ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(24), DOI: 10.17485/ijst/2015/v8i24/80193, September 2015* ISSN (Online) : 0974-5645

Real Time Implementation of Speaker Verification System

K. Mohanaprasad[1*], Jeet Kiran Pawani[2], Vedant Killa[2] and S. Sankarganesh[1]

[1]School of Electronics Engineering (SENSE), VIT University, Vellore - 632014, Tamil Nadu, India; kmohanaprasad@vit.ac.in, s.sankarganesh@vit.ac.in

[2]ECE, School of Electronics Engineering (SENSE), VIT University, Vellore - 632014, Tamil Nadu, India; jeet_pawani12@yahoo.com, vedant.killa@gmail.com

Abstract

Verification of speakers is the use of the voice pattern to check the authenticity of the individual. In this paper the speaker verification system has been implemented on a real time DSP processor TMS3206713 with a new technique is proposed. Feature extraction and speaker modeling are the two major steps involved in speaker verification systems. Being the most frequently used features extracted from the human speech signal, the Mel Frequency Cepstral Coefficients (MFCC) have been considered for feature extraction. In the speaker modeling, a new technique known as Multi Section Vector Quantization (MSVQ) is implemented in addition to normal Vector Quantization. These techniques MFCC and VQ are generally used to recognize the speaker from a database of trained users. Using Multi Section Vector Quantization we can compare the trained codebook of particular user with the utterance and the authenticity decision can be made. Thus the advantage of using Multi Section Vector Quantization in speaker verification system is being discussed and the efficiency of this system is presented. To make this real time system robust in most of the environments a noise filter is also added during the pre-processing stage.

Keywords: Codebook Mel, Frequency Cepstral Coefficients, Multi Section Vector Quantization, TMS320C6713, Vector Quantization

1. Introduction

Speaker verification is used to make sure if a user is authorized to access a device. A person makes an identity claim (e.g., by uttering a password with his voice say on a personal laptop) and it has to match with the enrolled codebook. For security applications and crime investigations, speaker verification is one of the best and reliable techniques. The importance of choosing voice instead of other systems is that it cannot be forged (unlike face detection where an image can be used, finger print scan where synthetic polymer can be used with obtained prints and so on). There are two major steps involved to develop any speaker verification system. They are Feature extraction and Speaker modelling which is shown in Figure 1. Many characteristics can be extracted from the human speech signal of which Mel frequency Cepstral coefficients are used most often and are reliable. To develop a speech model for each user these features are to be quantized and codebook has to be generated. Each codebook represents a user in which the quantized features are stored and each user is assigned an ID to claim this codebook. These codebooks are stored in a database. The user has to make his claim and the voice input has to be given. The authenticity check is made with the utterance and the codebook for which he claimed to be the authorised person. In general the MFCC and VQ techniques are used for speaker recognition or identification, which is to identify the speaker from a database of trained codebooks. In speaker identification,

* *Author for correspondence*

the MFCCs of the utterance are calculated and the distortion between the MFCCs and all the codebooks are found. The minimum of which is treated as the identified user.

In this paper a system where the MFCCs are compared with the respective codebook for which he claimed to be the authorised person. The distortion between the MFCCs and codebook is determined. If the result is less than an optimum value (fixed based on the size of codebook, Mel filter bank size, windowing size, number of sections in MSVQ) the user is treated as authorised user. Thus this system can be used in many security applications and commercial applications where the accessibility to database is limited and specified. Using normal Vector quantization the threshold value cannot be fixed since it does not consider the time evolution of a signal. Speech signals are time dependent so multi section vector quantization is implemented which considers the time evolution of signals.

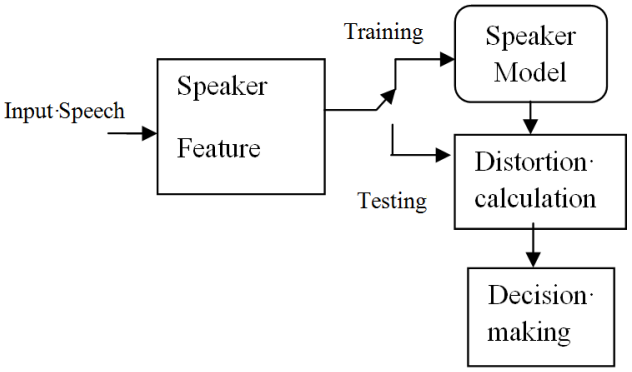

Figure 1. Block diagram of a typical speaker verification system.

