Volume 3, Issue 1 (2015) 216-220 ISSN 2347 - 3258
International Journal of Advance Research and Innovation

FPGA Implementation of Optimized Spiking Neural Network for Efficient Speak recognition System

K. Syedthathul Fathima, R. Joshua Arul Kumar
Department of Electronics and Communication Engineering, MAM College Of Engineering, Trichy, India

Article Info
Article history:
Received 19 January 2015
Received in revised form
15 February 2015
Accepted 28 February 2015
Available online 15 March 2015

Keywords
Speech Recognition,
Spiking Neuron,
FPGA,
Spikingneural Network Feature Extraction

Abstract
A field-programmable gate array (FPGA)-based speech measurement and recognition system is the focus of this paper, and the environmental noise problem is its main concern. To accelerate the recognition speed of the FPGA-based speech recognition system. Furthermore, the empirical mode decomposition is used to decompose the measured speech signal contaminated by noise into several intrinsic mode functions (IMFs). The IMFs are then weighted and summed to reconstruct the original clean speech signal. Unlike previous research, in which IMFs were selected by trial and error for specific applications, the weights for each IMF are designed by the genetic algorithm to obtain an optimal solution.

1. Introduction

A field-programmable gate array (FPGA)-based speech measurement and recognition system is the focus of this paper, and the environmental noise problem is its main concern. To accelerate the recognition speed of the FPGA-based speech recognition system. Furthermore, the empirical mode decomposition is used to decompose the measured speech signal contaminated by noise into several intrinsic mode functions (IMFs). The IMFs are then weighted and summed to reconstruct the original clean speech signal. Unlike previous research, in which IMFs were selected by trial and error for specific applications, the weights for each IMF are designed by the genetic algorithm to obtain an optimal solution resolution, zero crossings and mean square error. Different iterations are used to reduce calculation time and also mean square error.

2. Architectural Representation of Speech Recognition System

Speech recognition stands as a medium, or a communications Ambassador between machines and people, ever promising to deliver natural speech. As a result, Speech Recognition has the ability of unifying or incorporating many other current technologies, ultimately fusing many features and functions granted by today's technologies.

Audio signals (voice) most commonly used input for speech recognition systems. In the past few years, Speech Recognition has experienced many improvements that have enabled its machines, specifically computers, to perform elaborate tasks such as Dictation, and Command Recognition. However, speech recognition (by a machine) is a very complex problem. Vocalizations vary in terms of accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed. Speech is distorted by a background noise and echoes, electrical characteristics.

3. Speech recognition

Speech Recognition is important for machine to understand human voices and perform the action according to human commands. Speech recognition is highly research object and it is useful in area of pattern recognition, involving physiology, psychology, linguistics, computer science and signal processing, and many other fields, even to the people involved in body language.

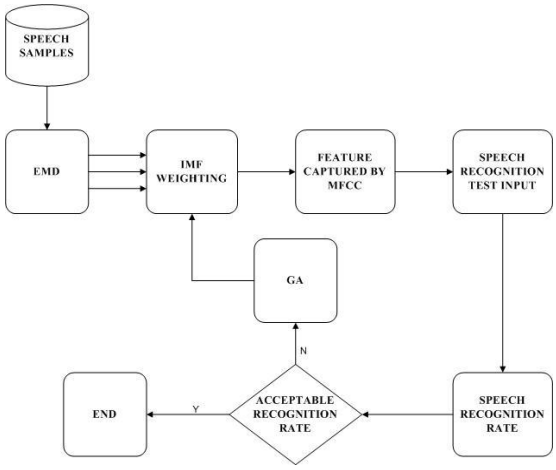


Fig: 1. Proposed Strategy for Speech Recognition

Speech recognition is the technology that makes it possible for a computer to identify the components of human speech. The process can be said to begin with a spoken utterance being captured by a microphone and to end with the recognized words being output by the system.

Research in speech recognition began somewhere in the late 1940's and the rapid progress in computer technology has been of great importance to the development of this field. Today, an increasing number of products incorporate speech technology. Some of these are found in industry. In situations where the hands and eyes of persons are engaged elsewhere, voice control can be a great advantage. Other applications are booking systems over the telephone, in which case speech is a more intuitive means of communication than pressing buttons. Yet other applications include various forms of aids for the disabled, dictation systems, as well as toys and games.

