HIDING VOICE DATA IN CENTER DENSITY OF SPEECH SPECTRUM FOR SECURE TRANSMISSION

Rupa Patel[1], Ruma Dhanawate[2]

[1]*Department of Computer Application, Shri Ramdeobaba College of Engineering & Management, Nagpur, 440013, India*

rupa2109@gmail.com

[2]*Department of Computer Science, Arihant College of Education, Pune, India*

rumadhanawate@rediffmail.com

Abstract

Speech data is becoming an effective and indispensable way for fast information transmission but associated with it are number of unprecedented threats. The idea of this paper is to present a robust speech watermarking method to realize secure voice data transmission. In this method carrier is transformed into frequency domain. Pre-processed normalized covert voice data undergoes exponential transformation which is then substituted in center density of high frequency, high energy subband of carrier. High frequency components are chosen for embedding watermark for two reasons. First, during transmission carrier might get contaminated with noise and filteration generally suppress low frequency components so embedding in high frequency component will keep watermark intact. Second reason is human ears are less sensitive to high frequencies so slight change in amplitude of high frequency components is imperceptible.. Technique uses frequency masking, invisible, and blind approach also. By applying reverse approach sensitive message can be extracted from the watermarked carrier. For embedding and retrieving watermark secret key have been used. Experimental results have shown that proposed method does not change the size of the cover signal even after embedding, does not degrade the quality of carrier and exhibit vigorous voice data hiding performance. Proposed work can be used in those applications where maintaining integrity and secrecy about the information against intentional or unintentional access is given utmost importance.

KeyWords: *Index Terms- Exponential, Frequency Masking, Musical sequence, Speech watermarking, secure voice data, signal to noise ratio*

----------***----------

I. INTRODUCTION

Technology advances have eased fast transmission of digital information. During transmission integrity of digital information is often threatened. As a result security and protection control mechanism is needed and one of the effective solutions is digital watermarking. Digital watermarking describes approach to embed data (watermark) into and extract from host signal (cover) to create watermarked signal. Fig. 1 shows general watermarking system.

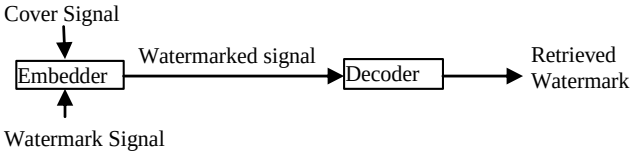

Fig.1 General watermarking system

Watermarking techniques are application dependent. According to type of media to be watermarked they are categorized as audio, image, text and video watermarking. Watermarking system exhibit number of properties namely imperceptibility, robustness, watermark bit rate, capacity of the channel, security, speed, asymmetry, cascadability of the watermark [1][2][3]. Audio watermarking is a technique which is used to protect audio information and can also be used for hiding speech data. Audio watermarking depends upon the human auditory system (HAS). For hiding information masking properties are used [3]. Audio watermarking uses frequency or temporal masking properties to embed watermark into host signal for hiding information and these masking properties are dominated by HAS[3]. When the sounds that human normally perceive is made inaudible by another sound, if both sounds are close to each other by frequencies then this masking in frequency domain is referred as frequency masking. The masking that occurs in time domain is referred as temporal masking. It occurs when a strong sound is preceded (pre-masking) or followed by weaker sound (post masking) of same or nearby frequency [3][5]. Audio watermarking techniques can be applied to speech watermarking problems [4].

The method presented uses simultaneous masking approach to cloak confidential information based on the fact that the loud sounds tend to mask out weaker sounds [3].

The rest of the paper is organized as follows. Section II describes various speech watermarking techniques. Presented approach is described in section III. Results are discussed in section IV followed by conclusion.

II. OVERVIEW OF SPEECH WATERMARKING TECHNIQUES

Several techniques for embedding watermark in host signal have been developed taking the advantage of human auditory system (HAS) perceptual properties [3]. In literature watermarking techniques are categorized on the basis of domain operations as time domain methods and transformation based methods [7], according to human perception as visible, invisible, and dual watermarks [2], according to extraction process as blind, semi-blind and non blind techniques [2].The speech watermarking techniques are discussed below

A. Least Significant Bit Coding

It is a time domain watermarking technique. In this technique modification of individual samples of carrier take place by substituting least significant bit (LSB) of the carrier signal with watermark bit pattern. The length of the carrier is usually greater than or equal to that of watermark. The advantage of this technique is that watermark channel capacity is high but random changes in LSB destroys coded watermark, introduces additive white gaussian noise. LSB works best only in pure digital environment [3] [6].

B. Spread Spectrum Technique

Using domain transformation techniques original host signal is first transformed to another domain and then watermark is spread over several frequency components of carrier such that the energy of any individual component is altered with very small value [6][7]. Reverse procedure is followed to retrieve watermark. Spread spectrum is easy to implement and robust against signal manipulations. But vulnerable to time scale modification attacks [6] [8].

C. Parity encoding

In this technique host signal is partitioned into small block and each watermark bit is encoded in a parity bit of these blocks. If watermark bit does not matches with the parity bit of selected block then LSB of the samples in the blocks are changed [9].

