Improvement Voiced and Unvoiced Classification Technique Based on Real Time Processing Using FPGA Board

Chalermkiat Sutacha[†] and Jakkree Srinonchat[*]

[†]Computer Technology, Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University.
2 Ratchathani Road, Mueang, Ubon Ratchathani 34000, Thailand

[*]Department of Electronics and Telecommunication Engineering
Faculty of Engineering, Rajamangala University of Technology Thanyaburi
39 M.1, Rangsit-Nakornnayok Road, Thanyaburi, Pathumthani, 12110, Thailand

Abstract

In speech signal processing, the amount of data analyzing requires a long time process. One of the pre-processing techniques to make the speech processing faster is the voiced and unvoiced (V/UV) classification. This article presents an improvement V/UV classification technique based on real time processing using Field Programmable Gate Array (FPGA) board. The Virtex-II Pro board which consists of XC2VP30 chip as central processor unit is used in this experiment. The XC2VP30 chip consists of 30,816 logic cells and it can operate with external memory. The experiment results show that this system can be function on real time system. It used only 4.17-50 ms for time processing which does not effect to delay time process especially in real time system. Moreover, the output speech signal quality is still similar to the original speech signal. This is the major point that the XC2VP30 chip can be develop to use in speech compression and speech recognition.

Keywords: Voiced and unvoiced classification, Xilinx XC2VP30, Real time processing system

1. Introduction

In speech signal processing, the amount of data analyzing requires a long time process [1]. To make a process faster, speech signal have to reduce the size of speech signal [2-3]. One of the pre-processing techniques to make the speech processing faster is the voiced and unvoiced (V/UV) classification. In [1] describes an analysis and comparison of V/UV classification techniques. That technique operates based on the weight distribution and frequency distribution. The results show that the best performance is the fifth technique which provides the least error. However, it is only a simulation result. If it operates in the real time system, the delay time might occur. This problem can be solved, if it uses the fast process. The Field programmable logic array (FPGA) is one of the technologies that can be used to solve this problem. FPGA implementation has been successfully applied for signal processing area such as in [4-5]. FPGA can operate a very high speed level, since it can carry out parallel processing. The standalone algorithm for real time speech enhancement can suppress stationary acoustic noise from speech by subtracting the spectral noise bias calculated during non-speech activity, while adding the unique option of dynamic moving averaging to it, it can now periodically upgrade the estimation and cope up with changes in noise level [4-5]. Also voice identification algorithms [6] are traditionally developed on-off the shelf digital signal processors with algorithm-hardware migration technologies and, in particular, the realization of speech features extraction using FPGA. Moreover, the voice activity detection algorithm implements on the FPGA which uses FPGA Discovery - III XC3S200F4 board [7] which the speech signal is computed in the Fixed-Point format and operated in real time system. The system uses several DPRAMs (Dual Port RAMs) inside the FPGA as parallel buffers in which speech data and the intermediate result are stored, to speed up processing. Processing modules [8] use sampling data is 16 KHz and the resolution of each sample is 16 bits and the intermediate result from the parallel buffers. The system can generate the VAD result every 15 ms of VAD per frame; it can be calculated in real-time.

This article proposes an improvement V/UV classification technique based on real time processing using FPGA board. According to the V/UV classification techniques in [1] require a long time processing, thus the FPGA processor is employed to implement and improve those technique in this article. The real-time processing module is implemented on Virtex-II Pro (XC2VP30) and operates with LM4550 AC97 audio codec. All modules

* Corresponding author, Email: Jakkree.s@en.rmutt.ac.th
Manuscript received: 30 September 2012: Selected Paper
Revised: 15 March 2013
KMITL-STJ; Special Section; Part C: CIT

are programmed using Very-High-Speed Integrated Circuit Hardware Description Language (VHDL).

