

Comparison of Classifiers for EMG based Speech Recognition

N Srisuwan¹, P Prukpatthanont² and C Limsakul²

¹Department of Electrical Engineering, Faculty of Engineering and Architecture, Rajamangala University of Technology Suvarnabhumi, 60 Moo 3 Asian Highway, Phranakhon Si Ayutthaya, 13000, Thailand

²Department of Electrical Engineering, Faculty of Engineering, Prince of Songkla University, 15 Kanjanavanich Road, Kho Hong, Hat Yai, Songkhla, 90112, Thailand

jivalin.eng@gmail.com

Abstract. In this paper, we propose a performance comparison of eight classifiers for speech recognition based on EMG signals to find an optimal classifier. An experiment was divided into two parts, 11 isolated Thai words classification and five Thai tones classification. The first part, EMG signals from five positions of the facial and neck muscles were captured while ten subjects uttered 11 Thai number words in both audible and silent modes. The second part, the subjects uttered 21 Thai isolated words including five tones for each word in audible mode only. Nine EMG features selected from RES index were employed and classification results of eight classifiers were compared in classification process. The results showed that a Fisher's least square linear discriminant (FLDA) and a linear Bayes normal (LBN) classifier yielded the best result, an average accuracy was 90.01% and 79.18%, for 11 isolated Thai word classification in the audible and the silent modes, respectively. Moreover, Logistic Linear (LOGL) classifier gave the best average accuracies, of 68.36% for five Thai tone classification.

1. Introduction

Automatic speech recognition based on electromyography signals or "EMG speech recognition" is a further expansion beyond traditional automatic speech recognition (ASR) using a human voice input. The first research of EMG speech recognition was reported in 1985 by Sugie et al. [1]. The purpose was to solve the problems of expanding numbers of patients with speech disorders. Three EMG signal channels from areas near to the mouth were captured to classify five Japanese vowels. Hence, that study provided the prototype for research in this field. In the early phases, studies concentrated on classifying a few isolated words [2], or a vowel [1, 3-4], a consonant, or a syllable [5]. In this reason, the potential applications were very limited. In 2006, continuous speech recognition based on EMG was explored by Jou et al. [6] to solve that limitation. Using classifiers trained to recognize phonemes based on acoustic models was employed which 108 English words were classified.

Because of distinctions in characteristics between languages, EMG speech recognition has been independently developed for a variety of languages, in a way like to the development of traditional ASR, which tends to be language specific. In our literature review we found that eight languages have been investigated: English [7-12], Chinese [13-14], Portuguese [15], Japanese [3, 16-17], Serbian [18], Arabic [4], Korean [19], and Spanish [5]. Nonetheless, the language most frequently explored is English due to widely using around the world. In our study, we investigated EMG speech recognition focusing on the Thai language. It is a tonal language and the meaning of a Thai word is powerfully



dependent on its tone. Hence, we recognized isolated words including their tones, which is a challenge unique to this study in the context of EMG speech recognition.

In our earlier work [20], the optimal sensing electrode positions for EMG data acquisition and the suitability of various features were explored. However, in [20] we used only an artificial neural network (ANN) for classification process because it was the most popular classifier types in this field. However, it has high computational demands. To get good classification performance with less processing time, in this paper, we further investigate the various types of classifier to find near optimal classifier, which have the potential to improve the accuracy of the system and low computational cost.

2. Methodology

In this section, the methodology will be described, i.e. EMG data collection, Pre-processing EMG data, Feature Extraction and classification

2.1. EMG data collection

Figure 1 shows the most proper electrode positions and according with muscle groups that were explored in our earlier work [20, 21]. The EMG signals had high amplitudes in all spoken cases, specifying that the EMG signals collected were related to Thai utterance. A Mobi amplifier (TMS international, the Netherlands) with 1024 Hz sampling rate was employed for EMG amplification and recording. Ag/AgCl surface electrodes (ARBO, Tyco Healthcare Group LP, Germany) were employed in EMG signal capturing.

