International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 13015-13021

Evaluation of Sparsification algorithm and Its Application in Speaker Recognition System

Satyanand Singh

Assistant Professor, Department of Electrical and Electronics Engineering, Fiji National University, Fiji, Island.

Orcid Id: 0000-0002-7707-031X

Abstract

This paper proposes spectral domain compression of the speech signal using novel sparsing algorithms. In a sparse algorithm, representation little quantity of coefficients holds a large proportion of the energy. Automatic Speaker Recognition (ASR) sparsity can play a major role to resolve big data issues in speech compression and its storage in the database, where the speech signal can be compressed before applying to ASR system and later can be used in speaker recognition. The Speech signal is converted to a spectral domain using Discrete Rajan Transform (DRT) and only first spectrum component has been retained forcing the remaining component to zero. The speech signal spectrum can be maximally compressed 8:1 ratio to the unique one. Spectrally compressed speech signal can be stored in the database and during training and testing time it can be synthesized using Inverse Discrete Rajan Transform (IDRT) in automatic speaker recognition. Acceptable speech signal spectral compression is 75% with Percentage of Identification Accuracy (PIA) of the speaker recognition system with sparsing is 95.3% and without sparsification 98.8% for TIMIT database respectively. In this paper the Novel Fuzzy Vector Quantization (NFVQ) feature matching technique was used, due to high accuracy.

Keywords: Vector Quantization, Fuzzy c-means Vector Quantization, Fuzzy Vector Quantization2, Novel Fuzzy Vector Quantization, Objective Function.

INTRODUCTION

Speaker recognition robustness in adverse condition has been investigated widely in recent years [1]. There are quite a number of factors affecting the automatic speaker recognition performance including channel/session variability and noise/reverberation. In real-world applications dealing with the mismatched condition is inevitable and any type of mismatch between training and test session will potentially result in degraded performance. Based on the type of the data in national institute of standards and technology (NIST) speaker recognition evaluations, the researchers in speaker recognition field have successfully developed techniques to deal with session/channel variability [2].

Although the state-of-the-art algorithms sensitivity to unseen channel or session variability is partially mitigated, they are highly vulnerable to additive noise and reverberant environment [3]. It has also been shown that even the performance of the state-of-the-art speaker recognition systems degrades substantially when limited speech is available in testing phase [3]. Although there are recent studies to handle reverberation and additive noise in feature and model domain for speaker recognition systems, the compensation techniques with respect to noise and reverberation for speaker recognition systems are still an open question.

Since our civilization, the speech is pure and natural means of human communication. Let us take an example for speaker recognition, a human recognizes a speaker regardless of the text spoken without of any effort for him/her to understand what exactly text spoken by different speakers. Human speech signal carries linguistic information as a major component as well as non-verbal information as a minor component. Based on speaker-specific features of acoustic speech signal a listener can identify his/her gender of the speaker, approximate age, and emotional state. In the human being, there is an effective way to automatically extract speaker-specific information from speech signals; the same concept has to be used in automatic speaker recognition by machine. The interference of redundancy in speech signal components hampers a speech signal or speaker recognition system performance [4].

The speech signal characteristically generated vocal tract which is a resonant system otherwise by physical impacts or occasionally together. Speech signal generated by vocal tract as resonant systems contain a number of redundant frequency components, therefore, if the speech signal which is going to be used in ASR is transformed into the spectral domain, a comparatively high degree of sparsity can be obtained. Taking into consideration that speech signal generated by physical impacts then it can be experimentally observed that the largest part of the speech signal is concentrated on time. This observation of speech characteristics permits superior sparse demonstration of the speech signal in time domain. In such kind of situation of the speech signal, Wavlets transform is most suitable for sparsification of speech signal. Hence, the perception of sparse representation and sparsity in speech processing and ASR is very effective [5].

The compressive sensing (CS) concept in sparsing can be utilized in a number of applications, particularly in speech signal processing that is speech pre-conditioning, SNR improvement, and speech coding [6]. Though sparsing is the latest technology, very little research has been done on the application of sparsing on speech signal and its utilization in ASR [7,8]. Speech is a common and pure natural way for communication among the persons however its processing is

troublesome in light of the fact that even on the off chance that we express the same word but can't create the same speech signal ever in all our years. In these manner, significant difficulties to apply sparsification in speech signal processing begins with, finding a good sparse basic and development most efficient measurement matrices [9].