Speaker verification and authentication system involves several stages. In this research work, each stage is elaborated and the results at each stage are being analyzed. Finally the efficiency of this system for authenticity check is calculated. As already told above, a noise filter is being added in the pre-processing stage to reduce the noise from the speech signal. This paper is structured as discussed hereafter. Analysis of development in speaker verification systems and DSP processors is demonstrated in section 2, section 3 gives a detailed account of Feature Extraction and Modeling of Speaker, proposed speaker verification system is introduced in the next section, Experimental Results are discussed in section 5 and Conclusion is briefed in section 6.

2. Review of Development in Speaker Verification and DSP Processors

There are several types of biometrics methods of authentication and identification. These include fingerprint method, recognition of face and iris, speaker verification, finger geometry and signature authentication[3]. Fingerprints of a person, including identical twins, are unique. It has therefore been used for a long time in law enforcement as a tool of biometric identification. Fingerprint systems can be used in both authentication and identification mode[3]. The biggest security issue with these systems is the vulnerability to artificial prints, "gummy fingers", which are cheap to produce and work well on both optical and capacitive sensors[5]. Face recognition uses facial images to identify subjects. There are several ways including the use of a normal camera using the visible spectrum or by the use of infrared cameras to capture the facial heat emission patterns. The biggest issues are detection of masks or photographs and handling the impact of lighting on the subject's face[3]. Iris recognition uses the unique features of a person's iris to identify them. Today's systems are accurate enough to work even in presence of eyeglasses or contact lenses. It works well for both authentication and identification purposes. Both face recognition and iris recognition have the advantage of not demanding any physical contact for recognition unlike fingerprint authentication systems[3].

Hand and finger geometry identification systems are similar to fingerprint systems but they measure physical characteristics like length, width, thickness and surface area instead[3]. Signature verification systems use a person's written signature for authentication. Speed, pressure and angle when the signature is produced are used to determine the validity of identity. Authentication systems like this are mostly used in e-business applications[3]. Speaker recognition uses the unique anatomy of an individual's throat and mouth to identify them by their voice[4]. The biggest advantage compared to all other biometric systems of identification and authentication is that it is the only biometric system that works over a telephone. Compared to other methods it's cheap because the major cost only comes from the software being used[1].

There are on the other hand a lot of problems to handle in speaker recognition. One of the biggest problems

is the variability produced by the talkers themselves and the variability produced by the transmitting and recording channels[6]. Sequence elementary acoustic sounds are as speech signal that characterize both the speech as well as the speaker. A large amount of data is generated in speech production while to represent the essential speech characteristics a small amount of data is required. Sufficient speech characteristics extraction in an amendable size and form is an important step in speech recognition for effective modellin[9]. While maintaining the discriminative information of a speaker voice signal[10], the feature extraction process reduces speech data. A comprehensive survey has been made by Picone[11] of signal modelling techniques in speech recognition. Picone concluded that satisfactory results for speaker recognition are provided by classical approach. Soria et al.[10], addressed this short coming by introducing a novel method called Mel-Frequency Cepstral Coefficients (MFC3) based on the cross correlation of MFCCs. A model comprising of joint probability functions of the feature vectors and the pitch was put forward by Ezzaidi et al. in[8]. Two pattern recognizers were used by them: Learning Vector Quantization with Single-Layer Perceptron (LVQ–SLP) and Gaussian Mixture Model (GMM). Their results showed an increase in identification rates. Li et al.[12] proposed a non-parameter procedure for speaker recognition which is dependent on Fisher Differentiation Vector (FDV). They concluded that this process is very efficient for text-reliant speaker identification. Nagorski et al. in[13] have presented a method based on Principal Component Analysis (PCA) to select limited and optimal speech data which represent maximally information for best training and testing of speech recognition systems. Nickel et al.[14] derived novel speech features from a PCA of speech segments, to improve the accuracy of text-dependent systems.

While performing PCA on the Mel Scale Spectrum Vector, Ding and Zhang[15] have proposed a different feature vector called Mel Frequency Principal Coefficient (MFPC). Correlated to how traditional MFCC was derived; the correlation information among different frequency channels was easily exploited by MFPC efficiently. The observation outcomes demonstrated that their suggested feature vector has peculiarity of closeness, lower redundancy and higher discriminability. Rosca et al.[16] introduced a speech synthesis model that deployed Independent Component Analysis (ICA) for text-independent speaker recognition. This method is robust to channel variability and invariant over time. In literature different techniques are available for feature extraction i.e. MFCC, LPC, Linear Predictive Cepstrum Coefficients (LPCC) and Perceptual Linear Prediction (PLP). MFCC are best features to express speech signal based on components of speech signal with low frequency[17].

