ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

EFFECT OF VOICED CONSONANTS ON EMG SIGNALS GENERATED IN ZYGOMATICUS MUSCLES

Himanshu Raina[1], Randhir Singh[2], Parveen Lehana[3]

M.Tech, Department ECE, Shri Sai College of Engineering and Technology, Pathankot, India[1]

HOD, Dept ECE, Shri Sai College of Engineering and Technology, Pathankot, India[2]

Associate Professor, Dept of Physics and Electronics, University of Jammu, India [3]

Abstract: Speech is natural form of communication between human. Speech signals are non stationery signals. During speech generation contraction and relaxation of the muscles are controlled by the nervous system. This action generates one of the biomedical signals known as EMG. These signals are complicated and depend upon the anatomical and physiological properties of muscles. Facial EMG is recorded for speech recognition and system automation. EMG signals are generally recorded using small surface electrodes placed near to each other. EMG activity is frequently recorded from specific muscles and plays a prominent role in the expression of elementary emotions and speech generation. The present research paper investigates the EMG patterns generated during the utterance of the unvoiced consonants. Six subjects in the age of 20-25 years were taken (three males and three females). Thirty eight vowel-consonant-vowel (VCV) syllables in Hindi were recorded along with the corresponding facial EMG signal. For each speaker, the means of log-spectral-distances (LSD) between the EMG signal of the VCVs and the reference EMG signal were computed. Analysis of the spectrograms and LSD showed that the EMG signals generated in the muscle vary with the subject and the VCV. Female subjects showed high value in standard deviation and mean value calculated for EMG signal. Hence, for automatic decoding of the EMG signals, the system should be trained using both the variants.

Keywords: EMG signal, PARCAS model, speech production, voiced speech signal.

I. INTRODUCTION

Electrical signal acquired from any organ of the body representing a physical variable of interest are biomedical signal. These are function of time and desirable in terms of its amplitude, frequency, and phase [1]. EMG is one of the biomedical signals that measure electrical currents generated in muscles during its contraction representing neuromuscular activities. Contraction and relaxation of the muscles are controlled by the nervous system. Hence, the EMG signal is complicated signal depending upon the anatomical and physiological properties of muscles [2]. It is also referred to as myoelectric activity. Surface EMG is the technique of collecting information from the muscle action potential. Invasive and non-invasive electrodes are the types of electrodes used to acquire muscle signal. Composite signal is obtained from electrodes mounted directly on the skin which is the combination of all the muscle fibre action potentials occurring in the muscle lying under the skin [3]. As these signals occur at random intervals, thus EMG can be positive or negative voltage. In order to acquire individual muscle fibre action potential, single wire or needle electrode can be place directly into the muscle [4]. Motor unit action potential (MUAP) is the combination of all the muscle potentials from all the muscle fibre of the single motor unit. It can be detected non-invasive electrodes located near this field, or by invasive electrode inserted in the muscle. EMG signal can be shown by simple model given in the equation [5]

$$x(n)=\sum_{r=0}^{N-1}h(r)e(n-r)+w(n)$$

where $x(n)$ is modelled EMG signal, $e(n)$ is the point processed, $h(r)$ represents the firing impulse, $w(n)$ is the zero mean addictive white Gaussian noise and N is the number of motor unit firings. Human body has same number of positive and negative charges thus making it electrically neutral. But in the resting state, the nerve cell membrane is polarized due to differences in the concentrations and ionic composition across the plasma membrane, thus causing to

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

develop potential. Information can be given off in the form of voice or speech from mouth. In the process of information passing, the facial muscles play a dominant role to accomplish the information acquiring and message transmitting because facial muscles can express emotions. We may, therefore, reasonably conclude that measuring the facial electromyography (EMG) will get much more useful information [6].

II. HUMAN SPEECH PRODUCTION

Speech is a natural form of communication for human beings. Figure 1 shows the human speech produced by vocal organs [7].