2. V/UV Classification Real Time Process

This article presents the improvement V/UV classification technique based on real time processing using FPGA board. The process is described as follow:

- Speech data configured to take analog sound input. The data is supplied to the codec and analog to digital conversion therefore the proposed system can select input from microphone and computer.
- Sampling data is stored to parallel buffers with window size.
- Speech is extracted feature for V/UV classification.
- V/UV classification operated with frame by frame.
- The data is supplied to the codec and digital to analog conversion provides the analog sound output from the speaker.

This technique uses equation (1-4) to identify V/UV signal. The Ratio 1 defines as the percent of adjust threshold range which is 1.5% of amplitude for V/UV. The V/UV classification technique with threshold is show in Fig 1.

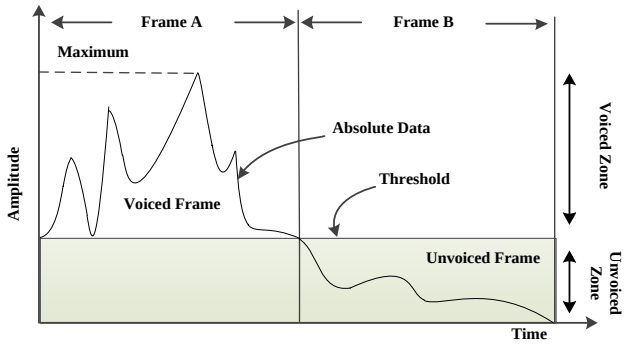


Figure 1. The threshold of V/UV classification

With this regard;

$$Th = Ratio1 \tag{1}$$

$$S = \begin{cases} 1, & x(n) \geq Th; voiced \\ 0, & x(n) < Th; unvoiced \end{cases} \tag{2}$$

$$E_s = \sum_{n=1}^{w} S(n) \tag{3}$$

$$\mathrm{Re} = \begin{cases} E_s \geq ratio2 & ; voiced\ frame \\ E_s < ratio2 & ; unvoiced\ frame \end{cases} \tag{4}$$

Where

Th	is the threshold
$Ratio1$	is the percent of adjust threshold
S	is result of voiced
E_s	is aggregate the sum of voiced
w	is the number of sampling in one frame
n	is the number of sampling
$Ratio2$	is percent of V/UV classification
Re	is result of V/UV classification

The speech data has converted to the absolute value according to reduce the processing steps. The sampling data is defined as voiced if it is greater than or equal to the Ratio1. The sampling data is defined as unvoiced if it is less than the Ratio1. The Ratio2 is defined as 50 (25%) for voiced frame and unvoiced frame classification. The frame is defined as voiced frame if the sum of data is greater than or equal to the Ratio2. The frame is defines as unvoiced frame if the sum of data is less than the Ratio2.

3. Hardware Implementation

The propose system is implemented on Virtex-II Pro. The audio codec on the Virtex-II Pro board is an LM4550 AC97 audio codec which uses 18-bit Sigma-Delta ADCs and DACs. The LM4550 provides 90 dB of dynamic range. The sample rate for the ADCs and DACs can be separately program with a resolution of 1 Hz to any rate in the range 4 - 48 KHz. LM4550 AC97 audio codec interfacing with FPGA is shown in Fig 2. The FPGA provides connection signals to the AC97 codec which is used as an interface with the AC97 controller. The system can select input from microphone and computer.

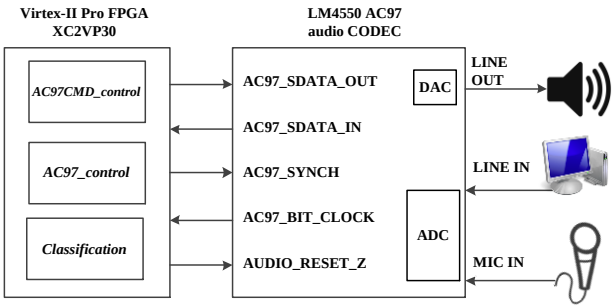
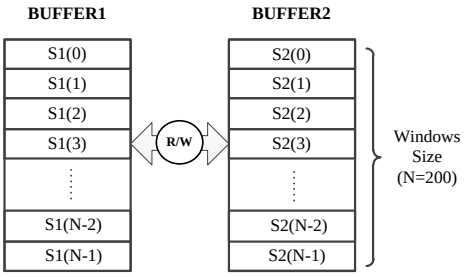


