Pak. J. Biotechnol. Vol. 13 (special issue on Innovations in information Embedded and Communication Systems) Pp. 121 - 124 (2016)

LOW POWER, SMALL FOOT PRINT EMBEDDED VOICE BIOMETRICS SYSTEM

[1]Puja Ramesh Chaudhari and [2]John Sahaya Rani Alex

School of Electronics Engineering, VIT University, Chennai Campus, Vandalur-Kelambakkam Road ,Chennai, India. [1]chaudharipuja7@gmail.com; [2]jsranialex@vit.ac.in

ABSTRACT

Biometrics is indeed becoming an important solution for any highly secured system. Voice is one of the biometric parameters that can use for a person identification and verification. In this paper, a small foot-print, low power embedded system is proposed and implemented using Beagle Bone Black (BBB). Hidden Markov Model (HMM) based speaker recognition system is implemented. Mel-Frequency Cepstrum Coefficients (MFCC) is used as features to identify the speaker. Each speaker is modelled as one HMM. The verification of the speaker voice is done using Viterbi decoder. The embedded system for Voice Biometric system is successfully implemented for a limited number of speakers and the accuracy is verified to be as almost 100%.

Keywords—Speaker Recognition, MFCC, Biometrics, Voice, HMM.

I. INTRODUCTION

II. In today's scientific world, transaction need to happen in a secured environment, the application based on the biometrics includes workstation and network access, single sign-on, application logon, data protecttion, remote access to resources, transaction security and web security purpose. Also human machine interfaces are becoming ubiquitous. For that purpose, the human has to authenticate to machines either through physiological (Ex:-Fingerprints, Face geometry, iris patterns and hand geometry) or through behavioral (Ex:- signature recognition)characteristics. The human voice is physiological characteristics[1] The biometric is the intrinsical property of a person. The customer requires a reliable authentication system. So, it is necessary to secure our transaction and network between man and machine from the potential attackers. The biometrics provides a robust connection between the individual and the advanced identity.

The speech is time varying non- stationary signal. The speaker voice is one of the biometric identities used for voice recognition. The person vocal tract and voice characteristic varies person by person so for that identify the person from his/her voice. Speech is a natural thing for the human communication purpose. Speaker recognition includes verification and identifycation. Automatic speaker verification, which uses a machine to verify a person identity from his/her voice. Speaker recognition system use Hidden Markov Model (HMM), Vector Quantization(VQ) based classifiers [2, 3,4] and the performance of the classifiers are compared [5]. The feature extraction of speech signal using Mel Frequency cepstral coefficients (MFCC) is done for speaker recognition system [2,6,7,8]. The MFCC is used for the collecting the characteristics of the speech signal. MFCC along with HMM provides the better result than Dynamic Time Warping (DTW). Recognition accuracy of the speaker recognition system is improved by having a modified MFCC method for verification along with VQ [9]. Speaker recognition system also implemented using neural network as classifier. Genetic Algorithm (GA) is used for optimazation of Radial Basis Function(RBF) network in implementing a speaker recognition system[10]. Gaussian Mixture Model (GMM) along with MFCC is also used for implementing speaker recognition system. All these implementations are done using MATLAB and a desktop. To enhance authentication in Automatic Teller Machine (ATM) machines, the hardware should be a small embedded system only for that purpose. Literature survey shows that Field Programmable Gate Array (FPGA) based speaker recognition system is implemented using MFCC,VQ [11]. Low power small embedded board is not used in literature to implement speaker recognition system. The objective of this paper is to implement a MFCC, HMM [12,13] based speaker recognition system on an embedded board. Beagle Bone Black is chosen because of its low power and small size. The rest of the paper is organized as follows; Section II gives the description of speaker recognition system. Section III gives the detail implementation processes of the MFCC and training part. Section IV gives the detail description of hardware. Implementation and result obtained are given in section V and conclusion is given in section VI.

II. SPEAKER RECOGNITION

Speaker recognition means the recognizing person based on human pitch. It depends on the speaker specific properties of the voice. The human speech contains different distinct features can be used for identifying the person. Human voice receives significant energy from zero frequency up to the 5 kHz. Speaker recognition parted into two parts: speaker verification and speaker identification. This is show in figure 1.