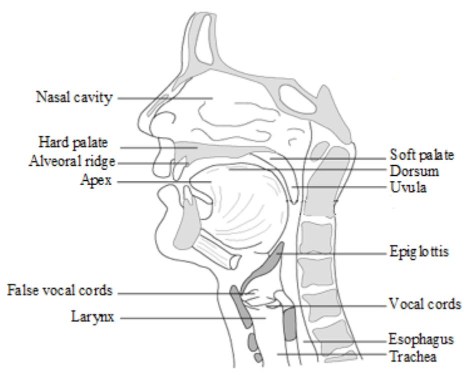

Fig. 1 The human vocal organs [7].

Lungs are the main energy source with diaphragm. They force air flow through the glottis between the vocal cords and the larynx to the three main cavities of the vocal tract, the pharynx and the oral and nasal cavities. From the oral and nasal cavities the air flow exits through the nose and mouth, respectively. Glottis is the V-shaped opening between the vocal cords is the most important part in the sound source in vocal system. During speech production vocal cords may act in several different ways, the most important function is to modulate the air flow by rapidly opening and closing which cause buzzing sound from which vowels and voiced consonants are produced [8].

Several artificial vocal tract models have been developed to produce human speech by many researchers [9-14]. In this research paper two artificial models are discussed. U. Laine in 1982 introduced and patented PARCAS (Parallel-Cascade) model for SYNTE3 speech synthesizer for Finnish. Figure 2 shows the PARCAS model consists of uniform function of vocal tract modelled with two partial transfer functions, each including every second formant of the transfer function. Coefficients k1, k2, and k3 are constant and chosen to balance the formant amplitudes in the neutral vowel to keep the gains of parallel branches constant for all sounds [15].

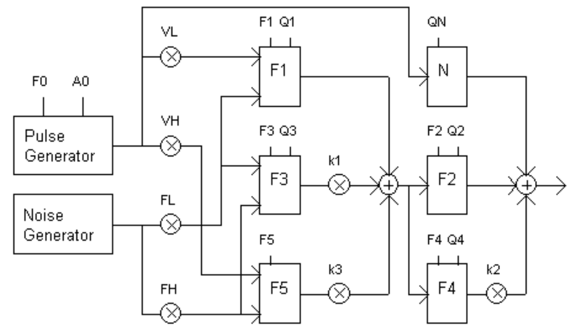

Fig. 2 Schematic of PARCAS model.

Sixteen control parameters are used by the PARCAS model

- F0 and A0 - fundamental frequency and amplitude of voiced component.
- Fn and Qn - formant frequencies and Q-values (formant frequency/bandwidth).
- VL and VH - voiced component amplitude, low and high.
- FL and FH - unvoiced component amplitude, low and high.
- QN - Q-value of the nasal formant at 250 Hz

Second model is the PSOLA (Pitch Synchronous Overlap Add) method was originally developed at France Telecom (CNET). It is not a synthesis method itself but allows concatenation of prerecorded speech samples. It also provides good control over pitch and duration. ProVerbe and HADIFIX commercial synthesis systems uses PSOLA model. PSOLA algorithm has many versions but their essences are same. Time domain version (TD-PSOLA) is commonly used due to its computational efficiency [16]. The basic algorithm consists of three steps: analysis step in which where the original speech is segmented; modification of each analysis signal to synthesis signal; synthesis step where these segments are recombined by means of overlap-adding. Small signals xm(n) are obtained from digital speech waveform x(n) by multiplying the signal by a sequence of pitch-synchronous analysis window hm(n):

$$x_m(n) = h_m(t_m - n)x(n)$$

where m is an index for the short-time signal. Hanning type windows are usually used which are centered around the successive instants tm, called pitch-marks. The segment recombination in synthesis step is performed after defining a new pitch-mark sequence.

III. METHODOLOGY

For investigating facial EMG patterns generated by uttering voiced consonants in VCV syllables, experiments were conducted with the six subjects (three males and three females) having age between 20-25 years. The speech signals and the corresponding EMG signals were recorded

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

using a data acquisition system at the sampling frequency of 16 kHz and 16-bit quantization. Electrodes were placed at zygomaticus minor, zygomaticus major, and mentalispoints on the face. Block diagram of the experimental setup for recording is shown in Fig. 4. The recorded signals were segmented and labeled manually into separate files for each of the VCVs. The signals were analyzed using time-domain patterns, spectrograms, and mean log-spectral-distances. For computing LSD, the first VCV of each subject was taken as the reference.