In this exploration, an experiment was divided into two parts. The first was the part of 11 Thai number words classification. The speaker uttered 11 Thai number words, “zero” through “ten” in Thai, in both audible and silent modes, while the EMG signals from five positions (channels) of speech production muscles were captured. Each subject spoke the words 50 times for each mode, and EMG data were collected from ten subjects (5 males, 5 females, age range 30-50 years with Thai as their mother tongue). Therefore, there were 550 segments of EMG signals for each mode and each subject, and in total 11,000 segments representing a single word (11 words x 50 records x 2 modes x 10 subjects).

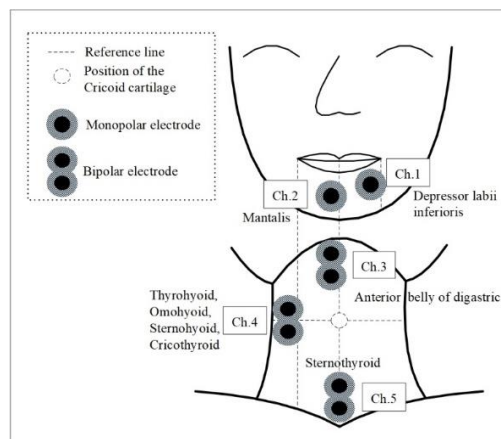


Figure 1. Electrode positions for EMG signal acquisition

21 Thai words (only mid-tone)		
Middle-class Consonants	High-class Consonants	Low-class Consonants
IPA/Thai/Pronunciation	IPA/Thai/Pronunciation	IPA/Thai/Pronunciation
kā: / กา / gah	phā: / เพอ / pur	ṇū: / งู / ngoo
cō: / โจ / jo	fā: / ฟา / far	jā: / ยา / yah
dū: / ดู / doo	thā: / ทอ / tau	nī: / นี / nee
bē: / เบอ / bay	khū: / ขู / koo	rūa: / เรือ / reua
ā: / ออ / aw	sī: / ซี / see	wūa: / วัว / wooah
tī: / ตี / tee	hō: / โฮ / hoh	mū: / มื้อ / meu
pā: / เปอ / pur	chē: / แชร / chae	lā: / ลอ / lau

Figure 2. The training words for 5 Thai tones classification. This figure shows 21 Thai words only a mid-tone. They are separated into three classes, middle class, high class and low class consonants.

Table 1. List of 22 EMG features used in evaluation process

Feature name		Feature name	
F1	Root mean square	F12	Median frequency
F2	Variance	F13	Mean frequency
F3	Waveform length	F14	The 1 st order of spectral moment
F4	Willison amplitude	F15	The 2 nd order of spectral moment
F5	Zero crossing	F16	The 3 rd order of spectral moment
F6	Slope sign change	F17	The 5 th order of spectral moment
F7	Mean	F18	Mean of the low part frequency signal
F8	Integrate absolute value	F19	Mean of the high part frequency signal
F9	Mean absolute value	F20	Power of the low part frequency signal
F10	The 1 st type of modified mean absolute value	F21	Power of the high part frequency signal
F12	The 2 nd type of modified mean absolute value	F22	Zero crossing of the high part frequency signal

Table 2. List of EMG features used in classification process.

Type	mode	Feature used
11 words classification	Audible	F2, F15, F10, F3, F9, F12, F4, F5, F6
	Silent	F8, F10, F1, F3, F12, F14, F5, F4, F6
5 tones classification	Audible	F12, F9_3 ^a , F10_3 ^a , F5_3 ^a , F1_3 ^a , F5, F10, F5_4 ^b , F4_4 ^b

^a “_3” indicates that the 3rd part of the segmented EMG signal was calculated.

^b “_4” indicates that the 4th part of the segmented EMG signal was calculated.

The second was the part of 5 Thai tones classification, the subjects uttered 21 Thai isolated words, shown in figure 2, including five tones for each word in audible mode only. Consequently, the number of words was 105. Each subject spoke 105 words 3 times for only audible mode, and EMG signals were collected from ten subjects same as the first part. Therefore, there were 315 segments of EMG signals for each subject, and in total 3150 segments representing a single tone (105 words x 3 records x 1 modes x 10 subjects). The EMG signals from only the fourth and the fifth positions (Ch. 4 and Ch. 5 in figure 1) were employed to classify the five Thai tones.

