

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia Engineering 32 (2012) 253 – 259

**Procedia
Engineering**www.elsevier.com/locate/procedia

I-SEEC2011

Feature selection for Thai tone classification based on surface EMG

N. Srisuwan^{*}, P. Phukpattaranont, C. Limsakul^{*}*Department of Electrical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla, 90112, Thailand***Elsevier use only:** Received 30 September 2011; Revised 10 November 2011; Accepted 25 November 2011.

Abstract

This paper aims to investigate the features of surface electromyography (sEMG) which can classify the Thai tonal sound for the EMG speech recognition and synthesis system. Signals were captured at seven positions on the strap muscles as a subject was uttering nine monosyllabic words which each of the words includes five tones. Eight features, i.e. Root Mean Square (RMS), Variance (VAR), Waveform Length (WL), Willson Amplitude (WAMP), Median Frequency (MDF) and three types of the Spectral Moment (SM), were computed and plotted on the scatter graph to cluster the tones. The results indicate that the EMG signal of the strap muscles can clearly classify the tones into three groups, i.e. a rising tone, a high tone, and the remainder clustered as one group. Moreover, RMS, VAR and WL can classify the high tone better than other features. All of the Spectral Moments yield similarly the results of classification, especially they can classify well the rising tone. For the remainder of the tones, while the scatter graphs are considered without the rising tone and the high tone, the low tone can be separated from their group when WL or WAMP is used for classification only.

© 2010 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of I-SEEC2011
Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

Keywords: Electromyography; Feature Extraction; Thai Tone; sEMG based-Speech Recognition

1. Introduction

Nowadays, Automatic Speech Recognition (ASR) technology has continuously been developed and applied with a variety of tasks. However, it still has many problems. The most important issue is that an input signal is drastically interrupted by external noises. This causes a system to malfunction. Moreover, when we want to use this system to communicate by private or secret data which cannot be disclosed to the public, a user and data may be unsafe. This system is inappropriate for this task. In addition, the automatic speech recognition is unable to work as well as usual when it is used in the unsuitable

^{*} Corresponding author. *E-mail address:* jivalin_eng@hotmail.com, chusak.l@psu.ac.th.

environment such as using for firefighter suit, HAZMAT suit and respirator mask. From these problems, many researches can overcome them by using Electromyography from the muscles producing speech to be input of system replace a voice signal. Research on EMG speech recognition first appeared in 1985. At that time, Sugie [1] captured EMG signal from around the mouth muscles to recognize 5 Japanese vowels and the result was an accuracy rate of 60%. Simultaneously, Morse and O'brien [2] used 4 channels to detect EMG signal from the neck and the forehead muscles, and a correct classification rate of 97% was achieved for 2 words. However, when ten words were recognized, the precision reduced to 60%. In 2001, A.D.C Chan [3] proposed method that can improve performance of ASR used in the cockpit which it has significantly noise. EMG and voice signal were an input of the ASR. The consequence was an accurate recognition rate of 98% for ten English isolated word classifications. After that, Manabe [4] first accomplished the unvoiced speech recognition in 2003. Three channels of the face muscle were captured to recognize 5 Japanese vowels, and classification rate was 94%. However, most researches have problem about session dependent; in other words, the classification rate depends on changing of electrodes' positions, subject and others when the data were collected at different time. In 2004, Maier Hein [5] proposed Normalization and Adaptation Method to overcome this problem. Since 1986 to 2004, all researches have been focused on isolated word recognition. Continuous speech recognition started in 2006 by Szu-Chen Jou [6]. The feature extraction method for continuous speech was proposed at that time. Recently, Tanja Schultz [7] proposed new approach to improve performance of this system which this method was Speaker-adaptive training (MLLR adaptation). The large corpus was used for this research, 13 speaker and 101 words. The result was a correct recognition rate of 68.92% for continuous speech.

The prior researches are based on English language which is a universal language used around the world. For the Asia, Japanese was used widely to recognize this system because the most of researcher is Japanese. For our research, we will develop this system based on Thai language recognition so that Thai user who cannot speak English can use it.