Sparse Representation with Discrete Rajan Transform (DRT) and Inverse Discrete Rajan Transform (IDRT)

RT demonstrates a function ϕ:G→H is a homomorphism nature if for all g1, g2 in G,(g1, g2)ϕ = (g1)ϕ (g2)ϕ and it is transformation invariant in speech signal. Due to homomorphism nature, it has many applications in image processing like detection of the curve, detection of lines, detection of contour, detection of edge and image point isolation. If signal sequences are highly correlated then error in reconstructed signal is less and vice versa with the application of DRT. Due to the highly correlated non-stationary nature of speech signal, the DRT plays a very important task in terms of spectral sparsification, compression, and original speech signal reconstruction.

A U dimensional speech signal vector "d" can be represented as U=2u with u being a nonnegative integer. Consider a speech signal d(u), apply DRT on signal then spectrum D(r) can be obtained after u steps. The time domain speech signal can be converted into spectral domain with a unique operating matrix of dimension (U/2r-1X U/2r-1) denoted as Yr. This unique operation matrix construction is defined as;

$$Y_r = \begin{bmatrix} I_w & I_w \\ -e_r^1 . I_w & e_r^1 . I_w \end{bmatrix} \quad (1)$$

I_w indicate the wth order identity matrix. For example at r steps the order of identity matrix is Wr=U/2r: r∈{1, 2, .., n} and e1ϒ is the "supplementary information" which indicates the equilibrium state condition of the signal during spectrum generation. There will be a certain inherent phasor relation with 'supplementary information' eϒ between the sample points 1st and 5th.

$$e_r^i = \begin{cases} -1, & d_r^i(w_r + 1) < d_r^i(1) \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

Where i ={1, 2, .., 2r-1}. At every steps r , let Fr denoted as output sequence and it is obtained as:

$$F_r = Y_r D_r = [f_r^1 \quad f_r^2 \ldots \quad f_r^i] \quad (3)$$

In Eqn. (3) at every steps Fr has got 2r pr elements When r = 1 , D1 = d, at every step the equilibrium segments are considered for r > 1, 2r-1 .

Yr is the operator matrix can be constructed at a r stage using supplementary information er. Additionally if r > 1 then the output can be restructured in equilibrium segments and it can be defined as:

$$D_{r+1} = [\bar{d}_{r+1}^1 \quad \Upsilon_r^1 \cdot \bar{d}_{r+1}^2 \cdots \quad d_r^{i-1} \cdot \bar{d}_{r+1}^i] \quad (4)$$

where $\Upsilon_r = [\Upsilon_r^1 \quad \Upsilon_r^2 \cdots \quad \Upsilon_r^{i-1}]$ and also,

$$\Upsilon_k^{i-1} = e_r^1 \times e_r^i \text{ for } r > 1 \quad (5a)$$

$$D_{r+1} = \begin{bmatrix} f_r^i(1) & f_r^i(2) \cdots & f_r^i(p_r) \\ f_r^i(w_r + 1) & f(w_r + 2) & f_r^i(2w_r) \\ \vdots & \vdots & \vdots \\ f_r^i(2^{r-1}w_r + 1) & \ldots & f_r^i(2^r w_r) \end{bmatrix}^T$$

$$= [\bar{d}_{r+1}^1 \quad \bar{d}_{r+1}^2 \cdots \quad \bar{d}_{r+1}^i] \quad (5b)$$

Dr+1 express that signal spectrum into equilibrium segments. Steps will be continuing till the final DRT spectrum is obtained after u steps. As explained already, DRT is a homomorphic function and it also exhibits the isomorphism property when the complementary phasor information is preserved. Since DRT is also viewed as an isomorphic function, one should be able to retry the original signal data from its DRT spectrum by means of its inverse transform. Indeed, the IDRT is used to retrieve the input data with the help of e1k and μk. Now the DRT operator Rk is obtained using the values of e1k and μk . The general expression used to retrieve intermediate signal data at every stage is as follows.

$$\tilde{D}_t = \frac{1}{2} [Y_t F_t] = [d_t^1 \quad d_t^2 \ldots \quad d_t^i]^T \quad (6)$$

Where t = {r, r-1, .., 1}. As on account of forward DRT calculation wherein the succession is part into balance portions, on account of IDRT calculation, the sections are recombined and data arrangement recovered iteratively.

When a = r, Ft = Fr then we can obtained final stage spectral domain signal and for t < r,

$$F_{t-1} = [\bar{F}_t(1) \quad \Upsilon_r^1 \cdot \bar{F}_t(2) \cdots \quad \Upsilon_r^{i-1} \cdot \bar{F}_t(i)] \quad (7)$$

$$\bar{F}_{a-1} = \begin{bmatrix} d_a^i(1) & d_a^i(2w_r + 1) \cdots & d_a^i(2^{r-1}w_r + 1) \\ d_a^i(2) & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ d_a^i(2w_r) & d_a^i(2^2 w_r) & d_a^i(2^r w_r) \end{bmatrix} \quad (8)$$

During the IDRT computation the original input speech signal can be obtained.

