ISSN (online): 2349-784X

An Efficient SVM based Speaker Verification System

P. Suganthi
Post Graduate Scholar
Department of Electronics and Communication Engineering
Christian College of Engineering & Technology
Oddanchatram, India

M. N. Srinithya
Assistant Professor
Department of Electronics and Communication Engineering
Christian College of Engineering & Technology
Oddanchatram, India

Abstract

The speaker verification system is to identify information from speaker voice and this system is mainly used to voiceprints and bio metric identification process. In the proposed system, the useful information from speech signal is extracted using DFT methodology. The SVM module will generate test vector for the input speech signal. The SVM module consists of Gaussian Kernel unit and Scaling unit. The Gaussian kernel unit consist of Gaussian kernel processing element which is used to evaluate kernel values for test vector and support vector. The existing system uses CORDIC circuit in exponential processing element to minimize architecture. This work uses several adders and shifters to minimize the shifting and rotation operation. The proposed system uses approximate multiplier circuit in the normal processing element to minimize architecture. The scaling unit performs multiplication and the remaining operations of decision value evaluation. Finally, the spoken words and claimed speaker is verified. The power and circuit complexity is reduced in this system.
Keywords: Gaussian kernel Processing Element(GK-PE), Scaling Unit, Discrete Fourier Transform(DFT), Coordinate Rotate Digital Integrated Circuit(CORDIC), low power

I. INTRODUCTION

Speaker recognition is an important branch of speech processing. It is the process of automatically recognizing who is speaking by using speaker-specific information included in the speech waveform. It is receiving increasing attention due to its practical value and has applications ranging from police work to automation of call centers. Speaker recognition can be classified into speaker identification (discovering identity) and speaker verification (authenticating a claim of identity). In a speaker identification system, an unknown speaker is identified as one of the speakers in the database. In a speaker verification system, a person ‘s identity is validated based on his/her speech feature. Speaker recognition has been extensively studied for the last decades. In speech verification system, the speech feature extraction and classification are the two essential issues. Speaker recognition has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioral patterns (e.g., voice pitch, speaking style).

Speaker verification has earned speaker recognition its classification as a "behavioral biometric". Speaker identification systems can also be implemented covertly without the user's knowledge to identify talkers in a discussion, alert automated systems of speaker changes, check if a user is already enrolled in a system, etc. In forensic applications, it is common to first perform a speaker identification process to create a list of "best matches" and then perform a series of verification processes to determine a conclusive match. Each speaker recognition system has two phases: Enrollment and verification. During enrollment, the speaker's voice is recorded and typically a number of features are extracted to form a voice print, template, or model. In the verification phase, a speech sample or "utterance" is compared against a previously created voice print. For identification systems, the utterance is compared against multiple voice prints in order to determine the best match(es) while verification systems compare an utterance against a single voice print. Because of the process involved, verification is faster than identification. Speaker recognition systems fall into two categories: text-dependent and text-independent. If the text must be the same for enrollment and verification this is called text-dependent recognition. In text independent systems both acoustics and speech analysis techniques are used. The various technologies used to process and store voice prints include frequency estimation, hidden Markov models, Gaussian mixture models, pattern matching algorithms, neural networks, matrix representation, Vector Quantization and decision trees. Some systems also use "anti-speaker" techniques, such as cohort models, and world models. Hidden Markov Models (HMMs) are, undoubtedly, the most employed core technique for Automatic Speech Recognition ASR). During the last decades, research in HMMs for ASR has brought about significant advances and, consequently, the HMMs are currently very accurately tuned for this application. Nevertheless, we are still far from achieving high performance ASR systems. Support Vector Machines (SVMs) are state-of-the-art classifiers. SVMs solution relies on maximizing the distance between the samples and the classification border.

This distance is known as the margin and by maximizing it, they are able to generalize unseen patterns. This maximum margin solution allows the SVM to outperform most nonlinear classifiers in the presence of noise, which is one of the longstanding problems in ASR.

