A VLSI Architecture of Tone Classification Function-Based Isolated-Word Speech Recognition

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

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

Abstract – Speech recognition in tonal languages such as Thai, Chinese, etc. classifies word meaning by using tone. Therefore tone classification function is extremely essential part for improving accuracy rate. This paper presents a novel VLSI architecture of tone classification function-based isolated word speech recognition. The architecture consists of two parts; feature extraction and tone classification function. In feature extraction part, voice detection, pitch period estimation and slope classification are introduced. The proposed pitch period is calculated by using parallel computation and 3-stage pipeline process. In the classification function, look-up table technique is employed to detect tone by using only \( F_0 \) characteristic information. This takes advantage in reducing the complexity of computation cost of the proposed architecture. Moreover, no training set is used. To evaluate the proposed architecture, the experiment is performed with 100 word vocabularies selected from 20-40 years old dependent-speakers. The architecture is implemented on Altera Cyclone II series FPGAs running at 50 MHz. The results reveal 88.25% accuracy rate and 8.27 ms/word processing time.

Keywords: tone classification function, VLSI implementation, pitch period, parallel computation, pipeline process, look-up table

I. Introduction

Speech recognition plays important role in human-machine interface application such as navigation system, mobile information terminal and disabled person equipment. In tonal language, for example: Chinese, Vietnamese, Thai and so on, use tones to classify word meaning. Each tone shows different purpose. Therefore, improving speech accuracy rate, we have to consider tone classification function efficiency. To achieve the performance of the function in the case of speech recognition equipments, we require recognition accuracy, low power consumption and small size equipments.

Tone classification function can be approached to isolated speech and continuous speech types. Generally, the function is developed by many mathematical techniques such as multilayer perceptron neural network, hidden Markov models and Fujisaki’s models and so on. The techniques are calculated for estimating the fundamental frequency (\( F_0 \)) and/or the \( F_0 \) feature vectors. Multilayer perceptron neural network technique is represented in isolated-tone recognition and also continuous-tone recognition. In isolated-recognition, the technique classified suprasegmental feature vectors of isolated Cantonese syllables [1] and recognized tone in Mandarin monosyllables [2]. In case of continuous speech, the technique was used to compare all tone feature and tone models using half-tone model [3]. Hidden Markov models was employed to classify Thai tones and showed that the tones are independent from the vowels [4]. Fujisaki’s model was presented an analysis-by-synthesis algorithm to recognize Thai tone in continuous speech [5]. Comparing two speech types, we have to consider tone behavior of neighboring syllables, especially in case of the continuous speech. Thus, continuous speech is more difficult to classify than isolated speech [3]. Although the papers show the recognition accuracy over 80% by using the software simulation, nobody shows the experimental results in hardware implementation.

In this paper, we focus on VLSI architecture of tone classification function for isolated-word speech recognition. The architecture is separated into feature extraction and tone classification function parts. The feature extraction part is divided into three processes. The processes can be presented by following steps. Firstly, voice detection is represented to find the short-time energy of voice speech in stationary speech signal. Secondly, pitch period estimation calculates the time between the peak positions from voiced speech. And the fundamental frequency is detected. Finally, slope classification finds the tone features by computing the slope of entire \( F_0 \) curve. And then, the maximum and minimum point of \( F_0 \) in the first and second half syllable is detected. In the tone classification function part, look-up table conditions are decided by using three feature parameters from feature extraction part. And then, the tone results are detected by searching the maximum vote from the table. Therefore, the accuracy rate of our proposed architecture is similar to use the speech algorithm programming. However, the processing time is improved.

The rest of the paper is organized as follows. Section II gives an ordinary tone classifier. Section III discusses the proposed architecture of tone classification function-based isolated-word speech recognition. In Section IV, the FPGA implementation is described. The architecture experiments and results are given in Section V. Finally, discussion and conclusion are presented in Section VI and Section VII, respectively.