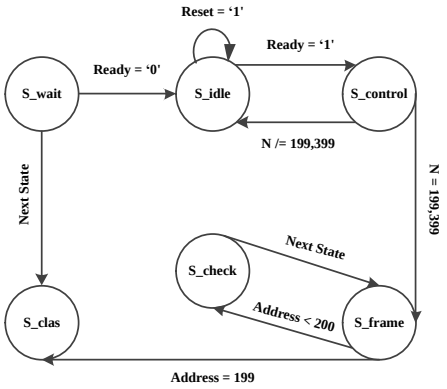
Figure 2. LM4550 AC97 audio codec interfacing with FPGA

The window of sampling data is design as 200 sampling per frame [9]. Fig 3(a) shows the parallel buffers where sampling data are stored. The buffers consist of 2 frames. Each buffer is a circular queue which performs a read or writes process to occur simultaneously. When it uses to read data, it needs to calculate and write sampling data simultaneously. The pointer indicates the current frame. When the next frame writing is finished, the pointer is updated to the next frame position.

The operation of V/UV classification is categorized into six stages: S-idle, S-control, S-frame, S-check, S-class, and S-wait. Fig 3(b) shows the finite state machine (FSM) diagrams which is used in this article. The S_idle state is the initial state. The Reset forces the FSM to enter the S_idle state if Reset is logic high. The FSM transfers from the S_idle state to the S_control state when Ready is logic high. At the S_control state, the buffers keep speech data and count N + 1 that control by address of buffers. The FSM transfers from the S_control state to the S_wait state when N is not equal to 199 and 399. The FSM transfers from the S_control state to the S_frame state when N is equal to 199 and 399. At the S_frame state and S_check, The state read data from buffers and evaluate the data for V/UV classification. The FSM transfers from the S_frame state to the S_clas and S_wait when Address is 199. The FSM transfers from the S_wait state to S_idle state when Ready is logic low.



(a) Parallel buffers



(b) Finite state machine

Figure 3. The frame of N sampling by windows size

4. EXPERIMRNT RESULTS

The system consists of three main entities: AC97CMD_control entity, Classification entity and AC97_control entity. The AC97_control entity generates timing signals for LM4550 AC97 audio codec interfacing with AC97CMD_control entity. Thus all timing signals are synchronized to the Bit_Clk signal which is an input to the AC97_control entity. The Classification entity is the finite state machine of V/UV classification that is shown in Fig 4. All the process modules are implemented using VHDL.

Figure 5 shows the improvement voiced and unvoiced classification technique based on real time processing module is implemented on Xilinx Virtex-II Pro (XC2VP30) FPGA device. The sampling rate of the implemented system is 4 KHz - 48 KHz and resolution of each is 18bits. The real-time processing systems take time 4.17 ms - 50 ms for voiced frame and unvoiced frame classification.

The simulation result shows S_idle, S_control, S_frame, S_check, S_class and S_wait of the finite state machine. Fig 6. shows the simulation results of V/UV classification process when N is equal to 199. Also Fig 7. shows the simulation results of V/UV classification process when Address is equal to 199.