DRT Sparsification, Compression and Decompression Application on Speech Signal

The experiments are performed on a speech signal is taken from TIMIT database. Speech signal of male and female taken for 3sec. The speech signal sampling frequency is 16 KHz. This experiment is conducted on MATLAB with i3 Intel Core Processor Clock Frequency at 2.53 GHz. The entire

speech signal is divided into a number of blocks. Every block contains 8 samples. Here, DRT will be applied to the speech signal, it will be Sparsified, compressed, stored and whenever the speech signal required IDRT will be applied to reconstruct the original speech signal.

Application of DRT on Speech Signal of 64 Sample Size

A 3 sec speech signal of female from TIMIT database has 62634 samples. Before applying DRT we need to take sample size which is divisible by 8. Let us take a sample of 48128 and divide it into 8X1 blocks. Now we have total 6016 number of blocks of size 8X1 and DRT applied on every block. A real-time speech signal d(u) of sample size 64 was taken and DRT is applied in block wise fashion, the corresponding spectrum of the blocks is obtained as D(r) .

For instance, let us consider a specimen real-time speech signal in discrete sequenced d(u) of length 64.

d(u)=0.1230, 0.1375, 0.1635, 0.1694, 0.1547, 0.1517, 0.1469, 0.1565, 0.1481, 0.1436, 0.1326, 0.1256, 0.1192, 0.1164, 0.1163, 0.0994, 0.0919, 0.0919, 0.0810, 0.0584, 0.0481, 0.03588, 0.0281, 0.01425, 0.0064, -0.0016, -0.0343, -0.0496, -0.0659, -0.0803, -0.1102, -0.1375, -0.1558, -0.1766, -0.1943, -0.2208, -0.2282, -0.2546, -0.2699, -0.2748, -0.2774,-0.2582, -0.2702, -0.2598, -0.2424, -0.2219, -0.2063, -0.1766, -0.1409, -0.1287, -0.1092, -0.0999, -0.0638, -0.0371, -0.0018, 0.0324, 0.0583, 0.0645, 0.0724, 0.0923, 0.0992, 0.1167, 0.1146, 0.1195.

Fig.1. (a) shows the plot of d(u) and (b) shows the plot of D(r) .

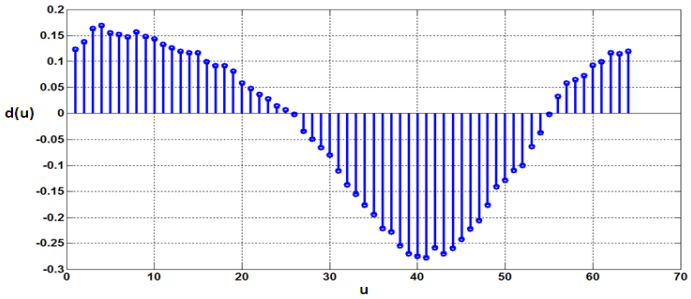

Figure 1: (a) Plot of d(u).

DRT is applied to d(u) in block-wise fashion and equivalent spectral blocks obtained as D(u).

D(u) = 1.2035, -0.0270, -0.0694, 0.0039, -0.0166, -0.0137, -0.0753, -0.0211, 1.0014, 0.0310, 0.0534, -0.0168, 0.0986, -0.0080, 0.0135, 0.0115, 0.4498, 0.0487, 0.0859, -0.0242, 0.1970, -0.0034, 0.0028, -0.0209, -0.4733, 0.0651, 0.1903, -0.0201, 0.3148, -0.0184, -0.0126, 0.0056, -1.7754, 0.0785, 0.1446, 0.0160, 0.2798, 0.0160, 0.0207, -0.0273, -1.9132, -0.0799, -0.0869, 0.0003, -0.2183, 0.0206, 0.0757, -0.0181, -0.5492, -0.0825, -0.1921, 0.0047, -0.4085, 0.0395, 0.0708, -0.0105, 0.7379, -0.0485, -0.0601, 0.0010, -0.1626, -0.0036, -0.0236, 0.0264.