A. Speaker Verification

If the speaker requirement of absolute equality and the voice is used for verifying this requirement, this is defined as verification. In another term speaker verification is a 1:1 match where one speaker's voice is matched to one template.

B. Speaker Identification

Speaker identification is the term of determining an unknown speaker's identity. Speaker identification is a 1:M match where the voice is compared for M templates.

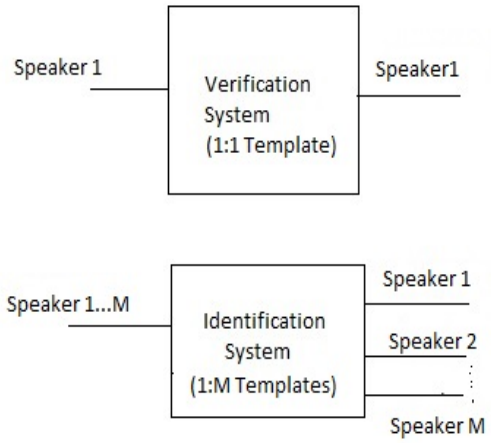

Figure1. Model of Speaker Recognition

Speaker recognition systems parted into two categories: text-dependent and text-independent.

A. Text Dependent

The text must be the same for enlistment and verify-cation this is called text-dependent recognition. In this case, the test comments are identical to the text used in the training phase. The test speaker has previous knowledge of the system.

B. Text Independent

Text-independent systems are most often used for speaker identification as they require very little if any collaboration by the speaker. In this scenario the text during enlistment is different. In fact, the enlistment may happen without the user's knowledge, as in the case for many forensic applications. As text-independent technologies do not compare what was said at enlistment and verification.

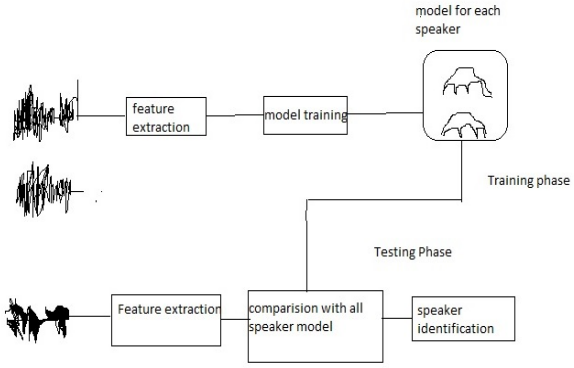

Figure2. Generalized Speaker Recognition System

Generalized speaker recognition system is shown in figure 2. Front end processing system takes speech samples from the speakers and features are extracted using feature extraction method, in this paper MFCC is used.. In the training phase, MFCC features are used to model the speaker. HMM of the speaker include statistical parameters such mean and variance. Using Viterbi decoding the trained HMM models are compared with unknown speaker features to recognize the speaker.

III. IMPLEMENTATION

A. Data Based Collection :

Speaker recognition system start with collecting the information from various speakers and recording by using the head mounted microphone for the various sampling rates such as 16 kHz, 44.1kHz and 48 kHz in .wav format.

B. Feature Extraction :

Certain attributes of the speaker is extracted from the speech signals of the speaker. It is basically represented the each speaker by extracting the small portion of data from the voice. Feature extraction for speaker must have some specific characteristics: it should be easily determinable from the set of known voices generated naturally and frequently in speech. It should be consis-tent for each speaker. Speech is the slowly time varying signal. When identified the voice signal over the short time period, then voice signal found to be stationary. However, when identified the signal characteristics tend to change over longer periods of time. This reflects the different speech utterances being spoken. Therefore, short-time spectral analysis is same way to differentiate a speech signal.

Many algorithms are used to extract speaker para-metric representation. They are Linear Predictive Cod-ing (LPC), Perceptual Linear Predictive (PLP), MFCC.

In this paper MFCC algorithm used for the feature extraction process.

