AU J.T. 8(3): 113-123 (Jan. 2005)

A Framework for Connected Speech Recognition for Thai Language*

Pratit Santiprabhob, Jirawat Chaiareerat, Ronnarit Cheirsilp, Nunmanus Dachapratumvan and Wanna Supasirirojana

Faculty of Science and Technology, Assumption University
Bangkok, Thailand

Abstract

Connected speech recognition problem for Thai language, like similar problem in other languages, involves three sub-problems: (i) syllable segmentation, (ii) syllable recognition, and (iii) syllable-based word recognition. This paper presents a framework upon which a speech recognition system can be built. The approach taken in our framework differs from a so-called word-based approach in which whole words are trained to be later recognized. Our approach attempts to recognize syllables based on their constituent phonemes; the recognized syllables are then grouped into words within a given context of discourse. The four constituent phonemes of Thai syllables are leading consonant, vowel, ending consonant and tones. The proposed framework utilizes several soft computing techniques in different parts. As for the signal-processing portion of the framework, Fuzzy System (FS) is used in the syllable segmentation part while the Neural Network (NN) and Hidden Markov Model (HMM) are used in the syllable recognition part. On the other hand, Genetic Algorithm (GA) and rule-based system techniques are used to develop alternative methods to recognizing words from given set of syllables

***Keywords**: Hidden Markov Model, neural network, fuzzy system, genetic algorithm, rule-based system.*

1. Introduction

Speech is a primary means of human communications. It is the most natural way for humans to convey ideas, to exchange information, to give instruction, etc. A speech is an intelligible group of words. Thus, the foundation for the understanding of human speech is the understanding of spoken words, which in turn requires the recognition of spoken words to first be achieved. Our proposed framework outlines methods that can be used to solve this spoken words (or speech) recognition problem. This is indeed an exciting yet challenging research area. Speech is seen as the way humans will interact with computers in the future. In general, humans can speak about two times faster than a proficient typist can type. In addition, this mode of man-machine interaction allows for hand-free operation such as giving on-board computer an instruction while driving a car.

Techniques for recognizing words as trained are widely commercially available. These words are not connected, individual words that can be encoded as templates. On the other hand, recognizing connected speech is a totally different problem with a magnitude of difficulty. Our proposed framework is conceptually depicted in Fig. 1. In the first step, the given speech is segmented into syllables. Then, in the second step, each syllable is attempted a recognition from its constituent phonemes. Eventually, in the third step, the recognized syllables are decoded into words within a given context of discourse.

Various researchers have developed different alternatives to the problem of Thai speech recognition. Different techniques are used such as Dynamic Time Wrapping (Penisiri and Jitapunkul 1995), Conventional Neural Network (Porsukchandra and Jitapunkul

* *This research is supported in part by the Thailand Research Fund.*