A METHOD FOR THAI ISOLATED WORD RECOGNITION USING ANT COLONY ALGORITHM

SARITCHAI PREDAWAN[1]
PRASIT JIYAPANICHKUL[2] *and CHOM KIMPAN*[3]

Faculty of Information Technology
Rangsit University
Muang-Ake, Paholyotin Road, Patumtani, 12000
THAILAND

CHAI WUTIWIWATCHAI[4]

National Electronics and Computer Technology Center
Ministry of Science and Technology
112 Thailand Science Park, Phahon Yothin Rd.,
Klong 1, Klong Luang, Pathumthani 12120
THAILAND

Abstract : -This paper presents Thai isolated word recognition system, which is based on the Ant-Miner algorithm. The system employs fundamental frequency (F0) extraction from input speech signal, analysis of F0 contour for feature extraction, and classification of each tone using the extracted features. In the F0 feature extraction, the polynomial regression functions are employed to fit the segmented F0 curve where its coefficients are used as a feature vector. All attributes are used for construct the classification rules by an Ant-Miner algorithm in order to classify 10 Thai digits utterances. For this experiment, the Ant-Miner algorithm is adapted, with a small change to increase the recognition rate. The result of this experiment is a 95% recognition rate.

Key-words :- Thai Tone, Fundamental Frequency (F0), F0 Feature Extraction, Ant Colony Optimization (ACO), Speech Recognition

1 INTRODUCTION

Intonations in Thai are used to distinguish the word's meanings[1]. Every Thai single-syllable word has its own intonation categorized into five groups, which are low("ake"), medium("saman"), falling("toe"), high("dtee") and rising("jattawa"). Research in Thai tone extraction presented by Ramalingam [4] is compared with our work. Besides, several Thai speech recognition approaches that have been elaborated for years such as in [2] and [5] are studied. The correctness values of these approaches are 74.9% [2] and 89.4% [5].

The feature of speech that was used to classify the tone is the shape of fundamental frequency (F0) contour, which shown in Figure 1. There are several parameters that also have the effect on the shape of F0 contour such as the gender and the age of speaker [8], the initial consonant, the final consonant and the duration of vowel (short or long). In our process, the F0 contour of input speech was automatically smoothed and segmented by the proposed algorithm in section 3.1. Then they were fit by the polynomial regression function, which we used its coefficients as the features of F0 contours. In the recognition process, we used Ant Colony Optimization (ACO) to classify the tones.

This paper is organized as follows. First, the overview of the speech recognition system is introduced. Second, the phonetics of Thai language. Third, F0 feature extraction, Ant Colony optimization and the proposed system of this experiment. Fourth, the experimental result are explained. The conclusion and discussion is given in the section 5.

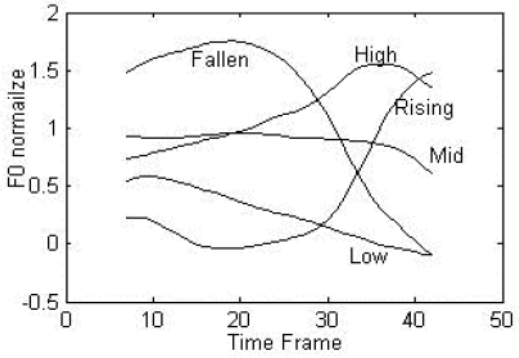

Figure 1 F0 normalize contour patterns of five Thai tones

2 PHONETICS OF THAI LANGUAGE

A Thai sentence consists of various words composed from one or more syllables. The smallest part of each syllable is called a phoneme. Every Thai syllable is organized from two up to six phonemes. Luksaneeyanawin [3] has classified Thai phonemes into two groups i.e. segmental phonemes, consisting of consonants and vowels, and supra segmental phonemes, for example tone and accent phonemes. In our study we divide a syllable into phonemes based on their time occurrences and characteristics. These phonemes are 1) the initial consonant, 2) the vowel, 3) the secondary vowel, 4) the syllable ending, and 5) the tonal. It can be concluded that the tonal is the most important phoneme embedding in every Thai syllable [3]. Our research also investigated that the tones always locate on the vowels and the resonant syllable ending phonemes. The tones always locate on the vowels and the resonant syllable ending phonemes [1]. In Fig. 2, the illustration of consonants and vowels of the a Thai numeral syllable "ku:a:w" (nine in English) uttered by six speakers are presented.