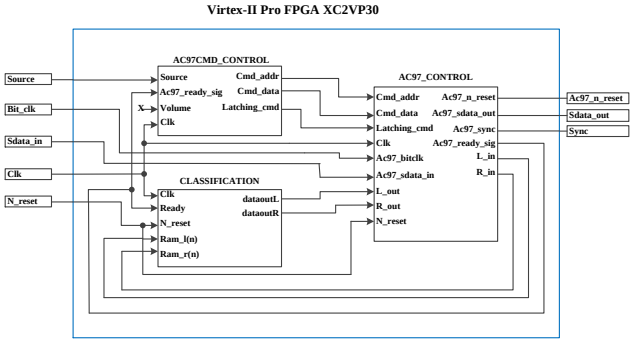


Figure 4. The frame of N sampling by windows size

Figure 5. The V/UV classification based on real-time processing system

TABLE I. DEVICE UTILIZATION SUMMARY

Name	Without U/V classification		With U/V classification	
	Used	Util.	Used	Util.
Slice Flip Flops	163	0.6%	191	0. 7%
occupied slice	124	0.9%	301	2.2%
4 input LUT	136	0.5%	502	4.0%
bonded	10	1.8%	10	1.8%
BUFMUXs	2	12.5%	2	12.5%

5. CONCLUSION

This article presents the improvement voiced and unvoiced classification technique based on real time

processing using FPGA board. The proposed system is implemented on Xilinx Virtex-II Pro (XC2VP30) FPGA device. The experiment results show this system can be function on real time system. It used only 4.17-50 ms for time processing which does not effect to delay time process on real time system. Moreover, the quality of the output speech signal is still similar to the original speech signal. The device utilization generates from Xilinx ISE 10.1 for the Xilinx FPGA devices show in Table 1. The implemented system inside Xilinx XC2VP30 FPGA device without U/V classification uses 0.6% of Slice Flip Flops, 0.9% of occupied slice, 0.5% of 4 input LUT, 1.8% of bonded and 12.5% of BUFMUXs respectively. The implemented system inside Xilinx XC2VP30 FPGA device with U/V classification uses 0. 7% of Slice Flip Flops, 2.2% of occupied slice, 4.0% of 4 input LUT, 1.8% of bonded and 12.5% of BUFMUXs. This is the major point that the XC2VP30 chip can be develop to use in speech compression and speech recognition.

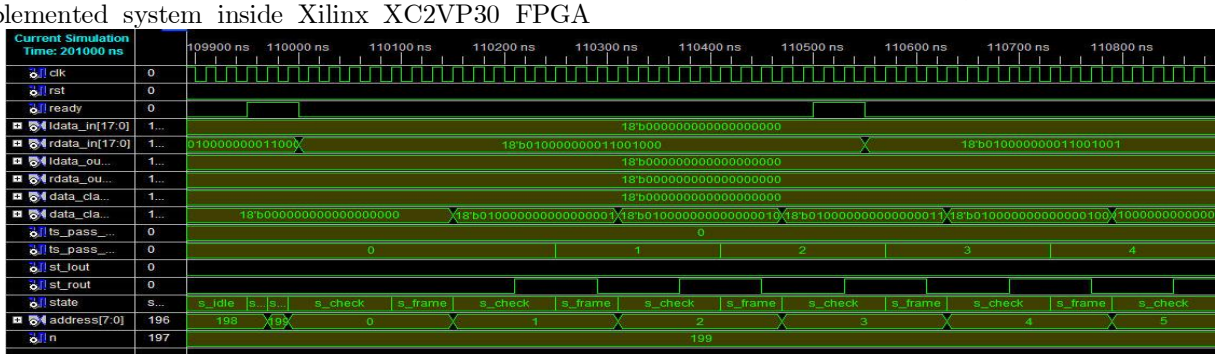


Figure 6. The Simulation result of V/UV classification process when N is equal to 199