Thai language has a characteristic that is different from other ones; namely, it is a tonal language which the meaning of a word will be different for the distinct tone. Hence, the first task that we must investigate is the classification of Thai tones which is proposed in this paper. EMG signals from the strap muscles were used for tone classification. This idea is derived from the research of Donna Erickson in 1993 [8]. He used the hooked-wire electrode to capture EMG signal from the strap muscles, i.e. Cricothyriod, Sternohyoid, Thyrohyoid, Cricothyroid, Sternohyoid and Vocalis. He found that pattern of EMG signal of each tone is distinct, and it directly relates with fundamental frequency (F0) contour of each tone voice. For this paper, we will study about tone classification by using the surface electrode to capture surface Electromyography (sEMG) from above muscles.

The next section, we will discuss about the theory used in this experiment, i.e. speech production, Thai tonal sound, Electromyography and feature extraction. Then, we will discuss the experiments which are mentioned the step of them, results, discussion and conclusion, respectively.

2. Methodology

2.1. Thai tonal sound

Thai speech consists of the consonant sound, vowel sound and tonal sound. Tonal sound is the main characteristic of Thai sound system. Tones arise from changing of the pitch which has many patterns as shown in fig.1. One pattern is one tone, and a syllable or a word of each tone will have a distinct meaning. Pitch is caused by stretching and tensing of the vocal cords [9]. These behaviors occur rapidly, so it can be considered to be vibration of the vocal cords. When the vocal cords vibrate with high

frequency, the pitch will be high, and vice versa. Changing of the subglottal pressure is another cause of the pitch. The high pitch will occur when pressure significantly changes, and vice versa. These processes result from the collaboration of the strap muscles. As a result, the muscle activities from these muscles were captured to classify Thai tones which have 5 levels, i.e. mid tone (/ ˊ/), low tone (/ ˋ/), falling tone (/ ˆ/), high tone (/ ˊˊ/) and rising tone (/ ˊˋ/). The symbols of these tones which represent a phoneme of the tonal sounds are defined by the International Phonetic Alphabet (IPA).

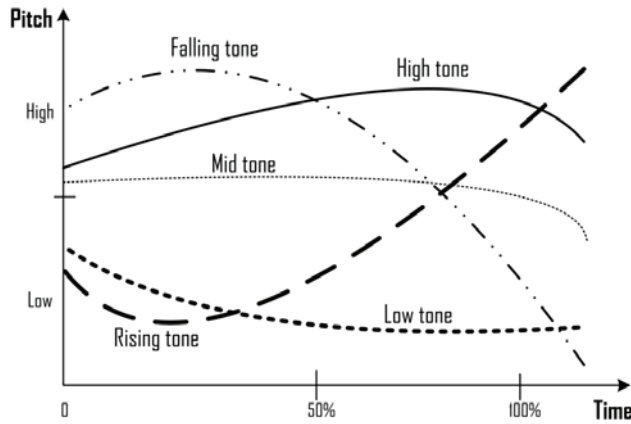


Fig. 1. F0 contours or pitches changing of five Thai tones

2.2. Electromyography and feature extraction

Electromyography will occur when the muscle fiber is activated by the central nervous system. A small current will be produced. It will move to the body tissues through inner electrical resistance of these tissues. This activity makes potential difference happen on the skin. EMG has very small amplitude, 0-10 mV_{p-p} or 0-1.5 mV_{rms} [10]. Thus, it is necessary to amplify the signal before preprocessing and filtering procedure. After that, the signal will be computed to determine the representation of the signal, known as feature extraction.

The feature extractions used in this experiment have 8 features which include both time domain feature and frequency domain feature, i.e. Root Mean Square (RMS), Variance (VAR), Waveform Length (WL), Willson Amplitude (WAMP) and three types of the Spectral Moment (SM1, SM2 and SM3).

Let x_n be the voltage sampled in time n , N is the total number of data, P_i is Power spectrum at i^{th} frequency, M is $(N/2) - 1$.

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2}, \quad VAR = \frac{1}{N-1} \sum_{n=1}^N x_n^2, \quad WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \quad (1), (2), (3)$$

$$WAMP = \sum_{n=1}^{N-1} f(|x_n - x_{n+1}|), \quad f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$\sum_{i=1}^{MDF} P_i = \sum_{i=MDF}^M P_i = \frac{1}{2} \sum_{i=1}^M P_i \tag{5}$$

$$SM1 = \sum_{i=1}^M P_i f_i, \quad SM2 = \sum_{i=1}^M P_i f_i^2, \quad SM3 = \sum_{i=1}^M P_i f_i^3 \tag{6), (7), (8)}$$

