A COMPARISON OF THAI SPEECH RECOGNITION SYSTEMS USING HIDDEN MARKOV MODEL, NEURAL NETWORK, AND FUZZY-NEURAL NETWORK

Visarut Ahkuputra, Somchai Jitapunkul, Nutthacha Jittiwarangkul,
Ekkarit Maneenoi, and Sawit Kasuriya

Digital Signal Processing Research Laboratory, Department of Electrical Engineering,
Faculty of Engineering, Chulalongkorn University, THAILAND
e-mail - jsomchai@chula.ac.th

ABSTRACT

The recognition of ten Thai isolated numerals from zero to nine and 60 Thai polysyllabic words are compared between different recognition techniques, namely, Neural Network, Modified Backpropagation Neural Network, Fuzzy-Neural Network, and Hidden Markov Model. The 15-state left-to-right discrete hidden markov model in cooperation with the vector quantization technique has been studied and compared with the multilayer perceptron neural network using the error backpropagation, the modified backpropagation, and also with the fuzzy-neural network with the same configuration. The recognition error on Thai isolated numerals using the conventional neural network, the modified neural network, the fuzzy-neural network, and the hidden markov model techniques are 26.97 percent, 22.00 percent, 8.50 percent, and 15.75 percent respectively.

1. INTRODUCTION

Various speech recognition techniques have been proposed and applied to English language. The hidden markov model (HMM) and the neural network (NN) are among the most popular approaches that have been implemented. The fuzzy techniques in cooperation with the neural network have also been applied on classification scheme. The implementation of these approaches to Thai speech recognition has been studied and compared. The 70 Thai polysyllabic words comprise single syllable words, double syllables words, and triple syllables words, 20 words in each set, and the other ten Thai numerals from zero to nine. The specific characteristics of Thai numeral utterances have some resemblance between each utterance, for instance, tonal levels and initial consonant sound, that make recognition more complicated than other languages. However, there are many researches on Thai speech recognition approaches, for example, dynamic time warping (DTW) [1],[2], LPC with backpropagation neural network [3],[4], and hidden markov model with vector quantization [6],[7], have been conducted. This paper is organized as follows. In Section 2, the Thai speech recognition approaches are introduced in details. In Section 3, the experimental results are compared and discussed. Finally, the conclusion of the speech recognition approaches with the recognition system configuration as shown in Figure 5 at the end of this correspondence.

2. THAI SPEECH RECOGNITION APPROACHES

The speech recognition approaches applied to Thai language—the hidden markov model, the neural network, the modified backpropagation neural network, and the fuzzy-neural network—are described in details as follows.

2.1. Neural Network

The neural network (NN) employed in Thai numerals recognition is the multilayer perceptron neural network (MLP-NN) using the error backpropagation algorithm in training. [3], [4] The neural network configuration has 330 input nodes with one 70-node or one 90-node hidden layer and 10 output nodes corresponding to 10 Thai numerals. The input speech sequence must be time-normalized to fit the input node of the network. The input sequences compose of 33 feature vectors, each comprises 10 linear prediction coefficients for the total of 330 coefficients for input nodes in the input layer. The neural network configuration is shown in Figure 2.

2.2. Modified Backpropagation Neural Network

The modified error backpropagation neural network has been proposed and applied to the Thai numeral recognition [4]. The classical error backpropagation algorithm has been modified to improve the slowness of convergence and the recognition accuracy of a conventional neural network. The modifications have been made on the slope adaptation of the activation function and the momentum weighting adjustment parameter has been added.

The sigmoid activation function used in the neural network is shown in (1). The slope α of the activation function has been modified to be adaptive to the error during training. The updating scheme of the activation function is shown in (2) and (3) using the gradient descent method. The μ_α is the momentum adjustment parameter for α, then, the momentum adjustment parameter for weight updating has been added as shown in the last term of (4). This will help the updated weight value not to be a local minima value and could lead to the local optimum value.

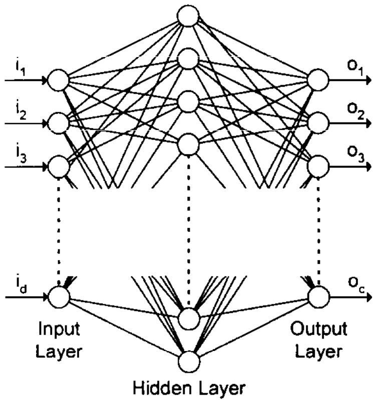

Figure 1: Left-to-Right Hidden Markov Model. [1]