2.2. Pre-processing EMG data

Before other processes were performed, the EMG signal had to be cleaned by eliminating noise, such as the motion artifacts, the 50-Hz power-line, and electrocardiography (ECG) noise. Three filters were employed for noise elimination. Firstly, a bandpass filter with cut off frequencies of 20-350 Hz was employed to eliminate the motion artifacts and the high frequency harmonics in the EMG signals from Ch.1-Ch.3. Secondly, a bandpass filter with cut off frequencies of 30-350 Hz [22] was employed to lower the ECG noise in Ch.4 and Ch.5. Both bandpass filters were invented based on a 5th order Butterworth filter which is simple to implement. Furthermore, as was shown in [23], this type of filter efficiently removes ECG noise from EMG recordings. Thirdly, a notch filter was used for all channels to remove the 50-Hz power-line interference.

2.3. Feature evaluation and extraction

Table 1 shows 22 EMG features widely used in the EMG speech recognition field [6, 24, 25] were employed in feature evaluation. A label of these features was defined as “F1-F22”. RES index [24-25] was used in this process to select the best EMG features. The selected features are shown in Table 2. They were grouped into three sets, the first two sets used for 11 Thai number words classification (audible and silent modes respectively) and the last set used for 5 tones classification. The label of

EMG feature specified “_” in the end of it was the feature that was calculated from the segmented EMG signal. In other words, EMG recording of each word was divided into 6 parts, and only some part of EMG was used. For example, the feature “F9_3” means the calculation of the mean absolute value (F9) feature using the third part of EMG recording. If the underscore symbol (“_”) is not specified, it means all parts of EMG recording were calculated. This technique was used only for 5 tones classification.

2.4. Classification

According to a review of related literature, the Artificial neural network (ANN) and the Hidden Markov Model (HMM) are the most attractive classifier types because they give good performance in both ASR and EMG speech classification researches. However, they both have high computational demands. To get good classification performance with less processing time, we therefore explored and compared the performance of 8 types of classifier, including ANN which was used in our earlier work [20]. The total classifiers employed in this paper are Nearest Mean (NM), K-Nearest Neighbor (KNN), Linear Bayes Normal (LBN), Logistic Linear (LOGL), Quadratic Bays Normal (QBN), Fisher’s Least Square Linear Discriminant (FLDA), Support Vector Machine (SVM) and Artificial Neural network (ANN). To evaluate the performance of the classifiers, we used ten-fold cross-validation. Datasets of each subject were separated into 10 parts, nine of which were used for training and the other for testing. This step was repeated ten times and the average accuracy across all 10 trials was the final performance.

3. Results and Discussions

Figure 3 shows the average accuracies across all 10 subjects and words in audible mode for the part of 11 Thai number words classification when the various types of classifier were employed. The FLDA classifier gave the best accuracies (90.01%) which it was higher than ANN classifier about 2.19%. However, the classifier that gave the lowest accuracies is NM classifier (48.21%). Moreover, figure 4 shows the average accuracies across all 10 subjects and words, but for words in the silent mode. The result shows that the best accuracy was obtained when LBN classifier was employed (79.18%). It is higher ANN classifier about 1.95%. The NM classifier gave the lowest accuracies same as the audible mode (31.52%). Moreover, figure 5 shows the classification result for the part of 5 tones classification. When the various types of classifier were employed, the result shows that LOGL classifier gave the best accuracies (68.36%). LOGL classifier had good performance close to ANN classifier. The NM classifier gave the lowest accuracies same as the part of 11 number words classification (40.09%).

Figure 6 shows a computational time for each classifier. The result indicated that SVM and ANN classifier much spent computational time more than other classifiers (more than 100 second). For other classifiers, they take very little time to calculate (less than 10 second).

These results shows that the top three classifiers of each experiment had a slightly different performance. However, when the computational time was considered, shown in figure 6, the ANN classifier is a computationally demanding and slow approach. For this reason, FLDA, LBN and LOGL classifier would be more attractive than ANN classifier.