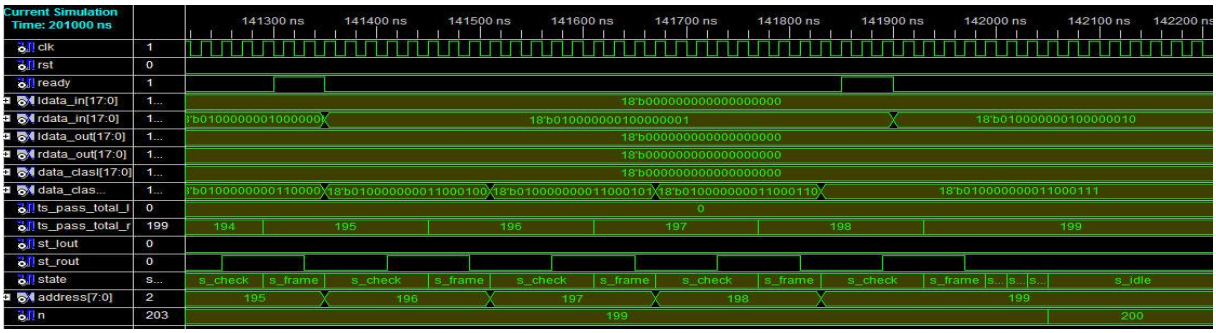


Figure 7. The Simulation result of V/UV classification process when *Address* is equal to 199

Acknowledgment

Authors would like to express my gratitude to all those who gave me the possibility to complete this research. We would like to thank the National Research Council of Thailand for giving a grant to research project.

References

[1] Kornsing, S., Pattanaburi, K., Sutacha, C. and Srinonchat, J. Analysis and Comparison of Voiced and Unvoiced Classification Techniques. Conference of Electical engineering network 2011 of rajamangala university of technology (EENET 2011), pp. 128-131.

[2] Srinonchat, J. Comparison of the Efficiency of Ordered and Disordered Codebook Techniques in Speech Coding. IEEE International Conference on Information and Communication Systems, 2005, pp. 195 – 198.

[3] Srinonchat, J. Enhancement Artificial Neural Networks for Low-Bit Rate Speech Compression system. IEEE International Symposium on Communications and Information Technologies, 2006, pp. 195-198.

[4] Jung, U, C., Lee, R, D., Park, H. J., Kim, H, S., Lee, C, H., Choi, S, Jung. and Jeon, W, J. An FPGA-Based Voice Signal Preprocessor for The Real-Time Cross-Correlation. International Conference on Control, Automation and Systems, 2007, pp. 793 – 797.

[5] Rahman, U, M, T. and Rashid, A. B. M. H. FPGA Implementation of Real Time Acoustic Noise Suppression by Spectral Subtraction using Dynamic Moving Average Method. IEEE Symposium on Industrial Electronics and Applications, 2009, pp. 365 – 370.

[6] Xu, J., Ariyaeeinia, A. and Sotudeh, R. User Voice Idenfication on Fpga. Perspectives in Pervasive Computing, 2005, pp. 75 - 83.

[7] Sutacha, C. and Srinonchat, J. Improvement Speech Signal Real Time Processing Techniques Using FPGA Board. Electical engineering network 2011 of rajamangala university of technology 2012, pp.585-588.

[8] Jung, J., Jin, S., Kim, D., Kim, H, S., Choi, K, S. and Joen, W, J. A Voice Activity Detection System Based on FPGA. International Conference on Control, Automation and Systems, 2010, pp. 2304 - 2308

[9] Pattanaburi, K. and Srinonchat, J. Enhancement Pattern Analysis Technique for Voiced/Unvoiced Classification. IEEE-International symposium on computer, consumer and control, 2012, pp. 389-392.

Chalermkiat Sutacha receives a M.Eng in Electrical Engineering from Rajamangala university of Technology Thanyaburi, He is currently lecturer and researcher at Computer Technology, Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University.

Jakkree Srinonchat receives a Ph.D in Electrical Engineering especially signal processing from Northumbria University, UK in 2005. He is currently lecturer and researcher at Signal Processing Research Laboratory at Dept. of Electronics and Telecommunication, Faculty of Engineering, Rajamangala university of Technology Thanyaburi.