Speaker modelling is a process which constructs a copy of speaker voice based on the features retrieved from speaker's speech sample in feature extraction step. There are two main approaches; Stochastic Modelling and Template for solving the classification problem. Template modelling is deterministic matching where training and testing data is compared using similarity measures. A probabilistic model of the speech signal is built in stochastic approach to describe its time-varying characteristics.

Pop in[19] has presented an approach for speaker verification task using single section Vector Quantization. As parameters they use LPC derived Cepstrum and MFCC. The results obtained in their experiments showed that the VQ method can be used for text-dependent speaker verification. Constantinou et al.[20] have also proposed a new type of VQ codebook design methods by introducing the concept of an Adjacency Map (AM). Fan and Rosca[21] have introduced heuristic weighted distance using a linear formula to bring up higher order MFCC feature vector factors. The experiments suggested that the new approach outperforms VQ-based solutions with 50% error education. Tunckanat et al.[22] have presented an approach based on neural networks for speaker recognition. The experimental results using text-dependent and text-independent recognition cases have been achieved 94% and 88% accuracies, respectively. The applicability of Probabilistic Neural Networks (PNNs) for medium scale speaker recognition was studied in[23] over fixed telephone networks. The authors presented two open-set text based self reliant systems based on PNN for speaker apperception and speaker authentication. The operation of repeated neural nets in an open-set text-reliant speaker identification job was addressed by Shahla and Philgreen in[24]. The objective was to find out recurrent neural net aptitude to take short-term spectral features. For 12 speakers the positive acceptance rate was upto 100% with a negative acceptance rate of 4% was achieved and for 16 speakers these rates were upto 94% and 7% respectively. Based on Template and Stochastic Modelling

approaches, feature modelling techniques are Vector Quantization (VQ), Nearest Neighbours (NN), Hidden Markov Model (HMM), Gaussian Mixture Modelling (GMM), Dynamic Time Warping (DTW), and Artificial Neural Networks (ANN). We selected VQ for speaker verification of our system.

Engineering and Science world is filled with different types of signals: images which were captured by remote space probes, the brain and heart's generated voltage, echoes by radar and sonar, seismic vibrations, and plethora of other uses. The science of utilizing computers to comprehend all forms of data is called Digital Signal Processing. This includes many different types of goals: recognition of speech, filtering, image enhancement, neural networks, data compression, and much more. DSP is one of the most powerful technologies that will shape the world's engineering and science in today's times[7].

Prior or even before to the discovery of individual DSP chips, most applications of DSP were executed or done with the bit slice processors. The 2920 was released by Intel as an "Analog signal processor" in 1978. It was not a great success in the market as it did not have a hardware but it came with an on-chip fabricated version of ADC/DAC with an inherent signal processor. The S2811 was released by AMI in 1979. It had to be initialized by the host and was designed as a microprocessor peripheral. The NEC μPD7720 and AT and T DSP1 – were introduced in 1980 which were the first standalone complete DSPs.

In 1983, Texas instruments launched its first DSP[8]. It had separate instruction and data memory and was based on the Harvard architecture. Thus, an exclusive instruction module was already present, with many directions like multiply-and-accumulate or load-and-accumulate. It needed around 390ns for a multiply-add operation and could work on less than or equal to 16-bit numbers. About five years later, the DSPs included hardware which accelerated tight loops and for storing two operands simultaneously had 3 memories, they also included an addressing unit capable of performing loop-addressing. This generation of DSPs began to spread widely. In the third generation, the main improvement was the application-specific unit's appearance and the data path instruction, or sometimes as coprocessors. Matrix operations and Fourier-transform are the simple but complex mathematical problems which were done by this unit using direct hardware acceleration. The modifications made in the instruction sets and the decoding/encoding of the instruction characterizes the fourth generation of DSPs in the best manner.