Corresponding Author,
E-mail address: vincysweety.11@gmail.com

Ideally, a speech recognizer would be able to accurately identify all possible words spoken by any person in any environment. In reality, however, system performance depends on a number of factors. Large vocabularies, multiple users, as well as continuous speech (as opposed to words spoken in isolation) are factors that complicate the task of recognition. The same is true for speech in noisy environments. In speech recognition, recognition performance is usually affected by the confusing set in the vocabulary

4. Empirical Mode Decomposition

The Empirical Mode Decomposition (EMD) was proposed as the fundamental part of the Hilbert–Huang transform (HHT). The Hilbert Huang transform is carried out, so to speak in 2 stages. First, using the EMD algorithm, we obtain intrinsic mode functions (IMF).

Then, at the second stage, the instantaneous frequency spectrum of the initial sequence is obtained by applying the Hilbert transform to the results of the above step. The HHT allows obtaining the instantaneous frequency spectrum of nonlinear and non-stationary sequences. These sequences can consequently also be dealt with using the empirical mode decomposition.

In contrast to the previously mentioned Fourier transform and wavelet transform, the EMD decomposes any given data into intrinsic mode functions (IMF) that are not set analytically and are instead determined by an analyzed sequence alone. The basic functions are in this case derived adaptively directly from input data.

EMD is based on three assumptions:

(i) The signal has at least one minimum and one maximum (non-monotonic function),
(ii) The time difference between successive extrema defines the characteristic time scale,
(iii) If there are no extrema but only inflection points, the data may be differentiated,

Then EMD applied and the result obtained by integrating the components. The pertinent results we obtained from this study include an analytic expression of the relationship between the energy density and the mean period of the IMF components derived from uniformly distributed white noise through EMD, an analytic expression of the energy density distribution and its spreading function. All the analytic expressions are tested against the results produced using numerically generated random data.Let the original signal be X (t) and a temporary signal,

Temp(t)=X(t) (1)

h(t)=Temp(t)–m(t) (2)

Check whether signal h (t) satisfies the conditions of the IMF or not. If it does, then the first IMF is obtained as imf_1 (t) = h (t) and proceed to the next step, or else, assign signal h (t) as Temp (*t*) and go back to step 1.

Calculate residue r_1 (t) as

$r_1(t)=Temp(t)-imf_1(t)$ (3)

Assign signal r_1 (t) as X (t), and repeat steps 1 and 2 to find imf_2 (t).

Repeat step 3 to find the subsequent IMFs as follows:

$r_n(t)r_{n-1}(t)-imf_n(t)$ (4)

n = 2, 3, 4 . . .

This step is completed when signal r_n (t) is constant or is a monotone function. After the EMD procedure steps 1–4 are finished, the following decomposition of X (t) is obtained:

$$X(t)=\sum_{i=1}^{n} imf_i(t) + r_n(t) \quad (5)$$

These methods have been useful for controlling the synthesizer parameters, such as information of the testing sample, duration, sampling rate, bit resolution, zero crossings and mean square error. In these empirical modes decomposition method to find the different types of iterations using different testing samples derived calculation time is reduced and also noise is removed. Different weightining functions are applied to the next chapter.

4.1 Intrinsic Mode Function

Any function having the same numbers of zero crossings and extrema and also having symmetric envelopes defined by local maxima and minima respectively is defined as an Intrinsic Mode Function. All IMF enjoys good Hilbert Transform:

C (t)= a (t) (6)

By applying EMD a signal can be decomposed into a set of mono component functions called Intrinsic Mode Functions (IMFs). A mono component function indicates an oscillating function close to the most common and basic elementary harmonic function. Therefore, the IMFs contain frequencies ranging from the highest to the lowest ones of the signal presented as amplitude and frequency modulated (AM-FM) signals, where the AM carries the envelope and the FM is the constant amplitude variation in frequency and calculated using a sifting process. To accomplish this, an IMF must satisfy two conditions:

i) The number of extrema (local maxima and minima) and the number of zero crossings must either equal or differ at most by one.
ii) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The first condition is necessary for oscillating data to meet the strict conditions needed to calculate the instantaneous frequency that presents the oscillation frequency of a signal at certain point of the time. It leads to a narrow-band signal. So using the EMD a rhythmic and harmonic analysis of the signal can be performed.

The second condition requires symmetric upper and lower envelopes of an IMF which makes the signal ready for modulation as the IMF component is decomposed from the original data.