II. PREVIOUS WORK

DSPs appeared on the market in the early 1980s. Since then, they have undergone an intense evolution in terms of hardware features, integration, and software development tools. DSPs are now a mature technology. This section gives an overview of the evolution of the DSP over their 25-year life span; specialized terms such as 'Harvard architecture', 'pipelining', 'instruction set' or 'JTAG' are used. The reader is referred to the following paragraphs for explanations of their meaning. DSPs have been used in accelerators since the mid-1980s. Typical uses include diagnostics, machine protection and feed forward/feedback control. In diagnostics, DSPs implement beam tune, intensity, emittance and position measurement systems. For machine protection, DSPs are used in beam current and beam loss monitors. For control, DSPs often implement beam controls, a complex task where beam dynamics plays an important factor for the control requirements and implementations. Other types of control include motor control, such as collimation or power converter control and regulation.

The speaker verification consists of Feature Extraction and Classification. The purpose of feature extraction is to convert the speech wave form to a set of features for further analysis. The speech signal is a slowly time-varying signal and when it is examined over a sufficiently short period of time, its characteristics are fairly stationary, whilst over long periods of time the signal characteristics change to reflect the different speech sounds being spoken. In many cases, short-time spectral analysis is the most common way to characterize the speech signal. Several possibilities exist for parametrically representing the speech signal for the speaker identification task, such as MFCCs, Linear Prediction Coding(LPC), Acoustic based Feature Extractions. Prior to classification. With this approach, however, efficiency of SVMs is limited by the errors in the segmentation stage. Other works cope with the variable time duration of the utterances embedding either, an HMM, or a Dynamic Time Warping algorithm, in the kernel of the SVM. It is not easy, however, to apply these last techniques to the problem of continuous speech because a previous segmentation in words Our solution to the SVM problems mentioned before consists in classifying each minimum unit of 25ms of voice (frame) as a basic class, a phone segment. With this approach we avoid the need to know where the words are and their time durations become unimportant.

In the paper SVM BASED SPEAKER VERIFICATION SYSTEM, the authors Jia-Ching Wang, Li-XunLian, proposed that speech recognition is a process of automatically recognizing who is speaking on the basis of the individual information included in speech waves. The speaker recognition system uses the acoustic features of speech that have been concluded to differ between individuals. This system is to design SVM classification architecture and this architecture is to improve the speaker verification system performance level. This system is used to identify the selected input voice signal information based on overall data base voice signal and this system is to effectively recognize the speech signal. This system is to focus the SVM classification architecture and this architecture consists of normal and exponent based processing element architecture and to optimize the kernel selection process. This paper does not identify the information clearly.

The recent results in SVM methods with the GMM super vector concept shows two natural methods for finding distances between GMM super vectors. One method is based upon an approximation to KL divergence between two GMM models, and the other method is based upon an L2 function space inner product. Both of these distances satisfy the Mercer condition typically required in SVM optimization [2].

The SVM was shown to provide complementary scoring information resulting in substantially lower error rates when it was fused with a GMM system. There are two types of trials target trials where the unknown speaker is the target and non-target trials where the unknown speaker is someone else. System errors for the first type are misses, the second type false alarms. System performance may then be characterized by the two error rate types: miss rate and false alarm rate. The requirement for likelihood scores for all trials using a common scale allows these two error rates to be determined multiple system operating point[3].

III. PROPOSED WORK

The proposed system consists SFE module, SVM module, Decision module in the Speech Feature Extraction module extract information in the speech signal. The SVM module classify the signal is the one speaker voice in the database. The decision module print the information depends upon the SVM module generate kernel values of test vector support vector both value distance is minimum then the spoken word is printed. This the block diagram of proposed system.