4. Conclusion

This paper has compared the performance of eight classifiers for 11 Thai number words and 5 Thai tones classification based on EMG signal. The best performance for the part of 11 Thai number words classification was obtained when FLDA (90.01%) and LBN (79.18%) was employed for audible and silent modes, respectively. Also, LOGL gave the best performance (68.36%) for the part of 5 Thai tones classification. ANN classifier had good performance close to these classifiers, however, it is a computationally demanding. For this reason, it is not attractive for our work. However, there were three best classifiers for three parts of classification. To reduce the complexity of system, it should have only one best classifier, which it is the future work explored.

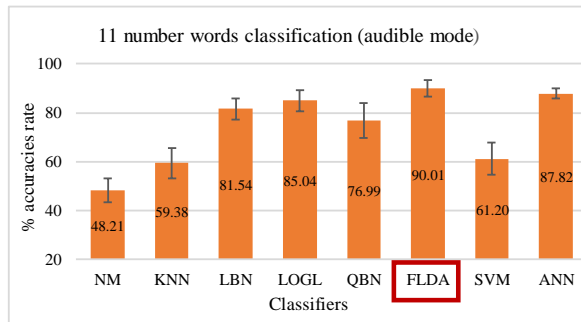


Figure 3. The average accuracy of 11 number words classification in audible mode, when the various types of classifier were employed.

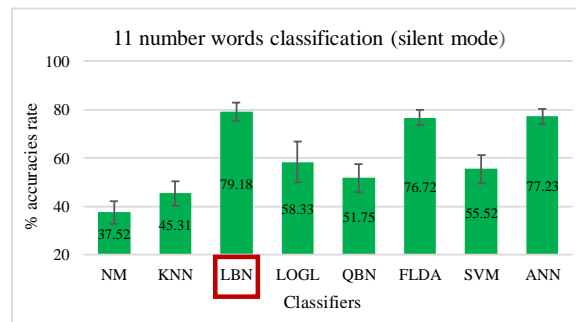


Figure 4. The average accuracy of 11 number words classification in silent mode, when the various types of classifier were employed.

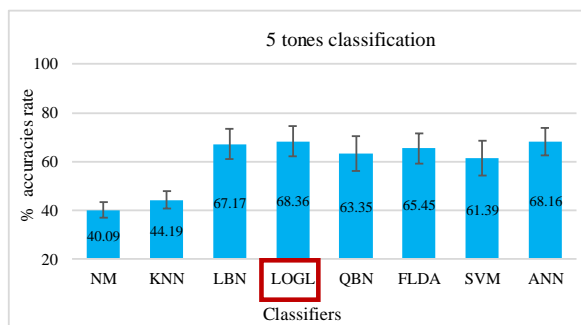


Figure 5. The average accuracy of 5 tones classification in audible mode, when the various types of classifier were employed.

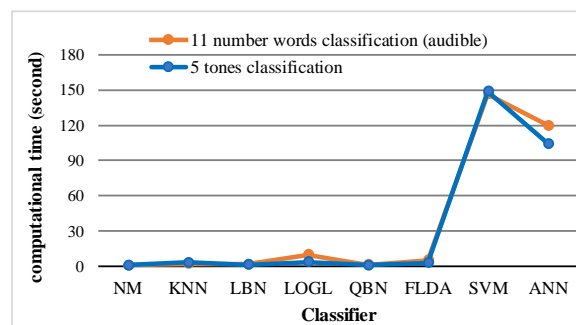


Figure 6. Comparison of computational time when the various types of classifier were employed.

Acknowledgments

This work was supported in part by the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0164/2552).