3. Experiments

3.1. Data collection

Positions of electrodes on the strap muscles were divided to two sets, 4 positions for the first set and 3 positions for the second set as shown in fig. 2. The electrodes which were used in this experiment are Ag/AgCl surface electrodes (ARBO, Tyco Healthcare Group LP, Germany). Their sizes are 2.4 cm of the outer diameter and 1 cm of the inner diameter. Signals were captured with sampling rate at 2048 Hz on these positions by the Mobi 6 amplifier (TMS international, the Netherlands) while a female subject was uttering the nine monosyllabic word, i.e. /ka:/, /jo:/, /du:/, /be:/, /ʔɔ:/, /ti:/, /pɜ:/, /ka:n/ and /ka:ŋ/. Each of these words consists of 5 tones. For one recording, there is one data set of 5 tones for each channel, and space between each tone is about 3 seconds. Signals were recorded at 3 sessions, 9 recording per session, so we have 27 recordings.

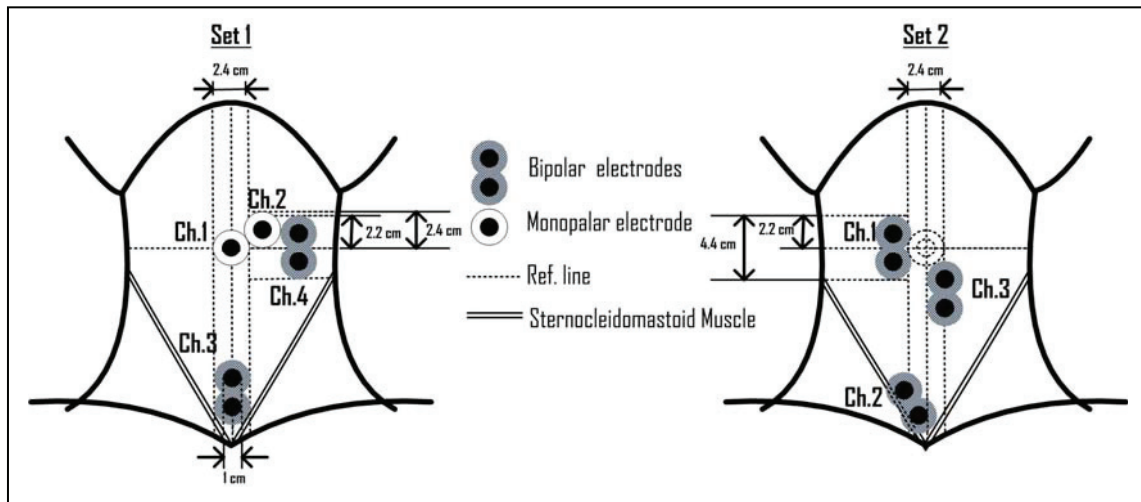


Fig. 2. Positions of electrodes for set 1 and set 2 experiments

3.2. Preprocessing and Feature Extraction

Because the captured signal was disturbed by noises such as motion artifact noise, harmonic noise, 50 Hz noise and ECG noise, the filtering is necessary for this task. Band pass filter, cut-off frequency is 60 to 500 Hz, and band stop filter, cut-off frequency is 48 to 52 Hz, were used for eliminating these interferences. Motion artifact, Harmonic and ECG noise were eliminated by band pass filter, and 50 Hz interference was eliminated by band stop filter. After filtering, the following step is segmentation. For

each recording, the data of 2500 samples from each tone were extracted. They include begin to end of speech. Then, these data will be used to compute the eight features, and these values will be plotted on the scatter graph by Microsoft Excel. Labview was used for the filtering and segmentation process.

3.3. Results

When surface EMG are considered in time domain, the signal that represents each tone will be noticeably different, as shown in fig.3. For the signal in Frequency domain, each tone will be slightly different except the mid tone and the rising tone, for these tones sometimes will distinctly differ from another tone; that is, they will have large amplitude.