Better performance was yielded by modern signal processors. This is because of both architectural and technological improvements like lower design rules, (E) DMA circuit, fast-approach two-level cache and a broader bus system. As in real time signal processing the extra range given by floating point is not necessary, and there is a cost benefit and comparable speed benefit due to lesser hardware complexity, most DSPs use fixed-point arithmetic. In circumstances where a wide dynamic range is needed, floating point DSPs are considered invaluable. Floating point DSPs are used by the Product developers might help reduce the expenses and complexity of software enhancement in exchange for more expensive hardware, since it is relatively easier to implement algorithms in floating point. DSPs are dedicated integrated circuits; however the DSP operation is sometimes realized by Field Programmable Gate Array. RISC processors used for general purpose are becoming increasingly DSP like in function. A DSP clocks at 1.2Ghz which is Texas Instruments C6000 and implements different data instruction and caches along with a 9Mib which is a 2nd level cache, and I/O speed is rapid thanks to its 64 EDMA channel sits. The upper models use VLIW (Very Long Instruction Word) encoding are capable of executing as many as 8000 MIPS (Million Instructions Per Second), per clock-cycle eight operations are performed and are consistent with a wide range of various buses (PCI/serial/etc.) and external peripherals. Free scale, Analog Devices, and NXP Semiconductors are other major players in the market that manufacture high end DSPs.

3. Extraction of Features and Speaker Modelling

3.1 Extraction of Features

Extraction of features or feature extraction is referred as getting the acoustic based characteristics of the speech/input/speech signal. With the help of extraction of features both recognition as well as training process is pursued.

Below listed steps are followed in order:

- Blocking of frames.
- Windowing.
- Wiener Filtering.
- FFT (Fast Fourier Transform).
- Wrapping of Mel-Frequency.
- Cepstrum generation (Mel Frequency Cepstral Coefficients).

As per investigations characteristics of speech signals stay stationary for amply minute interval of time (quasi-stationary state). Thus processing of voice signals are done in minute time gap. It is fragmented into numerous frames with sizes generally amidst 25-100 milli-seconds. A predefined size is reserved for each frame for overlapping itself with the previous frame. To smoothen the transition from frame to frame is the goal of overlapping scheme[27].

The windowing of frames is the second step. This is done for elimination the edges discontinuities in the fragmented frames. If the windowing function is assigned as w (n), $0 < n < N - 1$, where N is the sample count in each of resulting frames, the output signal is found to be:

$$y(n) = x(n) * w(n) \quad (1)$$

Hamming windows are generally used which is shown in Figure 2. A Wiener filter which is used to provide an evaluation of a expected or targeted random process by linear time-invariant filtering an observed (obtained) noisy measure, assuming both the pre-known stationary (constant) signal and noise spectra, and additive noise. It reduces the mean square error between the desired action and the estimated random action.

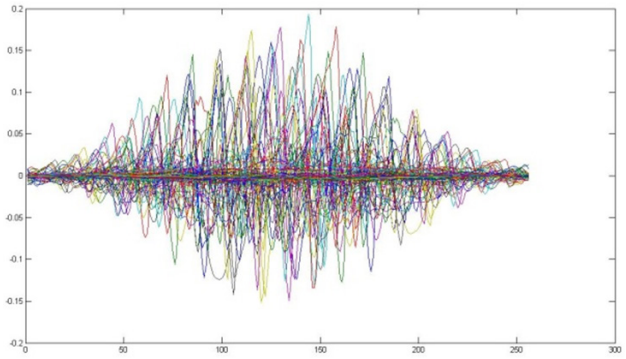

Figure 2. Signal after performing windowing.

Taking Fast Fourier Transform of each and every frame is the next step. To change a time respective domain signal to frequency respective domain signal is the base algorithm of Fourier transform, which is shown in Figure 3. This transformation changes the domain from time to frequency and is a quicker way of DFT (Discrete Fourier Transform)[27]. The reason why we are doing FFT is to simplify mathematical calculation as we know convolution in time respective domain is nothing but product in frequency domain.

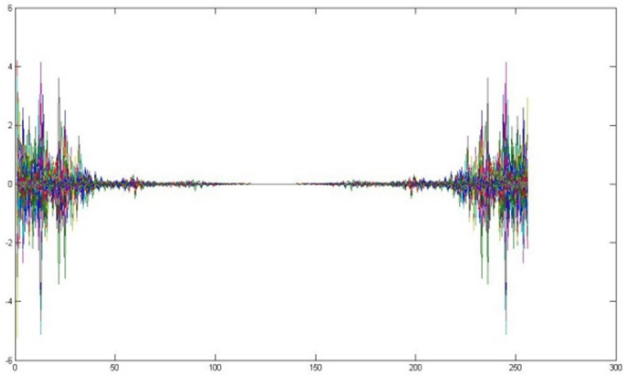

Figure 3. Signal after performing FFT.