D. Phase coding

In this approach watermark is embedded in the phase of host signal considering the fact that human ears are less sensitive to relative phase changes. The host signal is divided into blocks and watermark is substituted into the phase of the first block and in order to maintain the relative phase between blocks phase of subsequent blocks are adjusted. Watermark can be easily retrieved if the length of the block is known[6][9].

E. Echo hiding

This technique embeds single echo or multiple echoes as watermarks into discrete carrier to generate watermarked signal which retains statistical and perceptual characteristics of original host signal. The original signal is partitioned into small blocks and these blocks are then considered as individual signal for encoding with desired bit. Once encoding is done, the encoded signal blocks are concatenated to produce watermarked signal [6].Merit of this approach is that it can improve the quality of host signal but it is susceptible to malicious attacks.

F. Quantization Index Modulation

A blind watermarking scheme in which watermark is embedded by quantizing the host signal using set of quantizer. The watermark is the index for the quantizers. For retrieving watermark the distance metric to all quantizers is evaluated and the index of the quantizer with the smallest distance contributes to watermark [6].

G. Patchwork Method

Blind watermarking scheme in which the host signal is divided into two subsets called patches. Certain amount of value is added to samples of one patch while the same value is subtracted from samples of another patch. It uses statistical characteristics of the host signal to embed data so watermark can be easily detected. Decoding is performed by finding the differences between these two subsets. For unmarked signal the computed value will be zero otherwise it will be nonzero [6] [8].

III. PROPOSED METHOD

The proposed watermarking technique is summarized below.

A. Watermark Embedding Process

Fig.2 depicts flowchart of the embedding process.

Steps are as follows.

1. Record the confidential information w(y) at the sampling rate of 8 KHz with time duration of 3 seconds.

2. Voice Activity Detection is performed to detect the presence of voiced speech using short time energy and Zero Crossing Rate (ZCR)[10][11].

3. Interested in voiced segment of the input signal only. So crop the signal to get voiced segment. To reduce any unwanted component such as noise in the signal apply weiner filter to the segment. [12][13]This processed watermark is denoted by w (m).

$$w(m) = weiner(crop(w(y))$$

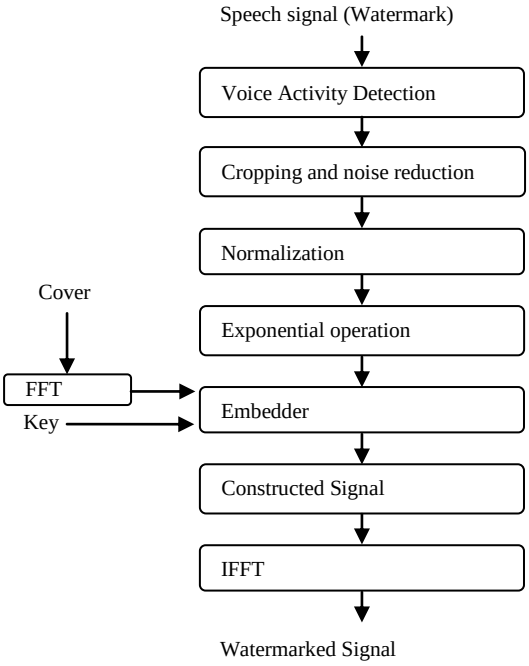

Fig.2 Flowchart of Embedding process

4. Watermark is normalized to alter energy and pitch of the signal [15].

5. For additional security apply exponential operation watermark
$$W(M) = \exp(w(m))$$

6. Original Cover signal c(n) is transformed into frequency domain using Fast Fourier Transform(FFT)
$$C(n) = FFT(c(n))$$

7. Embed watermark in high frequency components (having high energy compared to other frequency components) of the cover signal using watermark length as key and frequency masking concept provided number of samples in watermark is less than or equal to that of cover.
$$cw(n) = C(n) + W(M)\,; M \le n$$

8. Constructed signal cw (n) is transformed back into time domain using inverse FFT.
$$CW(n) = IFFT(cw(n))$$

B. Watermark Extraction Process

Watermarked signal is transformed into frequency domain and then reverse process is followed to retrieve embedded watermark as depicted in Fig. 3.
The speech watermarked signal is the input to extraction process. It is transformed in frequency domain. Two inputs to the decoder are transformed signal and watermark length as key. Decoder extracts the embedded watermark. The watermark is then operated by logarithmic function. The outcome of this model is a speech signal which is identical to that of input speech signal.