References

- [1] Sugie N and Tsunoda K 1985 A speech prosthesis employing a speech synthesizer vowel discrimination from Perioral muscle activities and vowel production. *IEEE Trans. Biomed. Eng.* **32** 485-90
- [2] Chan AD, Englehart KB, Hudgins B and Lovely DF 2006 Multiexpert automatic speech recognition using acoustic and myoelectric signals *IEEE Trans. Biomed. Eng.* **53** 676-85
- [3] Kubo T, Yoshida M, Hattori T and Ikeda K 2014 Towards excluding redundancy in electrode grid for automatic speech recognition based on surface EMG *Neurocomputing* **134** 15-9
- [4] Fraiwan L, Lweesy K, Al-Nemrawi A, Addabass S and Saifan R 2011 Voiceless Arabic vowels recognition using facial EMG *Med Biol Eng Comput* **49** 811-8
- [5] Lopez-Larraz E, Mozos OM, Antelis JM and Minguez J 2010 Syllable-based speech recognition using EMG *Proc. Inter. Conf. on IEEE Engineering in Medicine and Biology Society (EMBC)*
- [6] Jou SCS, Schultz T, Walliczek M, Kraft F and Waibel A 2006 Towards continuous speech recognition using surface electromyography. *Proc. Interspeech*
- [7] Deng Y, Colby G, Heaton JT and Meltzner GS 2012 Signal processing advances for the MUTE sEMG-based silent speech recognition system *Proc. Inter. Conf. on Military Communications (MILCOM)*

- [8] Wand M, Janke M and Schultz T 2014 Tackling speaking mode varieties in EMG-based speech recognition *IEEE Trans. Biomed. Eng.* **61** 2515-26
- [9] Betts BJ and Jorgensen C 2005 Small vocabulary recognition using surface electromyography in an acoustically harsh environment NASA TM-2005-21347
- [10] Zhou Q, Jiang N, Englehart K and Hudgins B 2009 Improved phoneme-based myoelectric speech recognition *IEEE Trans. Biomed. Eng.* **56** 2016-23
- [11] Arjunan SP, Kumar DK, Yau WC and Weghorn H 2006 Unspoken vowel recognition using facial electromyogram *Proc. Inter. Conf. on IEEE Engineering in Medicine and Biology Society*
- [12] Morse MS, Day SH, Trull B and Morse H 1989 Use of myoelectric signals to recognize speech. *Proc. Inter. Conf. on IEEE Engineering in Medicine and Biology Society*
- [13] Jia X, Wang X, Li J, Yang D and Song Y 2006 Unvoiced Chinese digital recognition based on facial myoelectric signal *Proc. Inter. Conf. on Communications, Circuits and Systems*
- [14] Lyu M, Xiong C and Zhang Q 2014 Electromyography (EMG)-based Chinese voice command recognition. *Proc. Inter. Conf. on Information and Automation (ICIA)*
- [15] Freitas J, Teixeira A and Dias MS 2012 Towards a silent speech interface for Portuguese *Proc. Biosignals*
- [16] Manabe H and Zhang Z 2004 Multi-stream HMM for EMG-based speech recognition *Proc. Inter. Conf. on IEEE Engineering in Medicine and Biology Society*
- [17] Bu N, Tsuji T, Arita J and Ohga M 2005 Phoneme classification for speech synthesizer using differential EMG signals between muscles *Proc. Inter. Conf. on IEEE Engineering in Medicine and Biology*
- [18] Topalović M, Damnjanović Đ, Peulić A, Blagojević M and Filipović N 2015 Syllable-based speech recognition using Electromyography and decision set classifier *Biomed. Eng. Appl. basis Commun.* **27**
- [19] Lee KS 2008 EMG-based speech recognition using hidden Markov models with global control variables *IEEE Trans. Biomed. Eng.* **55** 930-40
- [20] Srisuwan N, Phukpattaranont P and Limsakul L 2013 Three steps of Neural Network classification for EMG-based Thai tones speech recognition. *Proc. Inter. Conf. on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON)*
- [21] Srisuwan N, Phukpattaranont P and Limsakul L 2011 Feature selection for Thai tone classification based on surface EMG *Proc. I-SEEC2011*
- [22] Redfern MS, Hughes RE and Chaffin DB 1993 High-pass filtering to remove electrocardiographic interference from torso EMG recordings. *Clin Biomech.* **8** 44-8
- [23] Drake JD and Callaghan JP 2006 Elimination of electrocardiogram contamination from electromyogram signals: an evaluation of currently used removal techniques *J. Electromyogr Kinesiol* **16** 175-87
- [24] Phinyomark A and Phukpattaranont P 2009 A novel feature extraction for robust EMG pattern recognition *J Comput* **1** 71-80
- [25] Phinyomark A, Phukpattaranont P and Limsakul C 2012 Feature reduction and selection for EMG signal classification *Expert Syst Appl* **39** 742