1] *Mel Frequency Cepstral Coefficient (MFCC):*

MFCC takes human conception sensitivity with respect to frequencies into account which is good for speaker recognition. The MFCC basically performing the Fourier analysis depend on short–term power spec-trum. The MFCC spaced at low frequencies and logari-thmically at high frequencies for collecting important features of voice. The scale of Mel-frequency is the spectral analysis for linear frequency spacing below 1000Hz and for logarithmic frequency spacing above 1000Hz. Speech signal is framed and windowed for 25ms with an overlapping period of 10ms. Frequency component of the framed signal is passed through 24 triangular Mel scale filter banks. The filtered output is compressed using logarithmic and then the cepstral coefficients are decorrelated using Discrete Cosine Transform (DCT). The first 13 output of the DCT block is considered as static MFCC. From the Static MFCC, derivatives and double derivatives are calculated and used for speaker recognition. Figure 3 shows the block diagram of MFCC.

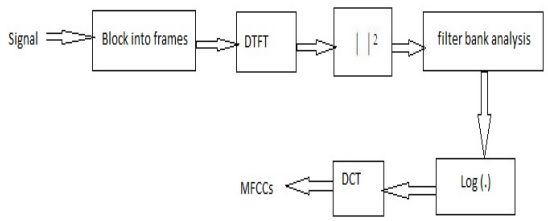

Figure 3: Feature Extraction of MFCC

C. Hidden Markov Model (HMM)

In the training phase, the HMMs are built using Hidden Markov Tool Kit (HTK) for every speaker. The observation is a probabilistic function of states. It depends on the acoustic model of various voices. For the output sequences derive the maximum various parameters of the HMM.

D. Viterbi Decoding

It is used for the decoding bit stream pattern of speech for the computing most probable path for the HMM module created for the each speaker output sequence. In this decoding checked for the comparing sequences of most probable path from the set of known speakers. The Viterbi Algorithm does a maximization step at every time point for the speaker recognition.

IV. HARDWARE

In this paper, Beagle Bone Black (BBB) embedded board is employed to implement speaker recognition. This is a low cost SitaraXAM3359AZCZ100 cortex A8 ARM processor, 1GHZ.

Figure: Beagle Bone Black (BBB) Board

A single 256Mb x16 DDR3L 4 GB (512MB) memory device is used. It will handle at a clock frequency of 303MHz. The board is outfitted with a single MicroSD connector to act as the secondary boot source for the board and, if selected as such, can be the primary boot source.

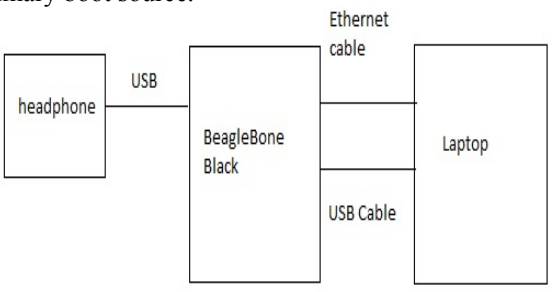

Figure 4: Board and Laptop connection

In this hardware interfacing as shown in figure 4, the cables connects from the Ethernet port of the laptop to the board RJ45 and then connect the mini USB cable provided to connect to the USB port of laptop for the power supply purpose. On Board USB port connect the headphone for recording the voice samples from the speakers. HTK is an open source software developed by Microsoft and Cambridge university. Viterbi decoder in HTK kit is licensed by Microsoft. All the feature extraction, Modeling and decoding is done using HTK.

V. IMPLEMENTATION AND RESULTS

The speakers are requested to read same set of documents for duration of 1sec. The audio file is stored .wav format. Using HTK, MFCC features are extracted and the speakers are modeled using HTK.

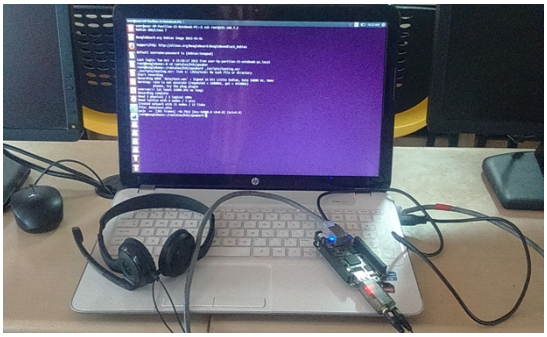

Figure 5: Speaker recognition output from the BBB

Among the speakers from the database one of the speakers is requested to say something for speaker identification and verification. Viterbi decoder takes the MFCC feature and did maximum likelihood probability matching among the speakers in the database to identify the unknown speaker. The result showed in figure 5 that the system recognized the person on the embedded board (BBB) with the 16 kHz sampling rate. The feature extraction by using MFCC method gives the accuracy nearer to 100%. This system implemented on board successfully.