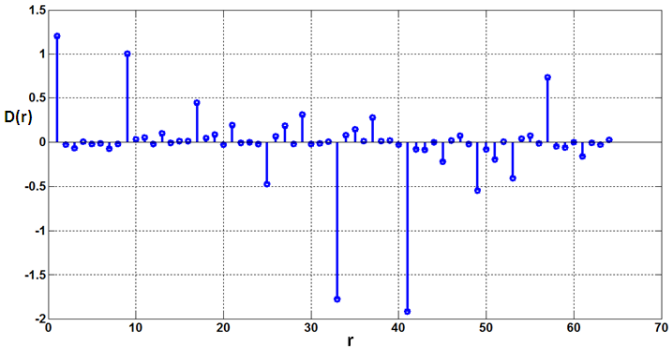

Figure 1:(b) Plot of D(r) the spectrum of d(u)

Sparsing of Speech Signal Data by Retaining CPI alone

D(r) is sparsed keeping the CPI alone in each block of length 8 and compelling remaining components to 0. At that point, the sparsed spectrum is Ds1(r).

Ds1(r) = 1.2036, 0, 0, 0, 0, 0, 0, 0, 1.0015, 0, 0, 0, 0, 0, 0, 0, 0.4500, 0, 0, 0, 0, 0, 0, 0, -0.4734, 0, 0, 0, 0, 0, 0, 0, -1.7755, 0, 0, 0, 0, 0, 0, 0, -1.9131, 0, 0, 0, 0, 0, 0, 0, -0.5493, 0, 0, 0, 0, 0, 0, 0, 0.7380, 0, 0, 0, 0, 0, 0, 0.

Compressing Speech Signal Data with CPI alone

Ds1(r) is the spectral domain sparsed speech data having only 8 non-zero elements. D's1(r) is the compressed version of Ds1(r) after ignoring all 56 samples of spectral components of zero values.

D's1(r) = 1.2036, 1.0015, 0.4500, -0.4734, -1.7755, -1.9131, -0.5493, 0.7380. D's1(r) can be stored in a database as a representative biometric vector of a speaker. The scale of compression and hence sparsity acquired by keeping the first component of the spectral is 12.5%.

Fig.2. (a) shows the plot of Ds1(r) the sparsed spectral sequence and (b) Compressed form of sparsed spectral sequence D's1(r).

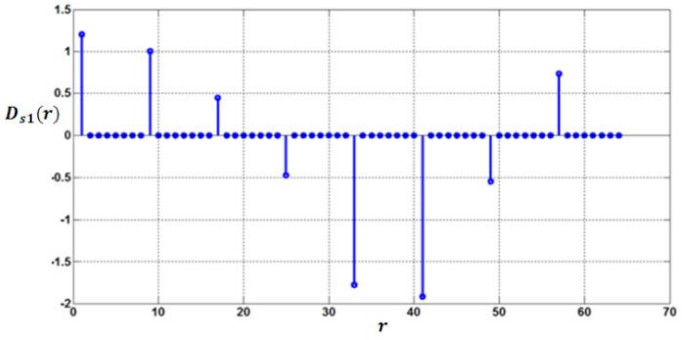

Figure 2: (a) Plot of Ds1(r) sequence .

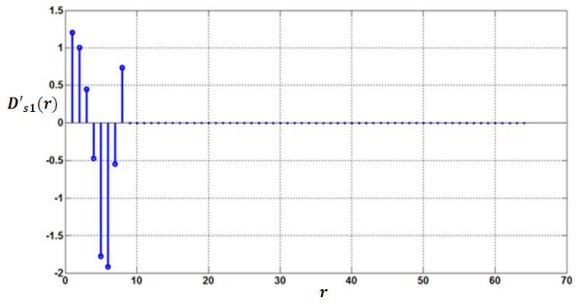

Figure 2: (b) Plot of D's1(r) sequence

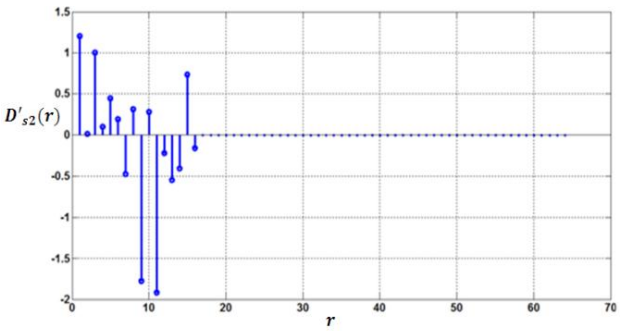

Figure 3: (b) Plot of D's2(r)

Sparsing of Speech Signal Data by Retaining CPI and mid Frequencies alone

In this case D(r) is sparsed by keeping the CPI and the mid frequency segment in each block of length 8 and driving remaining components to 0. At that point, the sparsed speech data sequence is Ds2(r).