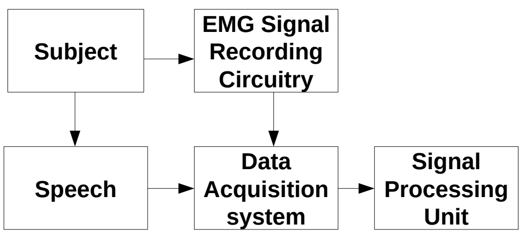

Fig.3 Schematic block diagram for EMG signal acquisition.

IV. RESULTS AND DISCUSSION

Investigations were carried out in time-domain and the corresponding spectrograms of the speech and facial EMG signals for two voiced consonants आगा and आजा are shown in Fig. 4 to Fig. 9 for six subjects (Sp1 to Sp3 are male subjects and Sp4 to Sp6 are female subjects). Here the x-axis represents the normalized time and y-axis represents the normalized frequency. The signals and the corresponding spectrograms show that the signals generated vary across the subjects and the syllables. Table I shows the means and standard deviations of voiced consonants आगा and आजा of all the six speakers. Analysis of the mean LSDs also suggests that the distances are a function of subject and VCV. Hence, training of the automated EMG recognition system needs both the EMG signals and the information regarding subjects.

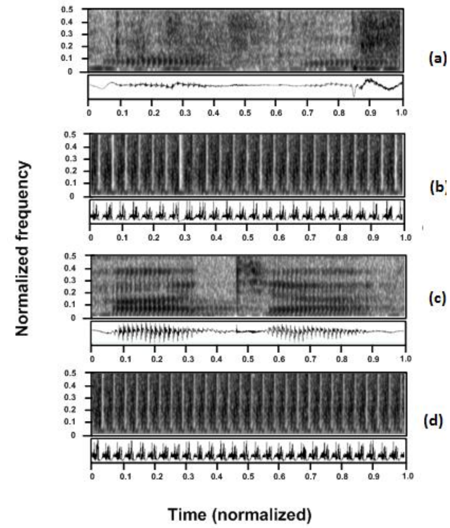

Fig. 4 Spectrogram of voiced consonant (a) Original speech signal (आगा) (b) Facial EMG signal (आगा) (c) Original speech signal (आजा) (d) Facial EMG signal (आजा) for Sp1.

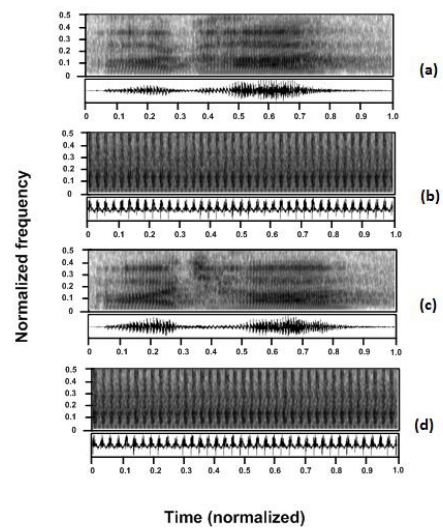

Fig. 5 Spectrogram of voiced consonant (a) Original speech signal (आगा) (b) Facial EMG signal (आगा) (c) Original speech signal (आजा) (d) Facial EMG signal (आजा) for Sp2.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