5. Genetic Algorithm in Speech Recognition

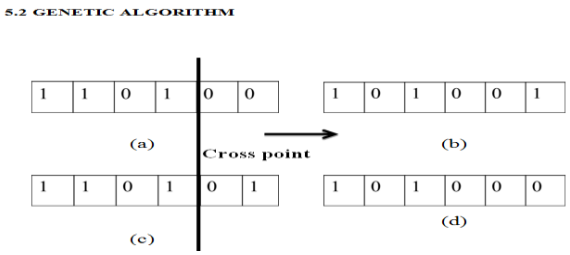


Fig: 2. Bit-string Crossover of Parents a & b to form of spring c & d

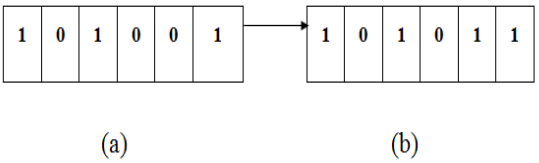


Fig: 3. Bit-flipping Mutation of Parent a to form off spring b

The most popular technique in evolutionary computation research has been the genetic algorithm. In the traditional genetic algorithm, the representation used is a fixed-length bit string. Each position in the string is assumed to represent a particular feature of an individual, and the value stored in that position represents how that feature is expressed in the solution. Usually, the string is evaluated as a collection of structural features of a solution that have little or no interactions. The analogy may be drawn directly to genes in biological organisms. Each gene represents an entity that is structurally independent of other genes.

The main reproduction operator used is bit-string crossover, in which two strings are used as parents and new individuals are formed by swapping a sub-sequence between the two strings as shown in fig 5.1. Another popular operator is bit-flipping mutation, in which a single bit in the string is flipped to form a new offspring string as shown in fig 5.2.

Varieties of other operators have also been developed, but are used less frequently. A primary distinction that may be made between the various operators is whether or not they introduce any new information into the population. Crossover, for example, does not while mutation does. All operators are also constrained to manipulate the string in a manner consistent with the structural interpretation genes.

5.1 Train codebook by using GA

A codebook for the quantization of speech signal features is generated by using GA. Thereafter, a modeled for speech recognition through the codebook in next section. It is well known that GA solves a optimization problem through the evolution process: selection, crossover, and mutation. It is important to define the chromosome in GA to meet the problem in hand. Each chromosome has multiple genes, and the number of the parameters in a problem will determine the number of genes. Encoding of genes can be divided into the following three ways: (a) binary encoding (b) real number encoding (c) symbol encoding as shown in fig 5.2

An effective way to get the local information which is vital to the non-stationary signals. Very often seriously degrades upper bound performance. The various sources of this mismatch include additive noise, channel distortion, different speakers, different speaking rate and modes, and so on. Model-based approaches, compensation is performed on pertained recognition model parameters, so that the modified recognition models will be able to classify the mismatched testing feature parameters collected in the application environment.

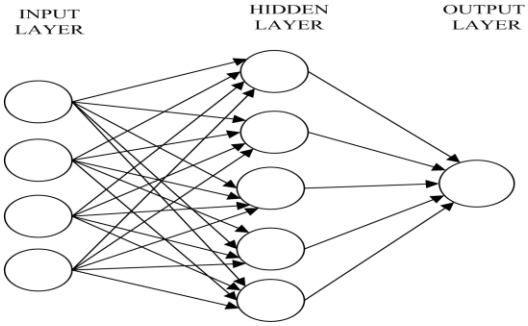


Fig: 4. Flow Chart

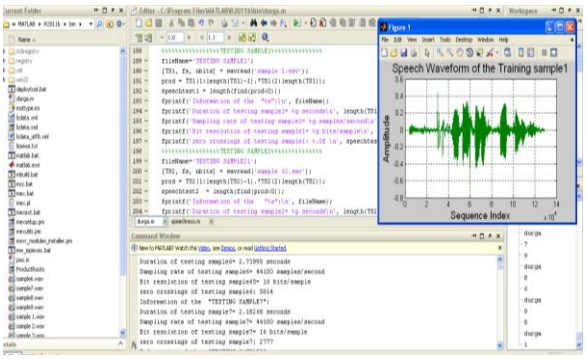
6. Explanation of Hardware Results

The advantages of a designed speech recognition system as implemented on an FPGA-based DSP platform with SOC architecture. All noise-affected speech is first decomposed into several IMFs, using the EMD process. Since each IMF contains a greater or lesser speech signal, these IMFs are then weighted by their corresponding weights and then summed to recover the original speech signal. It is noted that the weights are initially randomized and will be trained by the GA thereafter. The MFCC process is then performed on the recovered speech to extract its features. Also, in the training phase, the GA is used to train the weights to get an acceptable recognition rate. The stop criterion of the GA depends on the recognition rate and the number of generations evolved in the GA.