Ds2(r) = 1.2036, 0, 0, 0, 0.0166, 0, 0, 0, 1.0015, 0, 0, 0, 0.0986, 0, 0, 0, 0.4500, 0, 0, 0, 0.1970, 0, 0, 0,

-0.4734, 0, 0, 0, 0.3148, 0, 0, 0, -1.7755, 0, 0, 0, 0.2798, 0, 0, 0, -1.9131, 0, 0, 0, -0.2183, 0, 0, 0, -0.5493, 0, 0, 0, -0.4085, 0, 0, 0, 0.7380, 0, 0, 0, -0.1625, 0, 0, 0.

Compressing Speech Signal Data with CPI and mid Frequency alone

Ds2(r) is the spectral domain sparsed speech data having only 16 non-zero elements. D's2(r) is the compressed version of Ds2(r) after ignoring all 48 samples of spectral components of zero values.

D's2(r) = 1.2036, 0.0166, 1.0015, 0.0986, 0.4500, 0.1970, -0.4734, 0.3148, -1.7755, 0.2798, -1.9131,

-0.2183, -0.5493, -0.4085, 0.7380, -0.1625.

D's2(r) can be stored in a database as a representative biometric vector of a speaker. The scale of compression and hence sparsity acquired by keeping the first and mid frequency component of the spectral is 25%.

Fig.3. (a) shows the plot of Ds2(r) the sparsed spectral sequence with CPI and mid frequency component and (b) Compressed form of sparsed spectral sequence D's2(r).

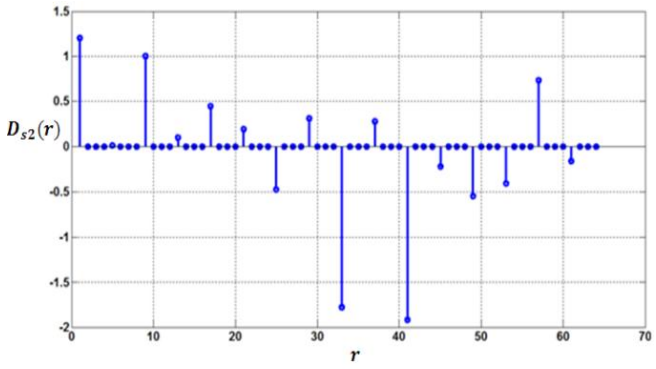

Figure 3: (a) Plot of Ds2(r)

Decompressing of the Speech Signal from D's1(r) and D's2(r)

Amid the speaker recognition testing stage, D's1(r) and D's2(r) is uncompressed to acquire Ds1(r), Ds2(r) in the case of 12.5% and 25% of sparsity respectively. Presently, IDRT algorithm applied to Ds2(r) and Ds2(r) to reconstruct the speech signal which can be used during testing and training phase of ASR. Here we can rename the reconstructed speech signal as d'1(u) and d'2(u) respectively.

d'1(u) = 0.1504, 0.1504, 0.1504, 0.1504, 0.1504, 0.1504, 0.1504, 0.1504, 0.1251, 0.1251, 0.1251, 0.1251, 0.1251, 0.1251, 0.1251, 0.1251, 0.0562, 0.0562, 0.0562, 0.0562, 0.0562, 0.0562, 0.0562, 0.0562, -0.0591, -0.0591, -0.0591, -0.0591, -0.0591, -0.0591, -0.0591, -0.0591, -0.0591,-0.2219, -0.2219, -0.2219, -0.2219, -0.2219, -0.2219, -0.2219, 0.2219, -0.2219, -0.2391, -0.2391, -0.2391, -0.2391, -0.2391, -0.2391, -0.2391, -0.2391, -0.2391, -0.2391, -0.0686, -0.0686, -0.0686, -0.0686, -0.0686, -0.0686, -0.0686, -0.0686, 0.0922, 0.0922, 0.0922, 0.0922, 0.0922, 0.0922, 0.0922, 0.0922.

d'2(u) = 0.1483, 0.1483, 0.1483, 0.1483, 0.1525, 0.1525, 0.1525, 0.1525, 0.1375, 0.1375, 0.1375, 0.1375, 0.1128, 0.1128, 0.1128, 0.1128, 0.0808, 0.0808, 0.0808, 0.0808, 0.0316, 0.0316, 0.0316, 0.0316,-0.0198, -0.0198, -0.0198, -0.0198, -0.0985, -0.0985, -0.0985, -0.0985, -0.18695, -0.18695, -0.1869, -0.1869, -0.2569, -0.2569, -0.2569,-0.2569, -0.2664, -0.2664, -0.2664, -0.2664, -0.2118, -0.2118, -0.2118, -0.2118, -0.1197, -0.1197, -0.1197, -0.1197, -0.0175, -0.0175, -0.0175, -0.01758, 0.0719, 0.0719, 0.0719, 0.0719, 0.1125, 0.1125, 0.1125, 0.1125.