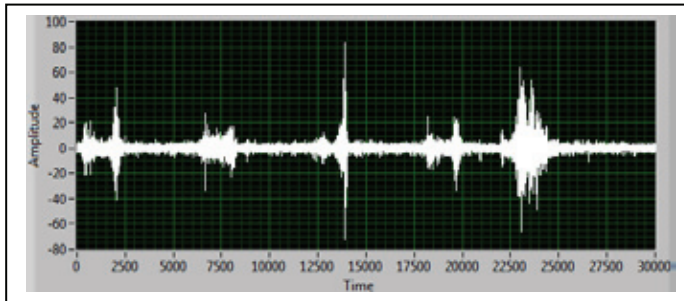


Fig. 3. EMG signal of one recording for words /du:/ with 5 tones, mid, low, falling, high and rising tone, respectively

After plotting the features on scatter graph which X axis denotes the feature of one channel and Y axis stands for same feature of another channel, these graph indicate that total data are divide to 3 groups, rising tone, high tone and the remainder clustered to be one group. All words of the high tone have similar value of feature, whereas the rising tone is different for total words. However, they are cluttered to be one group and meagerly overlap with other tone. For considering the features, Root Mean Square, Variance and Waveform Length are proper for the high tone classification because the graph of these features have a little overlapping for different tone. The rising tone can be classified by three types of the Spectral Moments well especially the third types because of the same reason mentioned above. For the remainder of tone, if scatter graph is considered only these tones by ignoring high and rising tone as shown in fig.4(c), the low tone can separate from this group when Waveform Length or Wilson Amplitude are used only. The channels that can classify these tone well are the third channel of set 1 (ch3), the first channel of set 2 (ch1(2)), the third channel of set 2 (ch3(2)) and the forth channel of set 1 (ch4), respectively.

3.4. Discussion

Results show that the surface EMG from strap muscles is affected by the different tone classification, and the used features must be proper. For this experiment, we select the features that are easy to compute and uncomplicated. These features include both time domain and frequency domain. We find that they can classify three tones, i.e. rising tone, high tone and low tone. For mid tone and falling tone, they cannot be classified in this experiment although two tones have different signal in time domain. The reason is that feature selection may be inappropriate for these tones. We will investigate it in the future work. For the electrodes' positions, there are four locations that yield well Thai tonal classification. However, we need only two positions for our research. Beside the Cricoids cartilage (ch1(2)) and on the lowest part of

the neck (ch3) are selected for this task because they yield best classification. Namely, there is a little overlapping for different tone. These positions include four muscles, i.e. Sternothyroid, Sternohyoid, Omohyoid and Thyrohyoid muscle. For overlapping, if we can minimize this problem, classification will be more accurate. This will be investigated in the next task.