The algorithms are first created on a personal computer by using high level program language and are then downloaded to the FPGA on Spartan 3 DSP. Thereafter, this board with these algorithms on the FPGA chip is used to receive speech from the MATLAB on the board, then recognize the speech, and finally display the recognition result on a seven segment light emitting diode(SEG7) and liquid crystal display (LCD) monitor.

After testing progress the output displayed on the board, speaker is recognized or speaker is not recognized. A push button is used to control the starting and the ending of the voice record, and a toggle switch is used to control the sampling rate of AUDIO codec. An AUDIO controller is used to receive speech data and an SEG7 and an LCD are used for the display of recognition results. As for the hardware implementation, static RAM (SRAM) and Flash RAM are used for the storage of source code and the testing of the signal, respectively. The bus Protocol is used to control the register of the platform.

6.1 Results



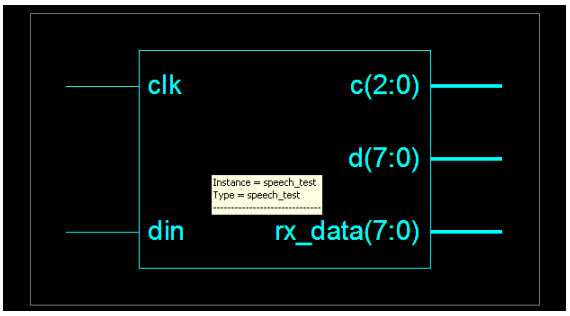


Fig: 6. RTL Schematic

Design Object List - I/O Pins

I/O Name	I/O Direction	Loc	Bank	I/O Std.
c<0>	Output	p169	BANK1	
c<1>	Output	p171	BANK1	
c<2>	Output	p168	BANK1	
clk	Input	p181	BANK1	
d<0>	Output	p167	BANK1	
d<1>	Output	p166	BANK1	
d<2>	Output	p165	BANK1	
d<3>	Output	p162	BANK1	
d<4>	Output	p161	BANK1	
d<5>	Output	p156	BANK2	
d<6>	Output	p155	BANK2	
d<7>	Output	p154	BANK2	
din	Input	p176	BANK1	
rx_data<0>	Output	p68	BANK5	
rx_data<1>	Output	p67	BANK5	
rx_data<2>	Output	p65	BANK5	
rx_data<3>	Output	p64	BANK5	
rx_data<4>	Output	p63	BANK5	
rx_data<5>	Output	p62	BANK5	
rx_data<6>	Output	p61	BANK5	
rx_data<7>	Output	p58	BANK5	

Fig: 7. Pin Assignment

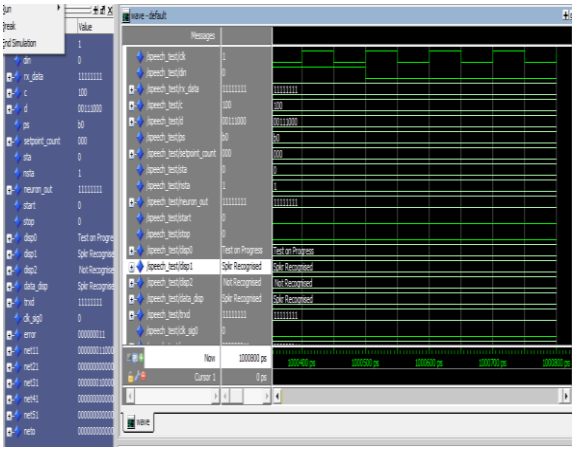


Fig: 8. Simulation Result

7. Conclusion and future work

I have developed a speech recognition system with reduced noise and less computation burden in FPGA using Empirical Mode Decomposition (EMD) and Genetic Algorithm (GA). Hardware implementation of proposed system is achieved by VHDL and synthesizer tools. The advantage of this mechanism is that it retains understandability of the speech. In this project, techniques such as EMD and GA have been applied to speech recognition. The simulation results of some of the speech are synthesized. These methods have been useful for controlling the synthesizer parameters, such as information of the testing sample, duration, sampling rate, bit resolution, zero crossings and mean square error. Different iterations are used to reduce mean square error calculation time. Feature based method using better recognition rate is obtained.