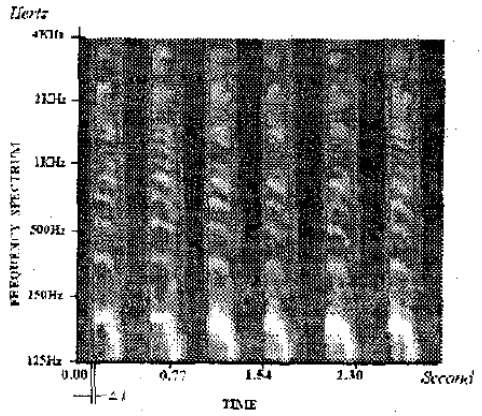

Figure. 2 The spectrograms of a Thai numeral syllable uttered by six speakers

In Fig. 2, The consonant and vowel area can be clearly distinguished. The consonant part of each spectrogram is located on the bottom of spectrogram and in the range of frequencies starting from 250 Hz are vowel areas that are represented in the form frequency resonance.

The data used in this experiment were only 10 Thai digits. All of these 10 Thai digits can occur in the syllable initial consonants, vowels, final consonants and tone as shown in Table 1.

Digit	Phoneme detail			
	Icons	Vowel	Fcons	Tone
0	/s/	/uu/	/n/	Rising
1	/n/	/ɑ/	/N/	Low
2	/s/	/@@/	/N/	Rising
3	/s/	/aa/	/m/	Rising
4	/s/	/ii/		Rising
5	/h/	/aa/		Falling
6	/h/	/o/	/k/	Low
7	/c/	/e/	/t/	Low
8	/p/	/ae/	/d/	Low
9	/k/	/ua/	/w/	Falling

Table 1 Thai digits used in this research [11]

3 F0 FEATURE EXTRACTION

The F0 feature extraction process has two procedures. The first is F0 smoothing and segmentation procedure. The second is polynomial curve fitting procedure.

3.1. F0 Smoothing And Segmentation Procedure

F0 from the F0 extraction process will be smoothed in the smoothing procedure by using median filtering. In the segmentation procedure, there is algorithm that was used to segment the smoothed F0. This algorithm will determine the beginning and the ending frame of the longest time that F0 at each frame has the value differ from the neighboring frame no more than ΔF_{max} =17 Hz.

3.2. Polynomial Regression

The objective of this procedure is to determine the coefficients b_k of a polynomial that fits the segmented F0 contour.

Let

$F = (F1, F2,...,FL)^T$

be a sequence of segmented F0 of length L ,

$F\hat{} (F\hat{}1, F\hat{}2,...,F\hat{}L)^T$

be an estimated vector of F

A d-dimension feature vector

$\beta = (\beta 0, \beta 1,..., \beta d-1)^T$ is the coefficient of (d-1)-order polynomial regression function

$$\hat{F}_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + ... + \beta_{d-1} t_i^{d-1} \quad (1)$$

where $t_i = i/L$ is a normalized time respect to F_i . Equation (1)

can be expressed in matrix form as

$$\hat{\mathbf{F}} = \mathbf{T}\boldsymbol{\beta}, \quad \begin{bmatrix} \hat{F}_1 \\ \hat{F}_2 \\ \vdots \\ \hat{F}_L \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{d-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{d-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_L & t_L^2 & \cdots & t_L^{d-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{d-1} \end{bmatrix} \quad (2)$$

A solution for in a least-squares sense (minimize the Euclidean distance between vectors F and F^) is obtained via forming the pseudoinverse of T, that is,

$$\boldsymbol{\beta} = (T^T T)^{-1} T^T F \quad (3)$$

However, if $T^T T$ is nearly singular, the numerical errors incurred in forming $T^T T$, and then forming the inverse, spawn a need for alternate approaches that are not plagued by numerical sensitivities.

3.3 Ant Colony Optimization

The ACO algorithm [6] is an essential system based on some agents that simulate the natural behaviors of ants. The natural behaviors are include mechanisms of cooperation and adaptation (Natural behavior of ants is shown in Fig. 3). It is based on the following ideas.

1) Each path is followed by an ant which is associated with a candidate solution for a given problem.

2)When an ant follows a path, the amount of pheromone deposited on that path is proportional to the quality of the corresponding candidate solution for the target problem.

3) When an ant has to choose between two or more paths, the first priority path is the path, which has a larger amount of pheromone.

The ACO is used to solve various kinds of trace route and congestion problems. As a result, the ant eventually converge on the shorts path (optimum or a nearest-optimum solution).