The human ear non-linearly perceives the frequencies. The scaling is linear up to 1000 Hz and log scaled above that, according to researchers. The human ear perceiveness of frequencies is characterized by the Mel-Scale (melody scale) filter bank. For this stage of identification it is used as a band pass filter. Band pass filter which is Mel-scaled is used to mimic/copy the human ear. As mentioned above, perception of a human being to the contents of frequency sounds for voice signals doesn't follow a range which is linear[28] as per psychophysical studies. The "MEL" scale with an absolute frequency for each tone, f, calculated in Hertz, a pitch which is subjective is measured on a particular scale. The Mel-frequency scale is similar to human ear. A referral to the pitch of a 1000 Hz accent, 40-45 dB beyond hearing threshold intuitive, is defined to be as 1000 mels[28].

$$F_{mel} = 1125 \ln (1 + f / 700) \quad (2)$$

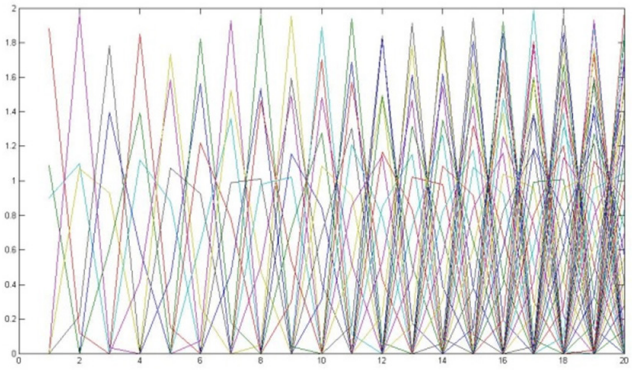

Figure 4. Mel-space filter bank (M = 20).

Using a filter based bank for simulation of the abstract spectrum is one approach, where one filter is used for each and every desired component of Mel-frequency. The frequency response of the filter based bank which is triangular band pass in nature and the Mel-frequency interval that is a constant which determines the spacing balance as well as the bandwidth which is shown in Figure 4. The counts of Mel cepstral coefficients, K, are chosen typically in the range of 20-25.

The final step is conversion of log Mel spectrum back to time domain. We get the bank of Mel Frequency Cepstral Coefficients (MFCC) as the output. A better representation (show) of the spectral properties which are local of the signal for the given frame analysis is given by the Cepstral representation of the speech spectrum[29]. As the coefficients of Mel spectrum and also their logs are real numbers, in turn they can convert to their time based domain using the Discrete Cosine Transformation (DCT). The Cepstrum is gained in a two-step process. A log scale power spectra is initially calculated and on that a DCT is performed. The real values of the real cepstrum are defined using the simple logarithm function whereas the cepstrum which is complex is defined using the complex logarithm function. The information of the magnitude is used by the real cepstrum of the spectrum whereas information about both phase and magnitude of the initial spectrum is held by the complex cepstrum, which in turn allows the reconstruction of the signal which was obtained. Finally we obtain coefficients which are namely Mel frequency Cepstral Coefficients (MFCC). Representation of audio based on perception is done by these coefficients. These coefficients have achieved a great success in speaker/speech recognition applications. The Fourier Transformation of the audio/voice clip helps give these very coefficients. The bands of frequencies are located logarithmically in this very technique, whereas in the case of Fourier Transformation the bands of frequencies are not located logarithmically as per what was seen in MFCC technique[29].

In MFCC the frequency bands are logarithmically positioned, the human based system response is approximated more clearly than any other system available. With the help of these extracted coefficients better examination of inputs is done. In the MFCCs the Mel Cepstrum's calculation is similar to that of the original Cepstrum besides that the warping of Mel Cepstrum's frequency scale is done in such a way that is used to align with correspondence to the Mel scale.

Finally Mel Frequency spectral coefficients were obtained for each speaker. But since it is practically not impossible to compare each and every coefficient while testing we need approximate the coefficients so for this we are using a technique called Vector Quantization technique where we will obtain codebook for each speaker where each codebook contains centroids which are approximations of the Mel Frequency coefficients.

3.2 Speaker Modelling

The Vector Quantization (VQ) paradigm also called as centroid paradigm is one of the simplest form of text independent speaker models. Its past is originally in data optimization and it was introduced in the 1980s for speaker recognition. Competitive accuracy is provided when it is combined with background model adaptation even though VQ is often used for practical implementations which are lightweight and computational speedup techniques[32].