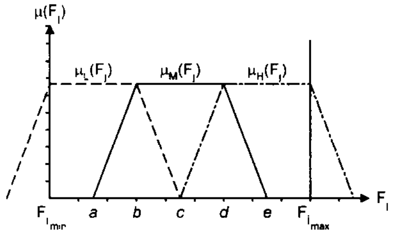

Figure 2: Trapezoidal Fuzzy Membership Function. [5]

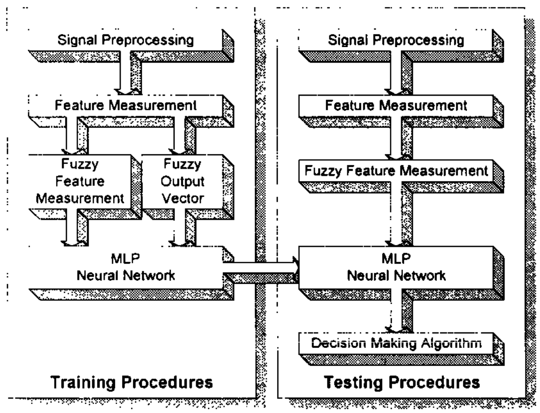

Figure 3: Fuzzy-Neural Network Processing Diagram. [5]

$$f(x)=\frac{1}{1+e^{-ax}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\delta_k\alpha_j^{[h]}=-\frac{\partial e_k}{\partial\alpha_j^{[h]}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$\begin{aligned}\alpha_j^{[h]}(t+1)=\alpha_j^{[h]}(t)+\varepsilon_\alpha\sum_{k=1}^{p}\delta_k\alpha_j^{[h]}+ \\ +\mu_\alpha\left(\alpha_j^{[h]}(t)-\alpha_j^{[h]}(t-1)\right)\end{aligned} \quad \ldots\ldots\ldots (3)$$

$$\begin{aligned}w_{ij}^{[h]}(t+1)=w_{ij}^{[h]}(t)+\sum_{k=1}^{p}\delta_k w_{ij}^{[h]}+ \\ +\mu_w\left(w_{ij}^{[h]}(t)-w_{ij}^{[h]}(t-1)\right)\end{aligned} \quad \ldots\ldots\ldots (4)$$

2.3. Fuzzy-Neural Network

The Fuzzy technique in cooperation with the Neural Network has been applied to recognize the speaker-independent Thai numerals recognition [5] compared to the conventional multilayer perceptron neural network. From the fuzzy set theory, a pattern "r" subset of the universe "R" has been created due to the grade of membership with the membership function $\mu_A(r)$ to a fuzzy set A as shown in (5). The modified overlapping trapezoidal fuzzy membership functions as shown in Figure 3 have been functioned to convert a 10-order LP feature vector into a 30-order fuzzy membership feature vector as shown in (6) and then pass to the neural network inputs. The class membership function has been modified to have the output value within the range of zero to one, [0,1], which indicates the degree of similarity to that class. The process of training and testing using fuzzy-neural network is shown in Figure 4 where the class membership function has been employed in the fuzzy feature measurement and fuzzy output vector.

$$A=\{(r,\mu_A(r))\},\quad r\in R,\quad \mu_A(r)\in[0,1] \ \ldots\ldots\ldots (5)$$

$$\bar{F}=\begin{bmatrix}\mu_L(F_1) & \mu_M(F_1) & \mu_H(F_1)\\ \mu_L(F_2) & \mu_M(F_2) & \mu_H(F_2)\\ & \vdots & \\ \mu_L(F_n) & \mu_M(F_n) & \mu_H(F_n)\end{bmatrix} \ \ldots\ldots\ldots\ldots (6)$$

2.4. Hidden Markov Model

The discrete hidden markov model (DHMM) applied to Thai speech recognition [6] is a 15-state left-to-right model as shown in Figure 4 with the property of state transition probabilities aij as shown in (7). The model configuration has been modified and adapted to accommodate the characteristics of Thai language and the polysyllabic words. The vector quantization algorithm in cooperation with the DHMM has been applied to replace a 10-order LP coefficient feature vector with a single scalar value of 256-vector codebook with minimum distortion. The K-Mean clustering algorithm has been employed in codebook training to create the 256-vector reference codebook for vector quantization using mean-squared error (MSE) distortion measure. The forward-backward procedure and the Baum-Welch reestimation procedure have been employed during training process. On unknown utterances testing, the Viterbi algorithm has been applied.

$$a_{ij}>0,\qquad j>i+2 \ \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

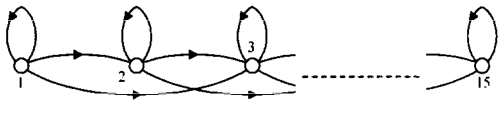

Figure 4: Left-to-Right Hidden Markov Model. [6]