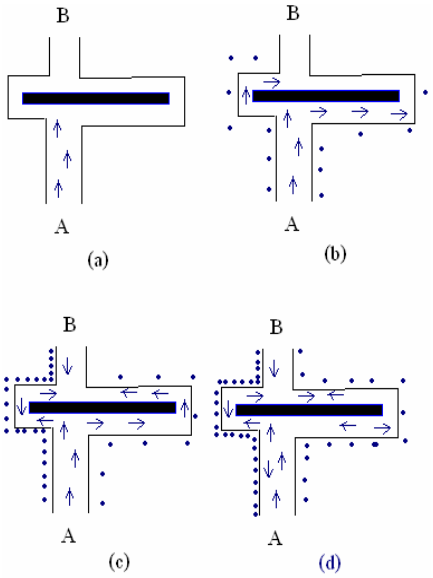

Figure 3 How real ants find a shortest path from A to B. (a) Ants arrive at a decision point. (b) Some ants choose the left path and others choose the right path. The choice is random. (c) The ants which choose the shorter path (left) arrive at the opposite decision point faster than longer path (right). (d) Pheromone accumulates at a higher rate on the shorter path. The number of points is approximately proportional to the amount of pheromone deposited by ants.

3.3.1 Ant-Miner algorithm

The Ant-Miner algorithm [7,10] has been proposed to the set of IF-THEN rule in the form of IF < term1 and term2 and … > and THEN < Class > in the work of data mining. Each term in the rule of antecedent part is a triple < attributes, operator, value >. Value is the possible value in a domain of each attribute. There is only an operator "=" used in this work such as < Month = January >. The consequent part specifies the class prediction in case that predicted attributes satisfy all the terms in the antecedent part. The set of rules, constructed by this algorithm cover all or almost all the raining cases. These rules have a small number of terms.

A small number of rules that is good for data mining. The high level of its algorithm is shown in Fig. 4.

Training set = all training cases;
Rule list = empty;
REPEAT
  i=0;
  Pheromone Initialization;
  REPEAT
    i=*i*+1;
    Ant_i constructs a classification rule;
    Prune the current constructed rule;
    Update the pheromone of the trail followed by Ant_i;
    The best rule is memoried;
  UNTIL (*i* = *No_of_Ants*) or (Ant_i constructed the same rule as the previous Ants continually *No_Rule_Converg* times)
  The best rule is added to the rule list;
  Remove the cases covered by the selected rule from the training set;
UNTIL (Number of cases in the Training set < *Max_uncovered_cases*)

Figure 4 High level of the Ant-Miner algorithm

Following the algorithm; after the pheromone is initialized, many rules are constructed in the inner Repeat-Until loop with the rule pruning and the pheromone updating method. The loop will stop when ants construct the same rule continually more than *No_Rule_Converg* times or the number of rules are equal to the number of ants. When the inner Repeat-Until loop is completed, the best rule will be added to the rule list. Then, all training cases predicted by this rule were removed from the training case set. Pheromone is initialized again. This cycle is controled by the Outer Repeat-Until loop.The Repeat-Until loop will be finished when the number of uncovered training cases is less than a threshold, called *Max_uncovered_cases*.

3.4 Propose Model

Over all of the implementation system is shown in Fig. 5 There are four essential steps in recognition process; speech signal processing, feature extraction, recognition engine and recognition result.

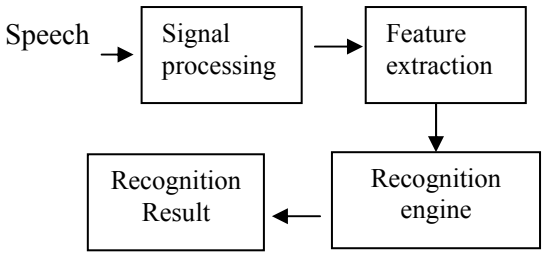

Figure 5 Overview of system implementation

The first block is preprocessing is a process of preparing speech signal for further processing. This process composes of signal pre-emphasis, smoothing window. The speech signal is pre-emphasized by first-order digital filter. The pre-emphasized signal is block into frame by a hamming windows to reduce the amplitude at the edge as shown in Fig. 6. The second block is the Feature extraction process, which determines the parameters that have sufficient information to describe the shape of F0 contour by the method of polynomial regression. The final block is the tone recognition algorithm that uses the parameters obtained from the previous process to determine the best matching tone for the input speech.