In vector quantization, a particular user's huge set of feature vectors are taken and a minute set of feature vectors is produced which is duty-bounded to present the centroids of the distribution, i.e. the spacing of points is in such a way to reduce the approximate distance to each and every other point. Vector quantization is typically used each and every time for the representation of each and every single feature vector in the generated feature space from the training utterance/speaking of the corresponding/particular speaker is impractical. Plenty of time is saved in the testing phase itself as only a very less feature vectors of a particular user are considered in spite of the VQ algorithm taking time to generate the centroids. Therefore, we can live with such an economic compromise. K-dimensional vectors are mapped by vector quantizer in the space of the vector R_k into a set of finite vectors $Z = \{z_i: i = 1,2,.....N\}$.

K dimensioning refers to the ample amount of coefficient of the features in each of vector features. A code book is the set of multiple code words or all the vectors of codes, every code given by the vector z_i.

The advantages of VQ are:

- There is a reduction of memory for information on spectral analysis.
- The calculation for determination of commonness of spectral analysis vectors is also reduced. Estimation

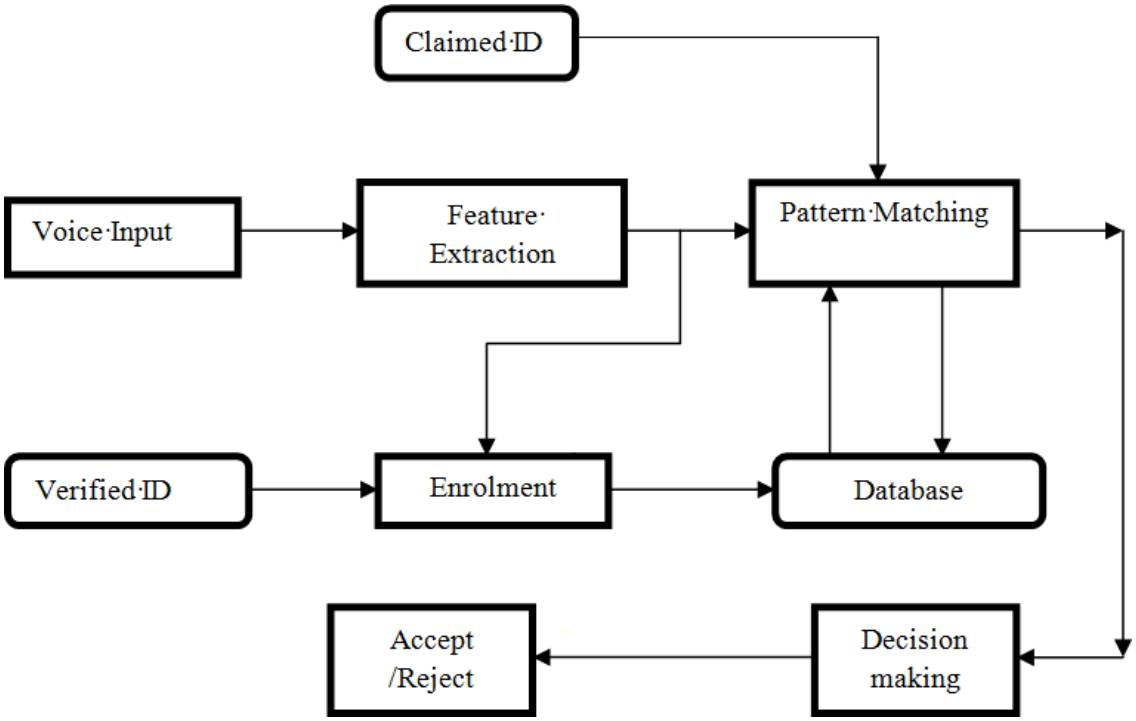

Figure 5. Proposed speaker verification system.

of spectral commonness amongst different pair of vectors in speech recognition is a major component. According to the VQ representation this is sometimes made into a table lookup of commonness amongst different pairs of codebook of vectors.

- Various accounts of sounds of speech.

4. Proposed Verification of Speaker System

The proposed verification of speaker system model is shown in Figure 4.

A group of trained vectors for a particular test speaker are extracted from the input speech of each speaker provided after the enrollment session. Building a genuine VQ codebook for every speaker using the trained vectors generated is the next most prominent task. A group of L trained vectors is clustered into a group of N codebook vectors using the LBG algorithm[31].