3. EXPERIMENTAL RESULTS

The recognition systems presented in this correspondence are based on the same configuration and speech databases but with different recognition approaches as shown in Figure 5. The 8,400 Thai utterances of 70 polysyllabic words from 60 male and female speakers are recorded twice at 16-bit and 11.025 KHz sampling rate. The speech samples are passed through the signal preprocessing--signal preemphasis, frame blocking, and smoothing window. The signal is preemphasized using the transfer function 1-0.95z-1 and smoothing windowed using 20 ms Hamming window with 5 ms frame shift. The 10-order Linear Prediction (LP) coefficient analysis is applied for speech feature extraction from each speech frame as a feature vector forming feature vector sequence of speech utterance for training and testing.

The recognition results on Thai numerals and polysyllabic word recognition using the hidden markov model, neural network, and fuzzy-neural network are shown in Table 1. The recognition error rate using the hidden markov model, the neural network, the modified neural network, and the fuzzy-neural network are 15.75 %, 26.97 %, 22.00 %, and 8.50 % respectively. From the recognition result, the hidden markov model has the highest accuracy among other techniques--not required to have the signal time-normalized compared to the neural network. In other words, the hidden markov model has no constraint on the length of the input sequences. In numerals recognition, error rate has been reduced by 41.60 % using HMM compared to the conventional neural network. The recognition error rate has been reduced by 18.43 % and 68.48 % using the modified neural network and the fuzzy-neural network respectively compared to the conventional neural network.

The substantial decline on the recognition error rate of the modified neural network over the conventional neural network results from the capability and flexibility of the slope adaptation to errors and the momentum parameters on weight updating in the error backpropagation algorithm during training. The modification on the backpropagation algorithm offers a significant improvement to the neural network.

The fuzzy-neural network has shown a vast improvement in the recognition accuracy over the modified neural network and the conventional neural network. The class membership function output value of the fuzzy-neural network has been modified over the conventional class membership value to be within the range of zero to one. This modification leads to a major improvement in recognition rate over the regular class membership value as well.

4. SUMMARY

This correspondence has presented the comparative review of different recognition approaches on the isolated Thai numerals and the Thai polysyllabic words recognition. The hidden markov model shows promising recognition accuracy over other approaches. Modification on error backpropagation algorithm and the fuzzy embedded into neural network have shown the substantial improvement in recognition rate over the conventional neural network. From these researches and experiments, the application of Thai speech recognition system could be realistic in the near future.

5. ACKNOWLEDGEMENT

The authors would like to acknowledge the Digital Signal Processing Research Laboratory, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University and the Telecommunication Consortium Scholarship, National Science and Technology Development Agancy (NSTDA) on the support of these researches.

6. REFERENCES

1. Pensiri, R. and Jitapunkul, S. "Speaker-Independent Thai Numerical Voice Recognition by using Dynamic Time Warping", *Proceedings of the 18th Electrical Engineering Conference*, 977-981, 1995.

2. Phatrapornnant, T. and Jitapunkul, S. "Speaker-Independent Isolated Thai Spoken Vowel Recognition by using Spectrum Distance Measurement and Dynamic Time Warping", *Proceedings of the 18th Electrical Engineering Conference*, 988-993, 1995.

3. Pornsukchandra, W. and Jitapunkul, S. "Speaker-Independent Thai Numeral Speech Recognition using LPC and the Back Propagation Neural Network", *Proceedings of the 19th Electrical Engineering Conference*, 977-981, 1996.

4. Maneenoi, E., Jitapunkul, S., Wutiwuwatchai, C., and Ahkuputra, V., "Modification of BP Algorithm for Thai Speech Recognition", *Proceedings of the 1997 International Symposium on Natural Language Processing*, 1997.

5. Wutiwiwatchai, C. "Speaker Independent Thai Polysyllabic Word Recognition Using Fuzzy Technique and Neural Network", *Master's thesis*, Chulalongkorn University, 1997.

6. Areepongsa, S. and Jitapunkul, S. "Speaker Independent Thai Numeral Speech Recognition by Hidden Markov Model and Vector Quantization", *Proceedings of the 1997 International Symposium on Natural Language Processing*, 370-378, 1995.

7. Ahkuputra, V., Jitapunkul, S., Pornsukchandra, W. and Luksaneeyanawin, S. "A Speaker-Independent Thai Polysyllabic Word Recognition Using Hidden Markov Model", *Proceedings of the 1997 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 593-599, August 1997.