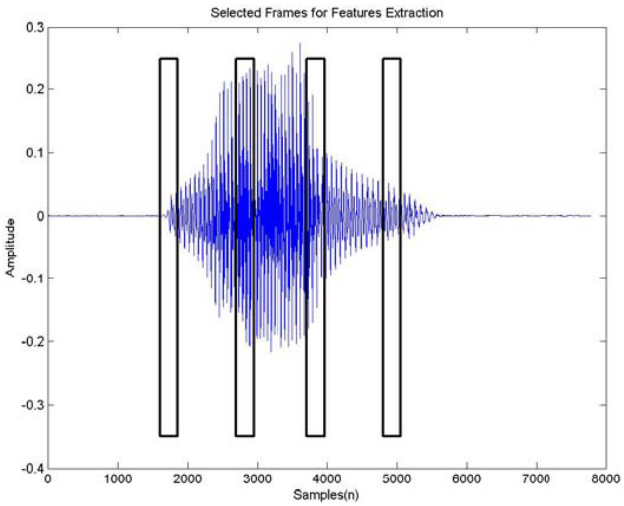

Figure 6 Selected frames for feature extraction

4 EXPERIMENT AND RESULT OF RECOGNITION

To evaluate the performance of the proposed speech recognition system, the speech material used in the experiment was a Thai isolated digit database produced by 30 speakers (15 males, 15 females, 900 wave files), within the range of 20-35 years old. Each speaker pronounced 0-9 digits 3 times. The speech utterances were recorded in a quite room. The recorded speech is 8-bits and 8 kHz sampling rate. The utterances from 20 speakers (10 males, 10 females, 600 wave files) were used as training data, and 300 wave files were use as test set.

Fig. 7 illustrates the stages in the automatic speech recognition system. Firstly, the end-point detection routine is applied to the raw data to find the region of interest for further processing (Fig. 7b). This is followed by the selection of frames based on the starting and ending point (Fig. 7c). The selected frames are then represented by F0 feature extraction

(Fig. 7d). Finally, the features of all frames are fed into the recognition system.

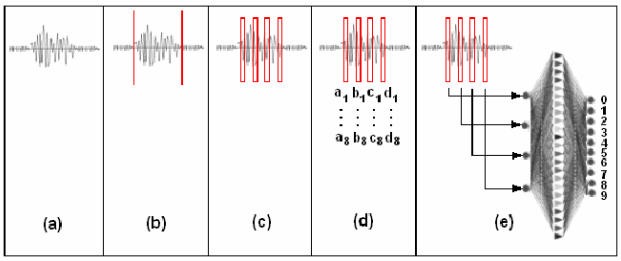

Figure 7 Stages in the Automatic Speech Recognition System (a) Raw data, (b) end-point detection, (c) frame selection, (d) feature representation, (e) feeding inputs to recognition system.

The same procedures are applied in the training and recognition stages.

A. Data

There are 10 Thai digits wave files from 30 speakers. Each speaker pronounced 0-9 digits 3 times. The total utterances are 900 wave files.

B. Pre-processing

In this step. A single syllable is its input. F0 was computed in this process from 256 samples speech frame with the overlapping of ¾ frame by using modified short-term autocorrelation with center clipping method.

C. Feature Extraction

The F0 feature extraction process, the features of each utterance are extracted which determines the parameters that have sufficient information to describe the shape of F0 contour by the method of polynomial regression. An example, which features extracted of "0" utterance has initial consonant /s/, vowel /uu/, final consonant /n/ and rising tone, Each frame consists of 256 sample data.

D. Recognition Process

In the this step,"0" which has 41 attributes for each data are classified by the Ant-Miner algorithm. The rule list from the algorithm is used as the recognition engine.

4.1 Recognition Engine and recognition Rule

In this experiment, the Ant-Miner algorithm is used for training the recognition system (to construct a rule list). The utilizing data was made from feature of a speech utterance described in section 3.

The original version of the Ant-Miner [7,10], the quality of the rule is computed by the equation (4)

$$Q(Rule) = \frac{TP}{TP+FN} x \frac{TN}{FP+TN} \qquad (4)$$

Where :

a) TruePos (TP) is the number of cases covered by the rule and having the same class as that predicted by the rule.
b) FalsePos (FP) is the number of cases covered by the rule and having the difference class as that predicted by the rule.
c) FalseNeg (FN) is the number of cases that are not covered by the rule and having the class as that predicted by the rule.
d) TrueNeg (TN) is the number of cases that are not covered by the rule which have a difference class from the class predicted by the rule.

Which the equation, the rules from the algorithm are short. It has a minimum number of rules in the rules list with a high accuracy. (Cover all or almost all in the training set). In this experiment, recognition all of the training data is needed. The equation of Q (Rule) is changed to (5)

$$Q(Rule) = \left(\frac{TP}{FP+1} \right) \qquad (5)$$

By the Q (Rule), the value of FP is converged to zero. This means that accuracy rate when testing, is converged to most data. The number of rules will be more than the original. The other parameters are No_of_Ant, No_Rule_Converg and Max_Uncovered_Cased. In the training step, the data of 600 wave files into 10 classes (all of digits), which has 60 data in each class and 41 attributes for each data. Finally data of each group are classified by the Ant-Miner algorithm. It was described in section 4. The rule list from the algorithm is used as the recognition engine.