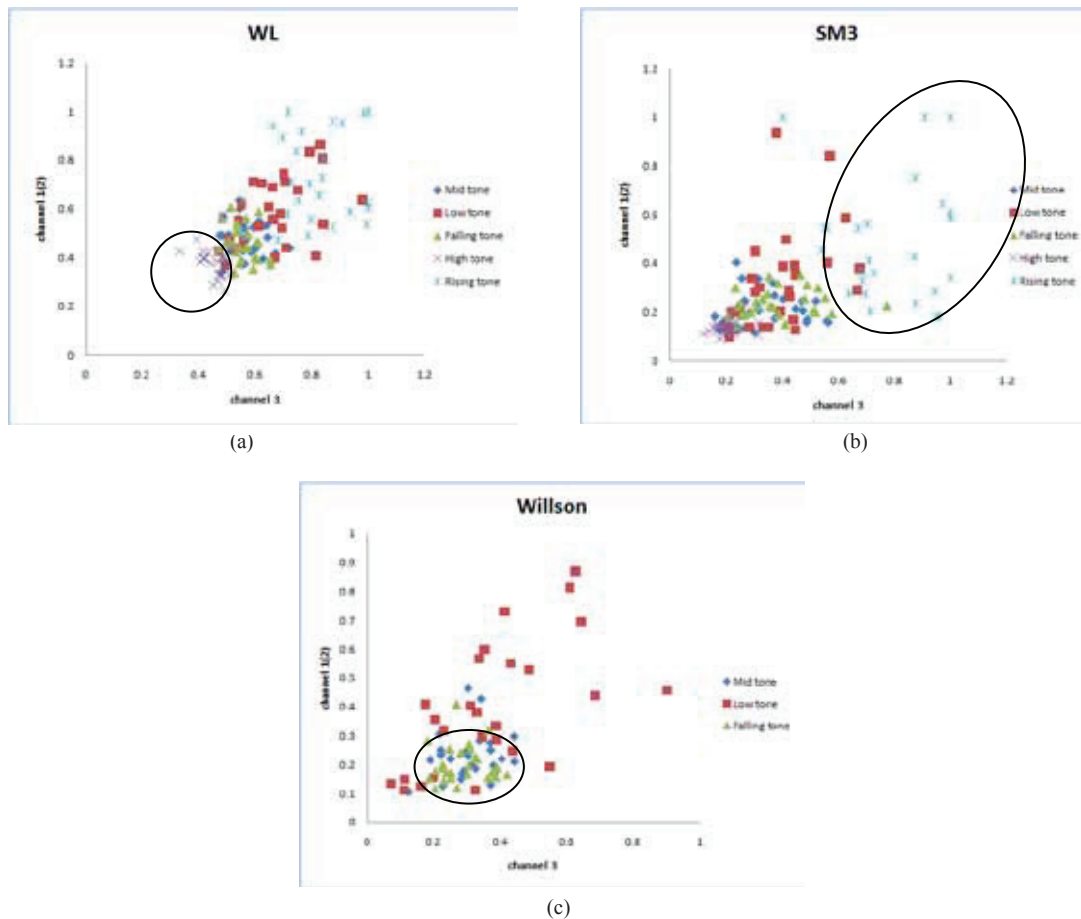


Fig.4. (a) Scatter graph of WL feature for the high tone classification; (b) Scatter graph of SM3 feature for the rising tone classification; (c) Scatter graph of Willson feature by considering only three tones for the low tone classification(the low tone is outside a circle)

4. Conclusion

This paper proposes the suitable feature selection for Thai tone classification based on surface Electromyography. Electrodes' positions which were selected for this task are beside the Cricoids cartilage and on the lowest part of the neck which include four muscles of the strap muscles. Eight types of the features were tested. We find that RMS VAR and WL can classify the high tone and 3 type of the Spectral Moment can classify the rising tone. Moreover, WL or WAMP only is appropriate for the low tone classification. For the mid tone, they cannot be separated from the falling tone although EMG signal

of each tone is different. Thus, the suitable feature extraction for such two tones will be investigated in the future work.

Acknowledgements

The authors would like to particularly thank the Royal Golden Jubilee Ph.D. program of the Thailand Research Fund and Prince of Songkla University that provides the scholarship for our research.

References

- [1] Sugie N, Tsunoda K. A speech prosthesis employing a speech synthesizer-vowel discrimination from perioral muscle activities and vowel production. *J IEEE Trans. Biomed Eng* 1985;**32**:485-490.
- [2] Morse MS, O'Brien EM. Research summary of a scheme to ascertain the availability of speech information in the myoelectric signals of neck and head muscles using surface electrodes. *J Comput. Bio Med* 1986;**16**:399-410.
- [3] Chan ADC, Englehart K, Hudgins B, Lovely DF. Myo-electric signals to augment speech recognition. *J Med. Biological Eng. Comput* 2001;**39**:500-504.
- [4] Manabe H, Hiraiwa A, Sugimura T. Unvoiced speech recognition using EMG-mime speech recognition. In: *Proc.ACM conf. Human Factors. Comput Systems*, New York, USA 2003, pp.794-795.
- [5] Marier-Hein L, Metze F, Schultz T, Waibel A. Session independent non-audible speech recognition using surface electromyography. In: *IEEE Workshop. Automatic Speech Recognition and Understanding*, San Juan, Puerto Rico 2005, pp.331-336.
- [6] Jou SC, Schultz T, Walliczek M, Kraft F, Waibel A. Towards continuous speech recognition using surface electromyography. In: *Proc. Interspeech*, Pittsburgh, PA 2006, pp.573-576.
- [7] Wand M, Schultz T. Towards Speaker-adaptive Speech Recognition based on Surface Electromyography. In: *Proc. BIOSIGNALS*, Porto, Portugal 2009, pp.155-162.
- [8] Erickson D. Laryngeal muscle activity in connection with Thai tones. *Annual Bulletin Research Institute Logopedics and Phoniatrics*, University of Tokyo 1993, 27, pp.135-149.
- [9] Ohala J. The physiology of tone. In: Hyman LM, editors. *Consonant types and tone. Southern California Occasional Papers in Linguistics*, University of California, 1972, pp. 1-14.
- [10] De luca CJ. Surface Electromyography: Detection and Recording. Tech. Rep. DelsysInc, 2002.