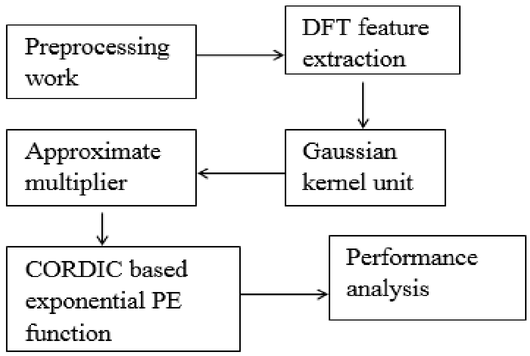

Fig. 1: Block Diagram of Proposed System

First to browse the given input voice signal in overall data base and to read the wave signal. Then to filter the selected voice signal and to reduce the noise level The Butterworth filter is one of the most popular analog filter design paradigms, first described by Stephen Butterworth. The basic philosophy of the conventional or integer order analog Butterworth filter is well practiced in various applications. Feature extraction is the process of collecting the information/data from the input voice signal. The feature is nothing but the data which consist of the several numerical values. Voice feature is another type of recognition system, which is used to recognize the unique person through the voice signal. Speech recognition system, Linear Prediction Coding Coefficients, Rasta PLP Coefficients, and Mel Cepstral Coefficients Linear predictive coding (LPCC) is used in representing the spectral envelope of a digital signal of speech in compressed form.

SVM locates a separating hyper plane in the feature space and classify points in that space It does not need to represent the space explicitly, simply by defining a kernel function. The kernel function plays the role of the dot product in the feature space. We use the approximate function in multiplier architecture. Because multiplier architecture is consisting of more circuit complexity level. Our work is mainly focused in DSP based classification process. So we don 't wants to accurate architecture. Because the signal multiplication process is expecting for approximate level based results. We design the 8* 8 bit approximate multiplier architecture level and to select the required test and support integer values. The CORDIC usually occupies a small area of hardware because it merely uses several adders and shifters. With the unfolding technique, CORDIC can also achieve high speed. The CORDIC architecture is used to reduce internal component count level in exponent based processing element architecture. We design kernel function logical architecture and it consists of normal processing element architecture and exponent processing element architecture. The speaker verification system is to identify the voice information effectively and to the hardware complexity. To reduce the circuit complexity level Finally this system is to identify the method. The processing element function is consisting of Euclidean distance based architecture between support vector and test vector.