Fig. 4. (a) shows d(u) and d'1(u) represented in the same plot.
Fig. 4. (b) shows d(u) and d'2(u) represented in the same plot.

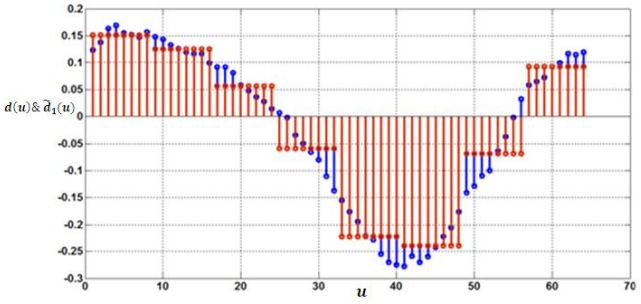

Figure 4: (a) Plot of d(u) and d'1(u)

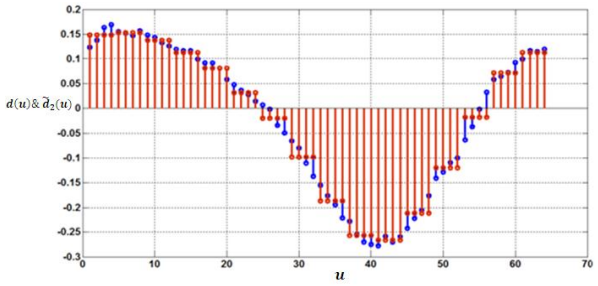

Figure 4: (b) Plot of d(u) and d'2(u)

Fig.5. (a) shows d(u) and d'1(u) exhibited in the same plot of 3 sec of speech data. Likewise, D(r) is sparsed by holding CPI values alone recurrence parts of all the 6016 blocks. Because of this sparsing, 12032 unearthly values would involve 12.5% of the real memory dispensed to oblige 48128 examples. Fig. 5. (b) indicates d(u) and d'2(u), the discourse signal reproduced from 25% of speech data. Reconstructed speech signal d'2(u) of a 3 sec of speech signal of a speaker from the sparsified information keeping the CPI and the mid frequency from each of block of length 8 and constraining different components to 0, that is D's2(r) is particularly closer to the original speech signal d(u).

Speech signal can be represented in so many different ways. The simplest way of representation is the plot of samples amplitude vs time waveform. In the universal representation of speech signal, the samples amplitude is normalized to reside between −1 and 1. All speech signal processing applications always used 16 bits/sample as bit resolution. The total number of quantization levels will in this manner be 2^{16} = 65536 and are observed to be ideal for saving speech data present in the simple form of the speech signal.

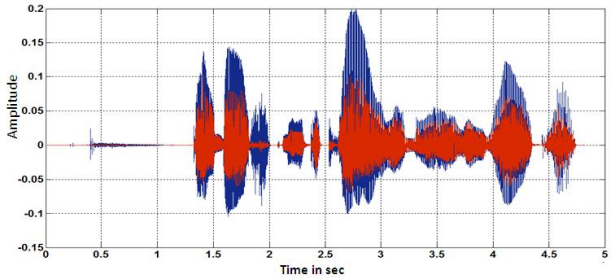

Figure.5: (a) Speech signal d(u) and d'1(u)

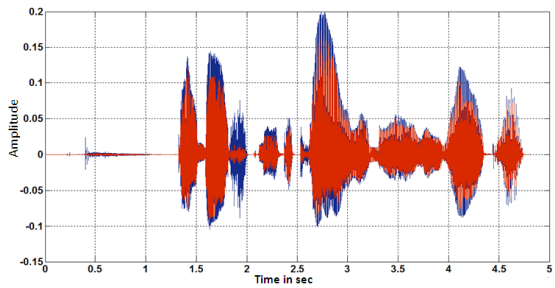

Figure 5: (b) Speech signal d(u) and d'2(u)

Error Dynamic Range (EDR) because of remaking from 25% of voice information is little when contrasted with the EDR because of reproduction from 12.5% of voice information and henceforth the previous is superior to the last in speaker recognition application.

Performance of Quality Measurement Matrix

Following three different performance parameters are used to measure the quality of reconstructed speech. Here we have measured the performance of d'1(u)and d'2(u) with reference signal d(u) [10,11].

The Mean Squared Error (MSE) is apparently the essential paradigm used to assess the quality of the reconstructed signal. The 3 sec original of speech signal and IDRT synthesized speech signal with "u" range time index covering the measurement intervals, then the MSE is defined as:

$$\text{MSE} = \sum_u \frac{(d(u)-\tilde{d}_1(u))^2}{u} \qquad (9)$$

In digital speech processing MSE represents the quantity by which IDRT reconstructed speech signal fluctuates from the original speech.