7.1 Future Work

Future work will include fault detection for various fault symptoms such as excessive noise, frequency distortion and automatic reduction of threshold parameters for magnitude and shape diagnosis when changes occur to the surrounding environment and the type of future work includes,

(i) Expansion of sound classes.
(ii) Integration of spotting system.
(iii) Researching on signal separation for a complicated environment with overlapped sound sources.

References

[1] Wang, H. Leung, A. P. Kurian, H. J. Kim, and H. Yoon, A deconvolutiv neural network for speech classification with applications to homeservice robot, IEEE Trans. Instrum. Meas., 59(12), 2010, 3237–3243

[2] L. Buera, A. Miguel, O. Saz, A. Ortega, E. Lleida, Unsupervised data-driven feature vector normalization with acoustic model adaptation 2572 IEEE Transactions on Instrumentation and Measurement, 61(9), 2012 for robust speech recognition, IEEE Trans. Audio, Speech, Lang.Process., 18(2), 2010, 296–309

[3] J. W. Hung, W. H. Tu, Incorporating codebook and utterance information in cepstral statistics normalization techniques for robust speechrecognition in additive noise environments, IEEE Signal Process. Lett., 16(6), 2012, 473–476

[4] L. D. Persia, D. Milone, H. L. Rufiner, M. Yanagida, Perceptual evaluationof blind source separation for robust speech recognition, Signal Process., 88(10), , 2008, 2578–2583

[5] J. Kim, B. J. You, Fault detection in a microphone array by intercorrelationof features in voice activity detection, IEEE Trans. Ind. Electron., 58(6), 2011, 2568–2571

[6] Y. Zhan, H. Leung, K. C. Kwak, H. Yoon, Automated speaker recognition for home service robots using genetic algorithm and Dempster–Shafer fusion technique, IEEE Trans. Instrum. Meas., 58(9), 2009, 3058–3068

[7] C. T. Ishi, S. Matsuda, T. Kanda, T. Jitsuhiro, H. Ishiguro, S. Nakamura, N. Hagita, A robust speech recognition system for communication robots in noisy environments, IEEE Trans. Robot., 24(3), 2013, 759–763

[8] C. W. Hsu, L. S. Lee, Higher order cepstral moment normalization for improved robust speech recognition, IEEE Trans. Audio, Speech, Lang. Process., 17(2), , 2009, 205–220

[9] A. Sankar, C. H. Lee, A maximum-likelihood approach to stochastic matching for robust speech recognition," IEEE Trans. Speech Audio Process., 4(3), 1996, 190–202

[10] C. H. Lee, On stochastic feature and model compensation approaches to robust speech recognition, Speech Commun., 25((1–3), 1998, 29–47
[11] M. J. F. Gales, S. J. Young, Robust speech recognition in additive and convolutional noise using parallel model combination, Comput. Speech Lang., 9(4), 1995, 289–307
[12] Y. Tsao, C. H. Lee, An ensemble speaker and speaking environmentmodeling approach to robust speech recognition, IEEE Trans. Audio, Speech, Lang. Process., 17(5), 2009, 1025–1037
[13] S. Windmann, R. Haeb-Umbach, Parameter estimation of a statespacemodel of noise for robust speech recognition, IEEE Trans. Audio, Speech, Lang. Process., 17(8), 2009, 1577–1590
[14] N. E. Huang, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A, 454(1971), 1998, 903–995
[15] X. Li, X. Zou, R. Zhang, G. Liu, Method of speech enhancement based on Hilbert-Huang transform, in Proc. 7th World Congr. Intell. Control Autom., Jun. 25–27, 2008, 8419–8424
[16] O. Cheng, W. Abdulla, Z. Salcic, Hardware–software code sign of automatic speech recognition system for embedded real-time applications, IEEE Trans. Ind. Electron., 58(3), 2011, 850–859
[17] C. C. Tsai, H. C. Huang, S. C. Lin, FPGA-based parallel DNA algorithm for optimal configurations of an omnidirectional mobile service robot performing fire extinguishment, IEEE Trans. Ind. Electron., 58(3), 2011, 1016–1026