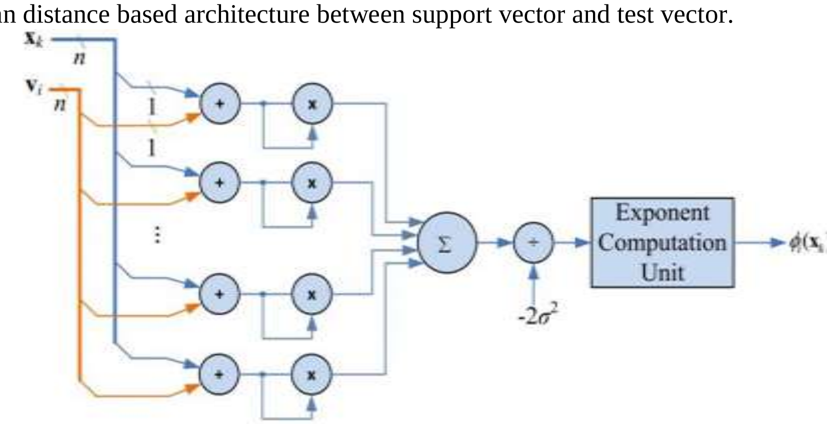

Fig. 2: Architecture of Gaussian Kernel Unit

The Gaussian kernel unit is designed to perform the Gaussian kernel evaluations of a test vector and all support vectors. The proposed architecture of the Gaussian kernel unit, which consists of four Gaussian kernel processing elements (GK-PEs) and five serial-to-parallel units (SPUs). For a test vector, each GK-PE performs its Gaussian kernel evaluation with one of the support vectors. Four GK-PEs are adopted in the Gaussian kernel unit so four support vectors can be processed simultaneously. The above-mentioned SFE module generates each dimension of an LPCC vector sequentially. An SPU is used to transform the serially input data to data that are input in parallel for GK-PE. One SPU receives the test vector while each of the other four SPUs takes its corresponding support vector. Each GK-PE incorporates a norm PE (Norm PE) and an exponential PE (Exp PE). The architectures of the two PEs are described in detail as follows.1) Norm PE: The Norm PE is responsible for calculating kernel value evaluation of vectors (test vector, support vector). compare the distance between the test vector and support vector is minimal distance then the signal is one of the desired signal.2) Exp PE: Among the several hardware implementation options in exponential operation, the CORDIC (Coordinate Rotate Digital Integrated Circuit) method is used. The CORDIC usually occupies a small area of hardware because it merely uses several adders and shifters. With the unfolding technique, CORDIC can also achieve high speed. The advantage of this paper is to reduce the circuit complexity level so the architecture will be optimized. The power consumption level will be reduced .so the classification performance level will be increased. The paper can be used for voiceprint, forensic and biometric solution applications.

IV. RESULTS

The architecture consists of a SFE module, an SVM module, and a decision module. The SFE module yields the LPC cepstrum vector. The SVM module evaluates all of the required kernel values, performs scaling multiplications, and completes the remaining operations of decision value evaluation.
The decision module computes the overall score of a test utterance to make an accepting or rejection decision.