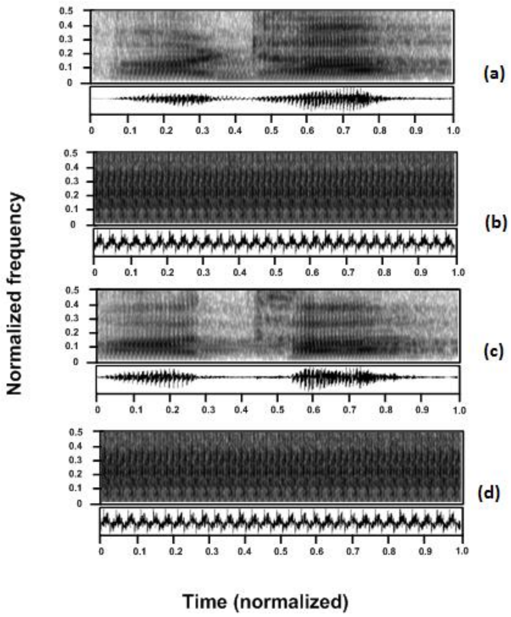

Fig. 6 Spectrogram of voiced consonant (a) Original speech signal (आगा) (b) Facial EMG signal (आगा) (c) Original speech signal (आजा) (d) Facial EMG signal (आजा) for Sp3.

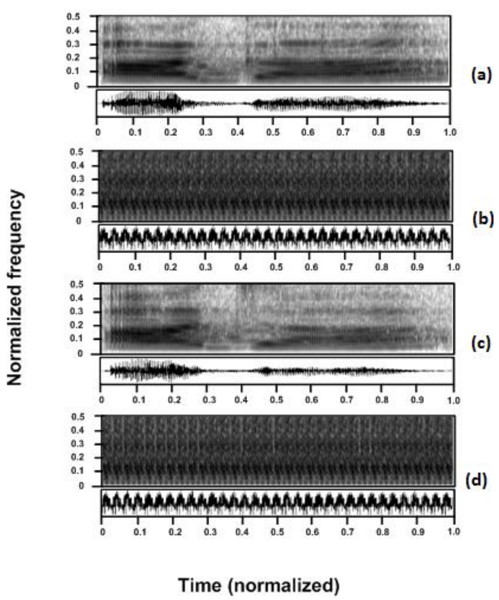

Fig. 7 Spectrogram of voiced consonant (a) Original speech signal (आगा) (b) Facial EMG signal (आगा) (c) Original speech signal (आजा) (d) Facial EMG signal (आजा) for Sp4.

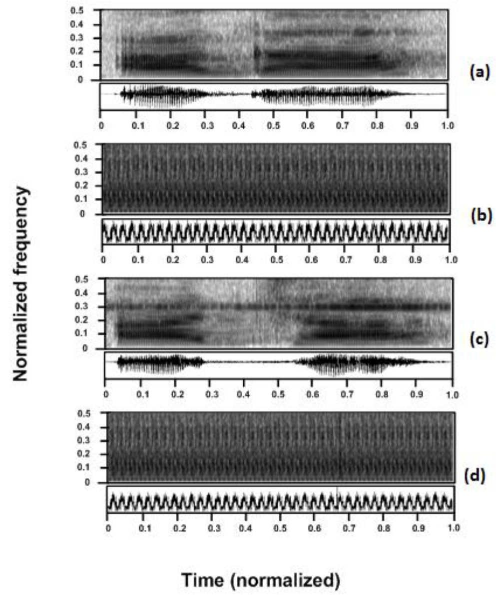

Fig. 8 Spectrogram of voiced consonant (a) Original speech signal (आगा) (b) Facial EMG signal (आगा) (c) Original speech signal (आजा) (d) Facial EMG signal (आजा) for Sp5.

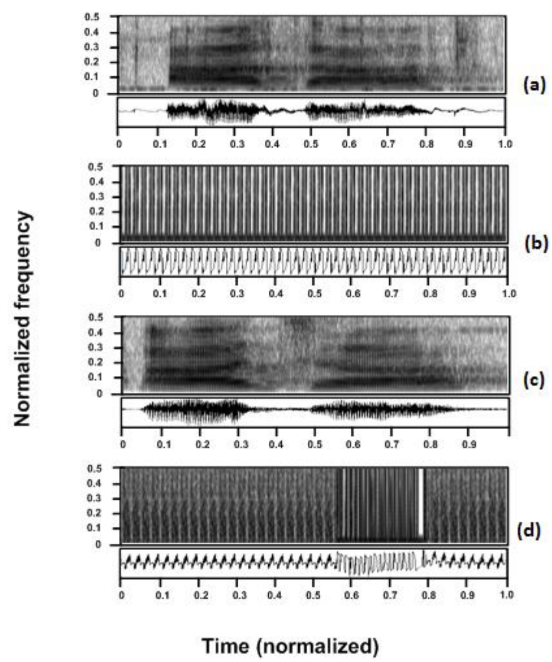

Fig. 9 Spectrogram of voiced consonant (a) Original speech signal (आगा) (b) Facial EMG signal (आगा) (c) Original speech signal (आजा) (d) Facial EMG signal (आजा) for Sp6.