Recognition Techniques	Recognition Error Rate (%)			
	Numerals	Single Sylalble	Double Syllables	Triple Syllables
Dynamic Time Warping [1]	20.75	-	-	-
Conventional Neural Network [3]	26.97	-	-	-
Modified Backpropagation Neural Network [4]	22.00	-	-	-
Fuzzy-Neural Network [5]	8.50	16.80	12.00	12.50
Discrete Hidden Markov Model [7]	15.75	13.25	7.62	3.75

Table 1: Thai Speech Recognition Results using Different Recognition Techniques

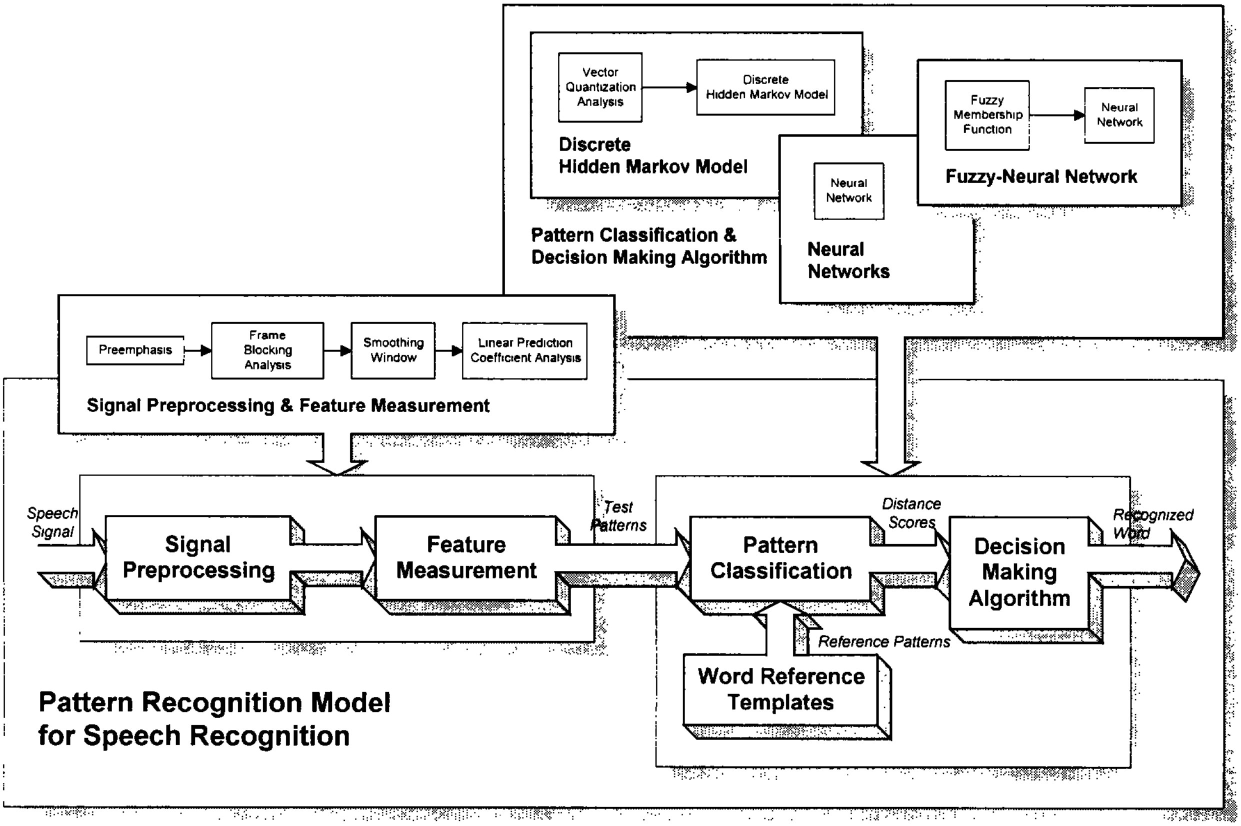
Vector Quantization Analysis
Discrete Hidden Markov Model
Discrete Hidden Markov Model
Fuzzy Membership Function
Neural Network
Fuzzy-Neural Network
Neural Network
Neural Networks
Pattern Classification & Decision Making Algorithm
Preemphasis
Frame Blocking Analysis
Smoothing Window
Linear Prediction Coefficient Analysis
Signal Preprocessing & Feature Measurement
Speech Signal
Signal Preprocessing
Feature Measurement
Test Patterns
Pattern Classification
Distance Scores
Decision Making Algorithm
Recognized Word
Reference Patterns
Word Reference Templates
Pattern Recognition Model for Speech Recognition

Figure 5: The Thai Speech Recognition System on Thai Numerals and Thai Polysyllabic Words