The steps that are formally executed by the following process[31]:

- Designing of a 1-vector codebook is done; it is the central element of the entire group of trained vectors and thus, iterations are not required.
- Codebook size is increased twice by dividing each current codebook by multiplying the centroid by (1+f) and (1-f) where 'f' is a splitting parameter.
- Nearest Neighbor Search: the needed code-word in the present codebook which is most similar (in forms of similarity measurement) to a particular code word is found and that vector is assigned to the particular cell (linked with the nearest code word).
- Central element Update: the code word in every cell is updated utilizing the central element of the training vectors allotted to that particular cell.
- Iteration 1: steps 3 and 4 are repeated until the mean distance drops lesser than a predetermined optimal level.
- Iteration 2: steps 2, 3 and 4 are repeated until a codebook vector of size of M is designed. An M-vector codebook is generated by the LBG algorithm iteratively. It initially starts by generating a 1-vector codebook, then uses a separation method on the codeword to begin the search for a multi-vector codebook and continues the separation procedure up till the required M-vector codebook is generated.

As said for the clustering of MFCCs we employ a clustering technique known as Vector Quantization. By clustering a codebook was obtained for each speaker with n number of centroids. But the problem with this vector quantization technique is that it is difficult to set a range for an authorized user during authentication check. Hence improvisation was done using multi-

section vector quantization where each utterance of a speaker will be split into M sections and codebook is obtained for each of these sections. Hence codebook of each trained speaker contains M sub-codebooks. Using this multi section technique the distortion gap between authorized user and unauthorized user get increased during authentication check hence it became easy to set the range for authentication check. Hence using this multi section vector quantization technique we can able avoids unauthorized activity into the system.

5. Research Outputs

In the research the algorithm being used is implemented in TMS320c6713 DSP processor. The DSK suits a wide variety of application environments since it contains a full complement of on-board devices. A Texas Instruments TMS320C6713 Digital Signal Processor which operates at 225 MHz with standard expansion connectors for daughter card use, interface/external emulator, single power supply voltage (+5V), an AIC23 stereo codec 512 Kilobytes of non-volatile/non-erasable Flash memory (256 Kilobytes usable in basic configuration), 16 Mbytes of synchronous DRAM, Program board configuration through registers executed in CPLD, 4 self-user accessible DIP switches and LEDs, reconfigurable boot-load options and JTAG emulation by on-board emulator of JTAG with the USB host is used. We designed this text dependent speaker verification system with Mel filter bank K value of 20 and total codebook size for each speaker is 24 with 3 sections each of size 8. The speech signal is sampled at 8000 Hz and 5 speakers are tested with ten different scenarios in security systems. The output characteristics are determined in terms of False Rejection Rates (FRR).

Table 1. False rejection ratio calculation for ten different T

	Authorized User	U1	U2	U3	U4	U5	FRR
T1	1	✓	x	x	x	x	100%
T2	3	x	✓	✓	x	x	75%
T3	2	x	✓	x	x	x	100%
T4	5	x	x	x	x	✓	100%
T5	4	x	x	x	✓	x	100%
T6	5	x	x	x	x	✓	100%
T7	3	x	x	✓	x	x	100%
T8	2	x	✓	x	x	x	100%
T9	4	x	x	x	✓	x	100%
T10	1	✓	x	x	x	x	100%

Here in the Table 1. T1 represents Test 1, U1 represents User 1. For the test T1 the assigned user is 1 which recognized correctly and FRR is 100%. For the Test T2 the assigned user is 3, but it is recognized as user 2 and user 3 with only 75% of FRR. Out of 10 tests only test 2 and test 3 went for 75% of FRR, whereas from the remaining test the FRR value is 100%.

6. Conclusion

In this paper Multi section vector quantization is implemented in a real time DSP Processor kit (TMS320c6713). The real time scenario works in all conditions which are checked with different locations as well as different people. The proposed techniques proved that the speaker is identified with improved accuracy and efficiency while compared with traditional MFCC and Vector quantization. This method which has been designed is efficient enough to get the proper output. Different test persons will have different coefficients in codebooks so comparison will be better and more real times efficient.

7. References

1. Joseph P, Campell JR. Speaker recognition: A tutorial. Proceedings of the IEEE; 1997.
2. Kinnunen T. Spectral features for automatic text-independent speaker recognition. 2003.
3. Podio FL, Dunn JS. Biometric authentication technology. 2001.
4. Furui S. Digital speech processing, synthesis and recognition. 1989.
5. Matsumoto T, Matsumoto H, Yamada K, Hoshino S. Impact of artificial 'Gummy' fingers on fingerprint systems. 2002.
6. Tosi O. Voice identification – Theory and legal applications. 1979.
7. Smith SW. The scientist and engineer's guide to digital signal processing. Available from: http://www.dspguide.com/whatdsp.htm
8. Lu X-C, Yin J-X, Hu W-P. A text-independent speaker recognition system based on probabilistic principle component analysis. 3rd International Conference on System Science, Engineering and Manufacturing Information; 2012 Oct 20-21; p. 255–60.
9. Ezzaidi H, Rouat J, O'Shaughnessy D. Combining pitch and MFCC for speaker recognition systems. A speaker odyssey - The speaker recognition workshop; 2001 June 18-22; Crete, Greece.
10. Soria RAB, Cabral EF. Combining neural networks paradigms and Mel-frequency cepstral coefficients correlations