Signal-to-noise ratio is defined as the ratio of the power of an original speech signal and the power of the error signal and mathematically defined as:

$$\text{SNR} = 10\log_{10}\left(\frac{\sum_u (d(u))^2}{\sum_u (d(u)-\tilde{d}_1(u))^2}\right) \qquad (10)$$

Table 5. and Table 6. show comparison of mean square Error (MSE) and signal to noise Ratio (SNR) of DRT, DFT, DCT, DWT for 100 different speeches randomly selected from TIMIT database.

Table 5: MSE of speech signal after applying different transform

Sparsification by retaining	MSE DRT	MSE DFT	MSE DCT	MSE DWT
Only CPI	0.001900	0.036900	0.014300	0.015400
CPI +Mid frequency	0.000647	0.034700	0.012200	0.011500

Table 6: SNR of speech signal after applying different transform

Sparsification by retaining	SNR DRT (dB)	SNR DFT (dB)	SNRDCT (dB)	SNR DWT (dB)
Only CPI	25.47	10.9125	15.03	14.7028
CPI +Mid frequency	27.39	11.7032	17.97	15.9656

MSE of DRT is least and SNR is more, therefore DRT is suitable for sparsification of the speech signal in ASR application.

PESQ is a generally utilized, upgraded perceptual estimation for voice quality in information transfers. By and large, speech quality appraisal can be categorized as one of two classifications; subjective and objective quality measures. Subjective quality measures depend on the examination of original and reconstructed speech signal by an audience or a board of audience members. The scope of PESQ varies from 0.5 to 4.5, with the lower values interpreted as poor speech quality.

It is observed that for the case of 25% data reduction the PESQ of the reconstructed speech data does not deviate so much from the standard value that is 3.2331. Indeed, for the case of 12.5% data reduction the PESQ of the reconstructed speech data deviates considerably from the standard value that is 2.1543.

EXPERIMENTAL RESULTS

Experimental assessment of the DRT sparsification algorithms continued with the 100 speakers from TIMIT database. The TIMIT database contains voice 630 individuals that are part in subsets as indicated by the Dialect Region to which each of them has a place. Each DR has been used in sparsification as well as to train and test ASR. Log-spectral vectors were extracted from a 30-dimensional Gaussian Mel-filter bank and modelling based on FCM, FVQ2 and NFVQ.

Table 7. Shows the PIA of TIMIT database for FCM, FVQ2 and NFVQ techniques with sparsification and without sparsification respectively.

Table 7: Efficiency of Speaker Recognition of TIMIT Database

PIA	FCM	FVQ2	NFVQ
without Sparse	98.1	98.3	98.8
With Sparse	94.1	94.5	95.3

Performance Evaluation of Speaker Recognition Systems

With a specific end goal to check the performance of NFVQ algorithms based ASR, we processed the real match scores with sparsification and without sparsification with the impostor match scores. The Detection Error Trade-off (DET) of every experiment has appeared in the accompanying figures. NFVQ Algorithm performance with TIMIT database an EER value of about 8 % and with sparsified speech the EER of about 13%. Fig. 6. (a) compares ASR system performance in three different modalities: that is FCM, FVQ2 and NFVQ modelling techniques without sparsification and Fig.6. (b) with sparsification.

Based on the analysis of the three DET curves of Fig. 6. (a) and (b) it is clear that by employing the sparsification algorithms we can compress the signal spectrum up to 25% without degrading much more the system performance of ASR.