Table 1. Mean and standard deviation of two voiced consonant.

Speaker	Voiced consonant (आगा)		Voiced consonant (आजा)	
	Mean	S.D.	Mean	S.D.
Sp1	7.59	2.04	8.77	2.52
Sp2	7.31	1.56	7.97	1.77
Sp3	5.48	1.22	6.12	1.08
Sp4	6.64	0.85	5.78	1.02
Sp5	5.69	1.15	6.73	1.05
Sp6	9.81	5.66	12.91	6.27

V. REFERENCES

[1] M. Gazzoni, D. Farina, and R. Merletti, “A new method for the extraction and classification of single motor unit action potentials from surface EMG signals,” *Journal of Neuroscience Methods*, vol. 136, pp. 165–177, 2004.
[2] C. L. Nikias, M. R. Raghuveer “Bispectrum estimation: A digital signal processing framework,” *IEEE Proceedings on Communications and Radar*, vol. 75, pp. 869-891, 1987.
[3] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, “Techniques of EMG signal analysis: detection, processing, classification and applications,” *Biological procedures online*, vol. 8, pp. 11-35, 2006.
[4] C. I. Christodoulou, and C.S. Pattichis, “Unsupervised pattern recognition for the classification of EMG signals,” *IEEE Trans. Biomedical Engineering*, vol. 46, pp. 169-178, 1999.
[5] J. V Basmajian, C. J de Luca., *Muscles Alive: The Functions Revealed by Electromyography*, The Williams & Wilkins Company, Baltimore, 1985.
[6] H. C. Ning, C. C. Han, C. H. Yuan, “The Review of Applications and Measurements in Facial Electromyography,” *Journal of Medical and Biological Engineering*, vol. 25, pp 15-20, 2004.
[7] J. Flanagan, *Speech Analysis, Synthesis, and Perception*, Springer-Verlag, Berlin-Heidelberg-New York, 1972.
[8] D. O. Saughnessy, *Speech Communication - Human and Machine*, Addison-Wesley, 1987.
[9] X. S. El Masri, Pelorson, P. Saguet, and P. Badin, “Vocal tract acoustics using the transmission line matrix (TLM) method,” in *Proc. of the Fourth International Conference on Spoken Language Processing*, Philadelphia, 1996, pp. 953–956.
[10] C. Lu, T. Nakai, and H. Suzuki, “Finite element simulation of sound transmission in vocal tract,” *J. Acoust. Soc. Jpn*. 1993, pp. 2577–2585.
[11] N. Umeda, and R. Teranishi, “Phonemic feature and vocal feature: Synthesis of speech sounds, using an acoustic model of vocal tract,” *J. Acoust. Soc. Jpn.*, pp. 195-203, 1966,
[12] T. Arai, “The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity,” *J. of the Phonetic Society of Jpn.*, 2001, pp. 31-38.
[13] T. Arai, E. Maeda, N. Saika, and Y. Murahara, “Physical models of the human vocal tract as tools for education in acoustics,” in *Proc. of the First Pan- American/Iberian Meeting on Acoustics*, Cancun, 2002.
[14] T. Lander and T. Arai, “Using Arai‘s vocal tract models for education in Phonetics,” in *Proc. ICPhS*, Barcelona, pp. 317-320, 2003.
[15] U. Laine “PARCAS, A New Terminal Analog Model for Speech Synthesis,” in *Proc. ICASSP*, vol. 82, 1982.
[16] R. Kortekaas and A. Kohlrausch “Psychoacoustical Evaluation of the Pitch-Synchronous Overlap-and-Add Speech-Waveform Manipulation Technique Using Single-Formant Stimuli.” *Journal Acoustical Society of America*, Vol. 101, 1997, pp. 2202-2213.