in a speaker recognition task. Proceedings of the 7th International Conference on Signal Processing Applications and Technology; 1996. p. 1725–9.
11. Picone J. Signal modeling techniques in speech recognition. Proceedings of the IEEE; 1993 Sep; Texas Instruments Inc; p. 1215–47.
12. Li B, Liu WJ, Zhong QH. Text-dependent speaker identification using fisher differentiation vector. International Conference on Natural Language Processing and Knowledge Engineering, 2003; Beijing, China. p. 309–14.
13. Nagorski A, Boves L, Steeneken H. Optimal selection of speech data for automatic speech recognition systems. 7th International Conference on Spoken Language Processing; 2002 Sep 16-20; Denver, Colorado, USA.
14. Nickel RM, Oswal SP, Iyer AN. Robust speaker verification with principal pitch components. IEEE Transactions on Signal Processing.
15. Ding P, Zhang L. Speaker recognition using principal component analysis. Proceedings ICONIP; 2001.
16. Rosca J, Kofmehl A. Cepstrum-like ICA representations for text independent speaker recognition. 4th Int Independent Component Analysis and Blind Signal Separation; 2003 Apr 1-4; Nara, Japan.
17. Pop PG, Lupu E. Speaker verification with vector quantization. International Workshop Trends and Recent Achievements in Information Technology; 2002 May 16-18; Cluj Napoca, Romania.
18. Watkins D. Fundamentals of matrix computations. 2nd edition. Wellesley: Wiley-Interscience; 2002.
19. Pop PG, Lupu E. Speaker verification with vector quantization. International Workshop Trends and Recent Achievements in Information Technology; 2002 May 16-18; Cluj Napoca, Romania.
20. Constantinou AD, Bull DR, Canagarajah CN. A new class of VQ codebook design algorithms using adjacency maps. Proceedings of SPIE - The International Society for Optical Engineering; Bristol , UK: Image Communications Group; University of Bristol.
21. Fan N, Rosca J. Enhanced VQ-based algorithms for speech independent speaker identification. Proceedings of Audio- and Video-Based Biometric Authentication (AVBPA 2003); Princeton, New Jersey; Siemens Corporate Research Inc.
22. Tunckanat M, Kurban R, Sagoroglu S. Voice recognition based on neural networks. Kayseri, Turkiye: Department of Computer Engineering, Faculty of Engineering, Erciyes University.
23. Ganchev T, Tsopanoglou A, Fakotakis N, Kokkinakis G. Probabilistic neural networks combined with Gmms for speaker recognition over telephone channels. 14th International Conference on Digital Signal Processing; 2002 Jul 1-3; Santorini, Greece. p. 1081–4.
24. Parveen S, Philgreen. Speaker recognition with recurrent neural networks. 6th International Conference on Spoken Language Processing; 2000 Oct 16-20; Beijing, China.
25. Brookes M. MATLAB voice toolbox. The Math works. Available from: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.htm
26. Beigi H. Fundamentals of speaker recognition. 2011.
27. Proakis JG, Manolakis DG. Digit Signal Process. New Delhi: Prentice Hall of India; 2002.
28. Campbell JP. Speaker recognition: A tutorial. Proceedings of the IEEE. 1997 Sep; 85(9):1437–62.
29. Childers DG, Skinner DP, Kemerait RC. The cepstrum: A guide to processing. Proceedings of the IEEE. 1977 Oct; 65(10):1428–43.
30. Gray RM. Vector quantization. IEEE ASSP Magazine; 1984 Apr. p. 4–29.
31. Linde Y, Buzo A, Gray R. An algorithm for vector quantizer design. IEEE Trans Comm. 1980; 28:84–95.
32. Soong F, Rosenberg E, Juang B, Rabiner L. A vector quantization approach to speaker recognition. AT and T Technical Journal. 1987 Mar-Apr; 66:14–26.
33. Makhoul J. Linear prediction: A tutorial review. Proceedings of the IEEE. 1975; 64(4):561–80.