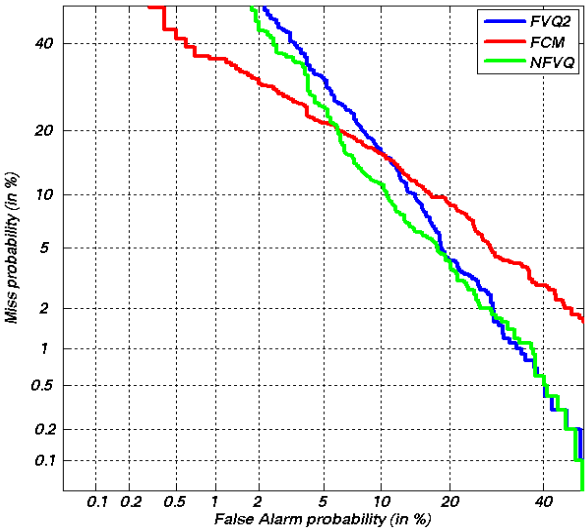

Figure 6: (a) DET curve of ASR with original speech signal

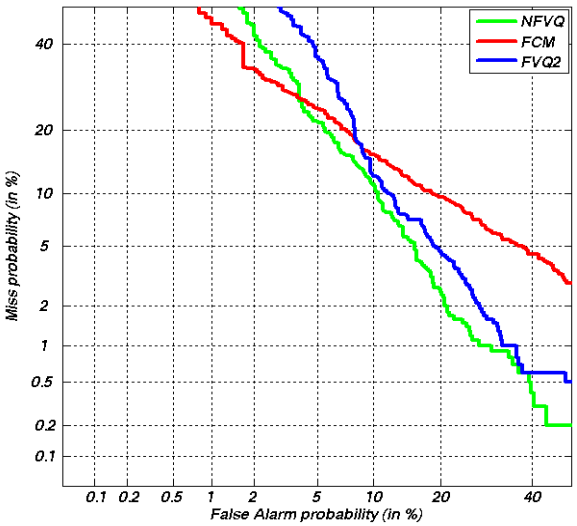

Figure 6: (b) DET curve of ASR with synthesize speech signal

CONCLUSIONS

The performance of speaker recognition accuracy of the ASR system proposed in this paper is 98.8% without sparsification with theTIMIT database. Highest attainable spectral compression of speech signal was about 75% and the performance of identification accuracy of the ASR system 95.3% and 93.5% respectively. A worthy level of trade-off between the performance identification accuracy of the framework and the capacity of the memory can be made according to the end client necessities, thus making the proposed framework strong, exact and adaptable. This framework has potential applications in common and military administrations where enormous database is required for putting away voice tests and train samples. Further research can be centered on enhancing the performance of identification accuracy of the automatic speaker recognition

systems by presenting more powerful sparsification procedures to lessen the trade-off element. Likewise, the impact of presenting efficient various leveled database in the framework can be examined.

REFERENCES

[1] S Singh, "Forensic and Automatic Speaker Recognition System," International Journal of Electrical and Computer Engineering, vol. 8, no 5, pp.2804-2811, 2018.

[2] S.Singh, Mansour. H. Assaf and Abhay Kumar, "A Novel Algorithm of Sparse Representations for Speech Compression/Enhancement and Its Application in Speaker Recognition System," International Journal of Computational and Applied Mathematics, vol. 11, no. 1, pp. 89-104, 2016.

[3] M. I. Mandasari, M. McLaren, and D. A. van Leeuwen, "The effect of noise on modern automatic speaker recognition systems", In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2012), 2012.

[4] Sadaoki Furui, "50 Years of Progress in Speech and Speaker Recognition Research", ECTI Transactions on Computer and Information Technology, Vol.1, No.2, pp 64-74, 2005.

[5] Plumbley M. D., Blumensath T., Daudet L., Gribonval R., Davies M. E. "Sparse Representations in Audio and Music: from Coding to Source Separation", In Proceedings of IEEE, vol. 98(6), pp 995-1005, 2010.

[6] Y. Wang, Z. Xu, G. Li, L. Chang and C. Hong, "Compressive Sensing Framework for Speech Signal Synthesis Using a Hybrid Dictionary", 4 th International Congress on Image and Signal Processing (CISP), Shanghai, vol. 5, pp. 2400-2403, 2011.

[7] S.Singh and Ajeet Singh "Accuracy Comparison using Different Modeling Techniques under Limited Speech Data of Speaker Recognition Systems" Global Journal of Science Frontier Research: F Mathematics and Decision Sciences, vol. 16, Issue 2, pp.1-17, 2016.

[8] S.Singh and Dr. E.G. Rajan "Application Of Different Filters In Mel Frequency Cepstral Coefficients Feature Extraction And Fuzzy Vector Quantization Approach In Speaker Recognition" International Journal of Engineering Research & Technology, vol. 2 Issue 6, 2013.

[9] Desai Siddhi and Nakrani Naitik, "Improved Performance of Compressive Sensing for Speech Signal with Orthogonal Symmetric Toeplitz Matrix", International Journal of Signal Processing, Image Processing and Pattern Recognition vol.7, no.4, pp.371-380, 2014.

[10] S.Singh. "Support Vector Machine Based Approaches For Real Time Automatic Speaker Recognition System," International Journal of Applied Engineering Research, vol. 13, no. 10, pp. 8561-8567, 2018.

[11] S.Singh, Mansour H. Assaf, Sunil R.Das, Emil M. Petriu, and Voicu Groza, "Short Duration Voice Data Speaker Recognition System Using Novel Fuzzy Vector Quantization Algorithms", IEEE International Instrumentation and Measurement Technology Conference, May23-26, 2016.