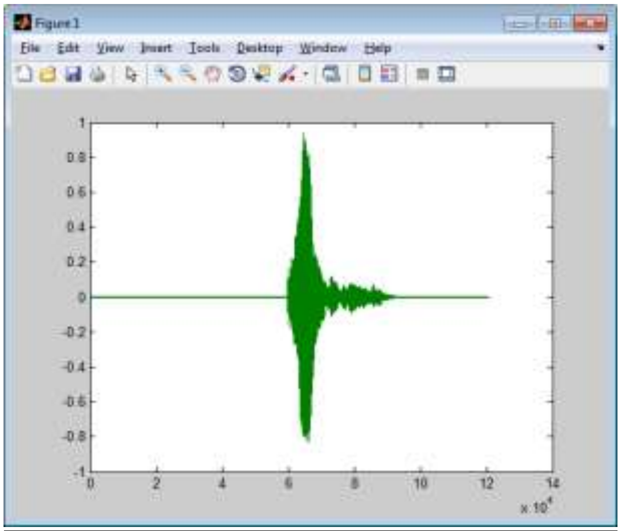

Fig. 3: Filtered Input Signal

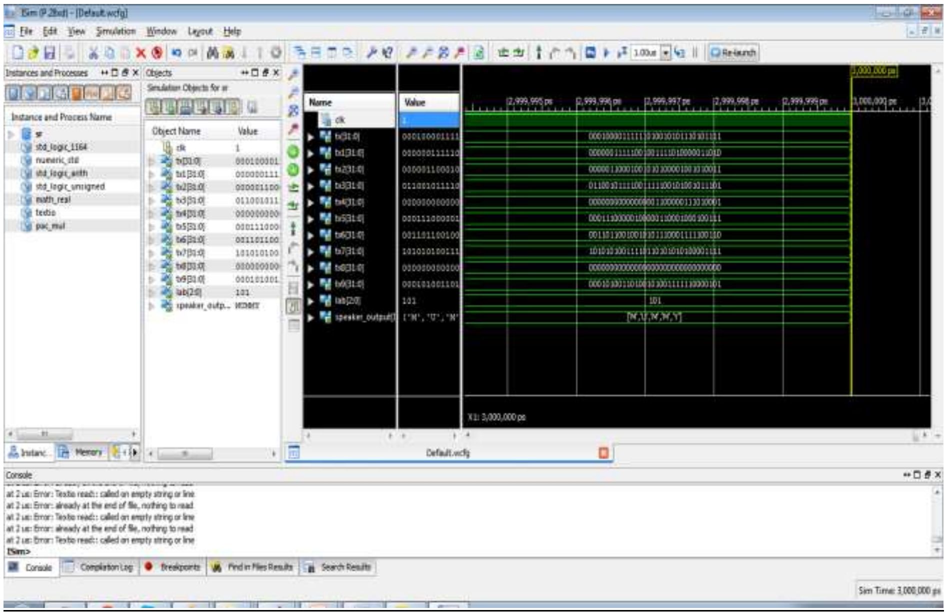

Fig. 4: Data write operation

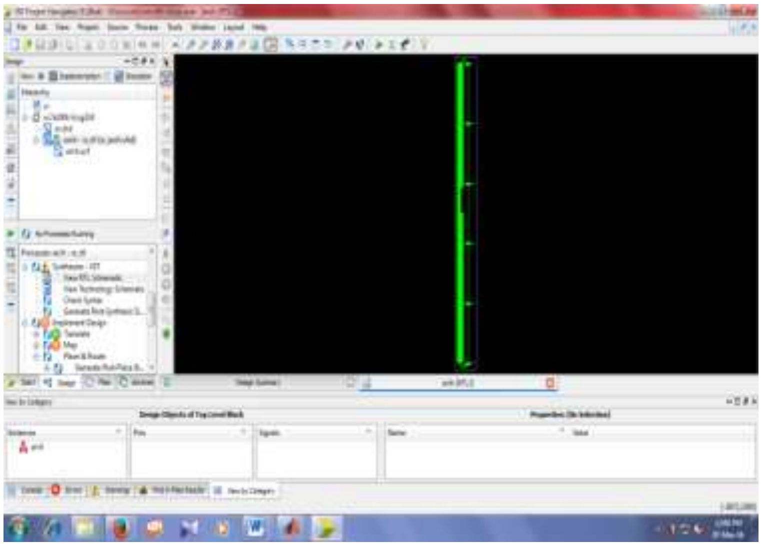

Fig. 5: RTL Diagram

Fig. 6: Power Estimation

Table - 1
Utilization Summary

PARAMETERS	*PROPOSED SYSTEM*
LUT'S COUNT	*17648*
FLIP FLOP'S COUNT	*8823*
MULTIPLIER COUNT	*20*
FREQUENCY(MHz)	*374.8*
POWER(mW)	*2240*

In our proposed system ,we reduced power and increase the frequency.so that the circuit complexity level will be reduced.Here the number of multiplier can be minimized.The multiplier count reduced by using approximate multiplier circuit.

Table - 2
Comparision Table

PARAMETERS	*EXISTING SYSTEM*	*PROPOSED SYSTEM*
LUT'S COUNT	*69120*	*17648*
FLIP FLOP'S COUNT	*69120*	*8823*
MULTIPLIER COUNT	*64*	*20*
FREQUENCY(MHz)	*100*	*374.8*
POWER(mW)	*4000*	*2240*

