Scalable Architecture of Tone Classification Function for Tonal Speech Recognizer

J. Chaiwongsai1 W. Chiracharit1 K. Chamnongthai1 Y. Miyanaga2 K. Higuchi3

1 Department of Electronic and Telecommunication Engineering, Faculty of Engineering, King Mongkut’s University of Technology Thonburi, Bangkok, Thailand
2 Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan
3 Department of Electronic Engineering, University of Electro-Communications, Tokyo, Japan

ABSTRACT

Tone classification function is used for improving recognition accuracy in tonal speech recognizer (TONE-SPEC). Although average magnitude difference function (AMDF) is generally used to find pitch period of fundamental frequency, there are many frame-repeated processes. This paper proposes scalable architecture of tone classification function for tonal speech recognizer. In the proposed architecture, the number of frames is reduced using vowel-AMDF (V-AMDF). Moreover, there is no frame iteration because the architecture converts series computation of conventional tone classification function into parallel. The parallel computation is designed to be able to reduce or extend the number of frame. Our architecture is set and evaluated with 10 Thai words selected from TV remote control commands and the words having the same phoneme but different tones. The experimental results show that the time consuming of general AMDF and series V-AMDF are improved 85.2% and 72.7%, respectively.

1. INTRODUCTION

Nowadays, many researchers developed speech recognizer for hardware implementation to be a small size equipment, low power consumption, real time processing and high recognition accuracy. However, in tonal languages such as Thai, Chinese, Vietnamese, etc., speech recognition needs more consideration. Tonal languages use tone to classify the word meaning. Therefore, tone is a serious problem of tonal speech recognition. Especially, in the case of a group of words which have same phoneme but different tone. The characteristics of tones can be distinguished by the shape of fundamental frequency ($F_0$) [1].

There are mathematical techniques available to classify tones such as multilayer perceptron neural network [2-4], hidden Markov models (HMMs) [5-6] and average magnitude difference function (AMDF) [7]. The nature of AMDF operations is suitable for special purpose hardware. AMDF has lower arithmetic complexity because it uses the number of multiply-operations less than autocorrelation function [8]. CM-AMDF [9] was implemented in frame-synchronous pitch feature extraction. The extraction system was equivalent to real time system. However, the CM-AMDF computes the pitch period by using the whole input speech. Thus, CM-AMDF has to repeat a lot of number of frames. Although V-AMDF [10] reduces processing time by using only vowel signal as an input of the classification function, the classification function is done in frame sequence.

This paper proposes scalable architecture of tone classification function for tonal speech recognizer (TONE-SPEC). The architecture improves AMDF processing time by computing only vowel speech, called V-AMDF. Tone classification function converts series classification function into parallel. We design parallel computation of the classification function to be able to reduce or extend the number of frame, called scalable architecture. Therefore, there is no frame iteration. The proposed architecture includes 4 processes; speech analysis, pitch detection, $F_0$ extraction and tone decision. Firstly, speech analysis process detects vowel signal by software simulation and then only vowel signal are sent through the pitch detection. Therefore, the number of frame in the pitch detection is reduced. Pitch detection process finds the pitch period by using 4-stage pipeline process of V-AMDF including absolute-operation, multiply-operation, and two add operations. Real peak point is determined by the first minimum peak which has lower value than a peak threshold. After that, pitch period is estimated and then $F_0$ is determined in $F_0$ extraction part. Finally, the decision process finds the maximum vote from look-up table and then tone result is classified.

2. OVERVIEW OF TONE CLASSIFICATION FUNCTION

Tone classification function is designed for improving recognition accuracy in tonal language speech recognizer.