau W.C. 2006. Vowel recognition of English and German language using Facial movement (SEMG) for Speech control based HCI.

[29] Arjunan S. and Kumar D. K. 2007. Recognition of Facial Movements and Hand Gestures Using Surface

Electromyogram (sEMG) for HCI Based Applications, DICTA. pp. 1-6.

[30] Yau, W.C., Arjunan, S.P. and Kumar, D.K. Classification of voiceless speech using facial muscle activity and vision based techniques, TENCON, 1-6, 2008.

[31] Lopez-Larraz E., Mozos O.M., Antelis J.M. and Minguez J. 2010. Syllable-based speech recognition using EMG, EMBC, 4699-4702.

[32] Janke M., Wand M., Nakamura K. and Schultz T. 2012. Further investigations on EMG-to-speech conversion, ICASSP. pp. 365-368.

[33] Wand M. and Schultz T. 2011. Analysis of phone confusion in EMG-based speech recognition, ICASSP. pp. 757-760.

[34] Maier-Hein L., Metze F., Schultz T. and Waibel A. 2005. Session independent non-audible speech recognition using surface electromyography, ASRU, 331-336.

[35] Schultz T., Fen, Gu and Wand M. 2010. Initialization Methods for an EMG-based Silent Speech Recognizer.

[36] Fraiwan L., Lweesy K., Al-Nemrawi A., Addabass S. and Saifan R. 2011. Voiceless Arabic vowels recognition using facial EMG, MBEC. pp. 811-818.

[37] Deng Y., Colby G., Heaton J.T. and Meltzner G.S. 2012. Signal processing advances for the MUTE sEMG-based silent speech recognition system, MILCOM, 1-6.

[38] Meltzner G.S. *et al.* 2008. Speech recognition for vocalized and subvocal modes of production using surface EMG signals from the neck and face, INTERSPEECH. pp. 2667-2670.

[39] Meltzner G.S., Colby G., Deng Y. and Heaton J.T. 2011. Signal acquisition and processing techniques for sEMG based silent speech recognition, EMBC. pp. 4848-4851.

[40] Freitas J., Teixeira A. and Dias M. 2012. Towards a silent speech interface for portuguese.

[41] Lam Y.M. 2012. Non-acoustic Communication with Speech Smoothing, SIPIJ. Vol. 3.

[42] Mendes J.A.G., Robson R.R., Labidi S. and Barros A.K. 2008. Subvocal Speech Recognition Based on EMG Signal Using Independent Component Analysis and Neural Network MLP, CISP. pp. 221-224.

[43] Andrew Binley, Andreas Kemna, DC resistivity and induced polarization methods, Hydrogeophysics Water Science and Technology Library Volume 50, 2005, pp. 129-156.

[44] Kyung Hwan Kim and Sung June Kim. 2000. Neural Spike Sorting under Nearly 0-dB Signal-to-Noise Ratio Using Nonlinear Energy Operator and Artificial Neural-Network Classifier IEEE Transactions on biomedical engineering. 47(10).

[45] I. Demirkol and W. Heinzelman. 2014. BaNa: A Noise Resilient Fundamental Frequency Detection Algorithm for Speech and Music. IEEE/ACM Trans. Audio, Speech, Lang. Process. 22(12): 1833-1848.

[46] B. Radish Kumar, J. S. Bhat, and N. Prasad. 2010. Cepstral analysis of voice in persons with vocal nodules. J. Voice. 24(6): 651-3.

[47] A. E. Rosenberg. 1971. Effect of Glottal Pulse Shape on the Quality of Natural Vowels. J. Acoust. Soc. Am. 49(2B): 583-590.

[48] D. P. K. Lun, T.-W. Shen and K. C. Ho. 2014. A Novel Expectation-Maximization Framework for Speech Enhancement in Non-Stationary Noise Environments. IEEE/ACM Trans. Audio, Speech, Lang. Process. 22(2): 335-346.

[49] Hong Hong, Zhengmin Zhao, X. Wang, and Zhiyong Tao. 2010. Detection of Dynamic Structures of Speech Fundamental Frequency in Tonal Languages. IEEE Signal Process. Lett. 17(10): 843-846.

[50] D. O'Shaughnessy. 2013. Acoustic Analysis for Automatic Speech Recognition. Proc. IEEE. 101(5): 1038-1053.