VI. CONCLUSION

In this paper, we represented the approach for the speaker recognition on the embedded board BBB. Being a low-cost and a small footprint device, this system could be employed as voice biometric system along with PIN entry in ATM machines. It is advantageous for the real time application such as transaction authentication for the toll fraud prevention and telephone brokerage, for access control in physical facilities, computer and data networking also for remote time and attendance logging home parole verification and prison telephone usage monitoring system etc. Further this work could be tried with different feature extraction methods and also with different classifiers.

REFERENCES

[1] Kartik, R., P. Prasanna and S. Vara Prasad, Multimodal biometric person authentication system using speech and signature features. TENCON 2008 - 2008 IEEE Reg. 10 Conf. Pp. 1–6 (2008)

[2] Clerk Maxwell, J., A Treatise on Electricity and

Magnetism, 3rd ed., vol. 2. Oxford: Clarendon Pp.68–73 (1892).

[3] Martinez, M.M., J. Perez, H. Escamilla and E. Suzuki, Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques. Electr. Commun. Comput. (CONIELE COMP), 22nd Int. Conf. Pp. 248–251 (2012)

[4] S. L. Malode, S.L. and A.A.Sahare, An improved speaker recognition by using VQ & HMM. Sustain. Energy Intell. Syst. (SEISCON 2012), IET Chennai 3rd Int., pp. 1–7, 2012.

[5] Yorozu, Y., M. Hirano, K. Oka, and Y. Tagawa, Electron spectroscopy studies on magneto-optical media and plastic substrate interface. IEEE Transl. J. Magn. Japan **2**: 740–741 (1987).

[6] Tripathi, S. and S. Bhatnagar, Speaker Recognition, Comput. Commun. Technol. (ICCCT) Third Int. Conf. Pp. 283 – 287 (2012).

[7] Tiwari, V., MFCC and its applications in speaker recognition. Int. J. Emerg. Technol. **1**(1): 19–22 (2010).

[8] Xi Zhou, Yun Ful, Mark Hasegawa-johnson and Thomas S. Huang, Robust Analysis and Weighting on MFCC Components for Speech Recognition and Speaker Identification. ICME Pp.188-191 (2007).

[9] Ajinkya Sunil Mukhedhkar and John Sahaya Rani Alex, Robust Feature Extraction Methods for Speech Recognition In Noisy environments. First International Conference on Networks & Soft Computing (2014).

[10] Sumithra,M.G., K. Thanuskodi and A. H. J. Archana, A New Speaker Recognition System with Combined Feature Extraction Techniques Department of Electronics and Communication Engineering **7**(4): 459–465 (2011).

[11] Zhou, Y. and Y. GU, Human Speaker Recognition Based on the Integration of Genetic Algorithm and RBF Network. Second Int. Conf. Intell. Human-Machine Syst. Cybern. Pp. 239–242 (2010).

[12] Li Jinghong, Research and implementation of speaker recognition algorithm based on FPGA. Control Decis. Conf. (CCDC), 24th Chinese Control and Decision Conference (CCDC) Pp. 1155–1158 (2012).

[13] study of speaker recognition systems. [Online]. Available at http://ethesis.nitrkl.ac.in/2450/1/Project_ Report. pdf. [Accessed: 17-Feb-2015].

[14] www.ee.columbata.edu/ln/LabROSA/doc/HTKBook21/ node3.html orr speech.ee.ntu.edu.tw/homework/DSP_ HW2-1/htkbook.pdf

[15] Lawrence R. Rabiner, A Tutorial on Hidden Markov Models ans Selected Application in Speech Recognition, Proceeding of the IEEE Vol. 77, No.2, February (1989).