Digit	1	2	3	4	5	6	7	8	9	0	Mean
No_Ant	Ant-Miner algorithm trained with 10 sets * data from male speaker and 10 sets * data from female speakers										
1000	99.1	91.2	97.3	99.1	96.5	94.7	93.8	92.9	91.2	94.7	95.0
1500	96.5	93.8	96.5	98.2	95.6	96.5	97.3	94.7	93.8	90.3	95.3
2000	99.1	94.7	98.2	99.1	96.5	93.8	95.6	95.6	95.6	92.9	96.1
2500	100.0	94.7	100.0	97.3	96.5	94.7	98.2	94.7	92.9	94.7	96.4

Table 2 Recognition Rate at Different number of Ants

The training set of 600 wave files are input in the Ant-Miner algorithm for learning. In order to classify each group of ten Thai digits by construction of a set of rules. The result is show in Table 2 and averages are calculated by Numbers of Ant in each group. Table 3 compare the recognition rate of the proposed system with the KLT+LVQ [9].

Recognition Engine	Recognition rate
Proposed system	95%
KLT+LVQ [9]	90%

Table 3 Recognition rate of proposed system and KLT+LVQ[9].

5 CONCLUSION

In this paper, the approach of using Ant-Miner algorithm for Thai isolated word recognition has been studied. We used the coefficient of polynomial regression function as a feature vector of the segment F0 contour. In the training phase, the feature vector was used to determine the statistical parameters of the model for each class. In the testing phase, the feature vector will be passed to the automatic recognition process. The result shows that the system can recognize 95% of the testing set. The recognition rate of the propose system is more than the KLT+LVQ [9]. There are many advantages of the proposed system. First, when there are some feature parts which are incorrect or missing, it will use the other appropriate features. It is the solution for the problem[7]. The second is giving a higher recognition rate. We believe that this method can be applied to other types of recognition.

REFERENCES

[1] R.Kongkachandra, S.Pansong, T.Sripramong and C.Kimpan. : "Thai Intonation Analysis in Harmonic-Frequency Domain," The 1998 IEEE APCCAS. Proceeding, (1998) pp. 165-168. 1998.

[2] E.Maneenoi, et al.: "Modification of BP Algorithm for Thai Speech Recognition," NLPRS'97 (Incorporating SNLP'97) Proceeding (1997), pp. 287-291

[3] Luksaneeyanawin, S.: "Intonation in Thai," Ph.D. Thesis, University of Edinburgh, 1983

[4] Ramalingam, H.: "Extraction of Tones of Speech: An application to the Thai Language," Master Thesis (TC-95-5), Asian Institute of Technology, Thailand, 1995

[5] W.Pornsukjantra, et al: "Speaker Independent Thai Numeral Speech Recognition Using LPC and the Back Propagation Neural Network," NLPRS'97 (Incorporating SNLP'97) Proceeding, pp. 585-588. 1997.

[6] Marco Dorigo, Thomas Stutzle. "Ant colony optimization," A Bradford Book The MIT Press, Cambridge, Massachusetts, London, England, 2004.

[7] S. Airphaiboon. "Recognition of Hand-written Thai character considering the head of character," Masters Thesis, Department of Electrical Engineering, King Monkut's Institute of Technology Ladkraband, Bangkok, Thailand, 1998.

[8] P. Charnvivit, S. Jitapunkul et al., "F0 Feature Extraction by Polynomial Regression Funcition for Monosyllabic Thai Tone Recognition," Eurospeech 2001, Scandinavia.

[9] S. Predawan, P. Jiyapanichku. C, Kimpan and C. Wutiwiwatchai. "Tone Analysis in Harmonic-Frequency Domain and Reduction using KLT+LVQ for Thai Isolated Word Recognition," 2006 WSEAS International Conference, May 2006.

[10] P.Phokharatkul, K.Sankhuangaw, S.Phaiboon S.Somkuarnpanit, and C.Kimpan "Off-Line Hand Written Thai Character Recognition Using Ant-Miner Algorithm," Transactions on ENFORMATIKA on Systems Sciences and Engineering, vol. 8, pp. 276-281, October 2005.

[11] A. Deemagarn and A. Kawtrakul. "Thai Connected Digit Speech Recognition Using Hidden Markov Models," SPECOM'2004 : 9th Conferenece Speech and Computer, September 2004.