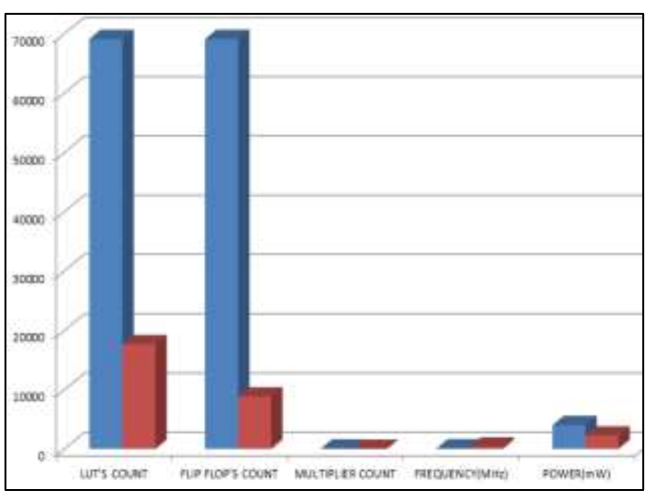

Fig. 7: Performance Graph

V. CONCLUSION AND FUTURE WORK

This proposed work provides promising result than existing system which concentrates on reduction of power, thus better performance is achieved than the existing design of SVM module with exponential processing element architecture. This proposed work uses normal processing element with approximate multiplier that reduces circuit complexity. This project aims at efficient SVM-based speaker verification for deaf and dumb people. The proposed work can be extended for image classification and it will be implemented in real time FPGA Spartan 3E Processor.

REFERENCES

[1] Clarkson.T.G, C. C. Christodoulou, Y. Guan, D. Gorse, D. A. Romano-Critchley, and J. G. Taylor, Speaker identification for security systems using reinforcement-trained pRAM neural network architectures,IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 31, no. 1, pp. 65–76, Feb. 2001
[2] Campbell .W.M , D. E. Sturim, and D. A. Reynolds, ―Support vector machines using GMM super vectors for speaker verification,‖IEEE Signal Process. Lett., vol. 13, no. 5, pp. 308–311, May 2006.
[3] Doddington.G.R, M. A. Przybocki, A. F. Martin, and D. A. Reynolds The NIST speaker recognition evaluation—Overview, methodology, systems, results, perspective IEEE Signal Process. Mag., vol. 31, no.2- 3, pp. 225–254, Jun 2000.
[4] A. Ganapathiraju, J. E. Hamaker, and J. Picone, "Applications of support vector machines to speech recognition,"IEEE Trans. Signal Process., vol. 52, no. 8, pp. 2348–2355, Aug. 2004.
[5] Juang.B.H and T. H. Chen, ―The past, present, and future of speech processing, IEEE Signal Process. Mag., vol. 15, no. 3, pp. 24–48, May 1998.
[6] Jiang.H and L. Deng, ―A Bayesian approach to the verification problem: Applications to speaker verification,IEEE Trans. Speech Audio Process., vol. 9, no. 8, pp. 874–884, Nov. 2001.
[7] Jia-Ching Wang, Li-Xun Lian, Yan-Yu Lin, and Jia-Hao Zhao ―VLSI Design for SVM-Based Speaker Verification System IEEE Transactions on very large scale integration (vlsi) systems, vol. 23, no. 7, july 2015.
[8] Li.Q ―A detection approach to search-space reduction for HMM state alignment in speaker verification IEEE Trans. Speech Audio Process., vol. 9, no. 5 pp. 569–578, Jul 2011.
[9] Pellom.B.L and J. H. L. Hansen, ―An efficient scoring algorithm for Gaussian mixture model based speaker identification, IEEE Signal Process. Lett., vol. 5, no. 11, pp. 281–284, Nov. 1998.
[10] Reynolds.D.A and R. C. Rose, ―Robust text-independent speaker identification using Gaussian mixture speaker models,‖ IEEE Trans. Speech Audio Process., vol. 3, no. 1, pp. 72–83, Jan. 1995.
[11] Wang .J.C C. H. Yang, J. F. Wang, and H. P. Lee, ―Robust speaker identification and verification,‖ IEEE Comput. Intell. Mag., vol. 2, no. 2, pp. 52–59, May 2007.
[12] T. W. Kuan, J. F. Wang, J. C. Wang, P. C. Lin, and G. H. Gu, "VLSI design of an SVM learning core on sequential minimal optimization Algorithm,"IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 20, no. 4, pp. 673–683, Apr. 2012.
[13] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," Comput. Speech Lang., vol. 20, nos. 2–3, pp. 210–229, 2006.
[14] Xiang. B and T. Berger, ―Efficient text-independent speaker verification with structural Gaussian mixture models and neural network, IEEE Trans. Speech Audio Process., vol. 11, no. 5, pp. 447–456, Sep. 2003.