Following steps are followed

1. Received watermarked signals are transformed into frequency domain.
$$RCW(n) = FFT(CW(n))$$

2. Embedded information denoted by RW (M) is extracted from highest frequency components using watermark length as key.

3. To get back the watermark , logarithmic operation is performed and retrieved watermark is represented by rw (m).
$$rw(m) = \log(RW(M))$$

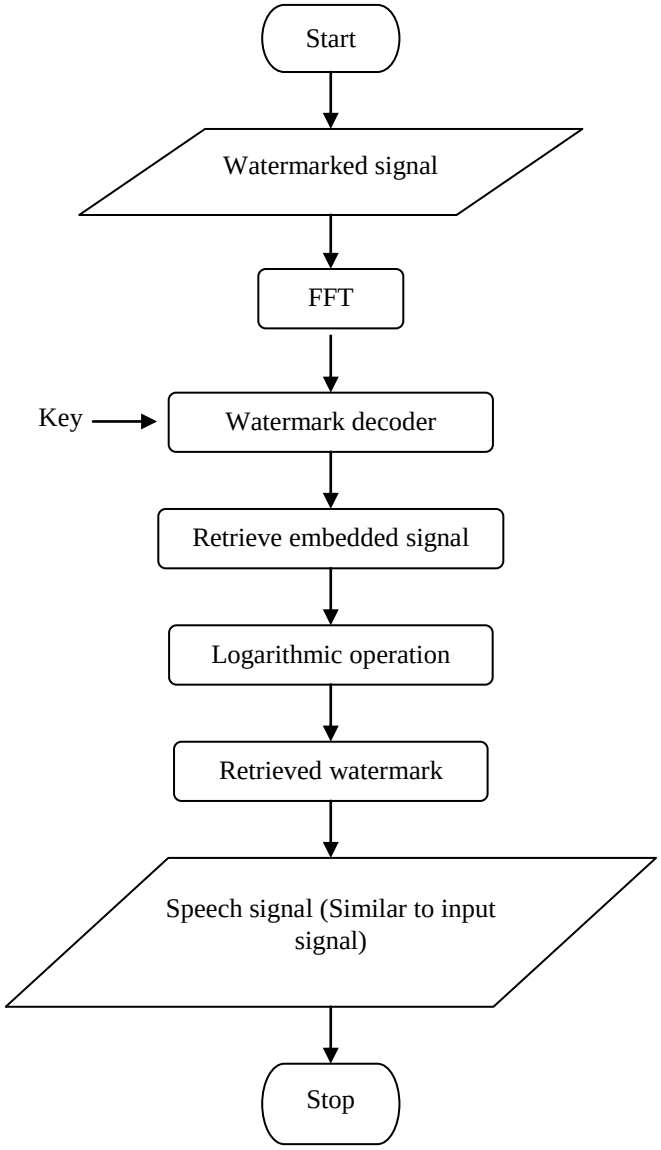

Fig. 3 Flowchart of Extraction Process

IV. RESULTS AND DISCUSSION

MATLAB have been used to implement algorithm. For experimental purpose one word information was recorded. Table I below shows the test parameters.

Table I. Test parameters

Language	English
Speakers	7
Speech Type	Word
Recording Condition	Openroom (Noisy)
Sampling Frequency	8KHz

A. Embedding Process

To perform Voice Activity Detection frame of 20ms is chosen with 50% overlapping and hamming window. The results of VAD are summarized in Table II [15].

Table II. Voice activation detection result

Signals	Energy (db)	ZCR
Case 1	0.6085	0.0791
Case 2	0.3633	0.1004
Case 3	0.6866	0.0713
Case 4	1.0293	0.0752
Case 5	0.1542	0.0907
Case 6	0.4785	0.0803
Case 7	0.4398	0.0749

Fig.4 shows the energy and ZCR plot of the input signal for case 1. This indicates that the input signal is voiced speech signal.

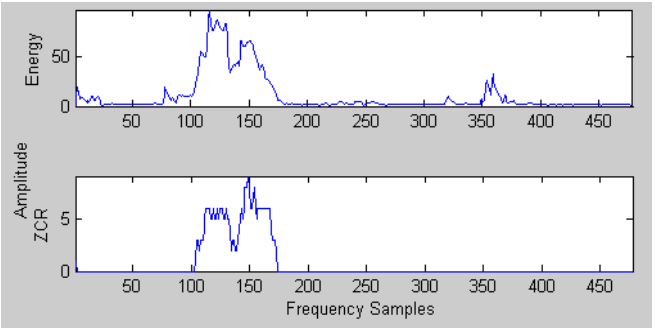

Fig.4 Energy and ZCR plot of the speech signa for case 1

The isolated word from the signal is determined and cropped as shown in Fig. 5.

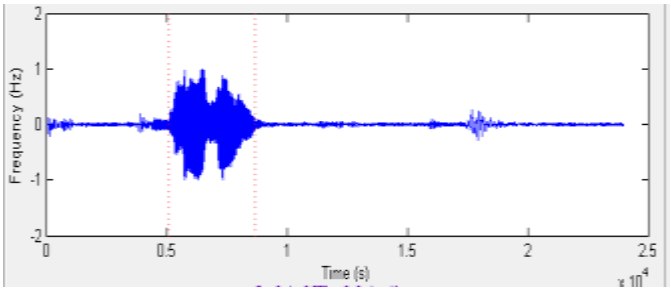

Fig.5 Isolated word detection for case 1

Similar plots were obtained for other cases. The desired signal is often observed in noise, so it is filtered using weiner filter. The implementation of filter often requires estimate of Signal to Noise ratio (SNR) which is obtained from the estimates of the power spectra of the signal and noise [13]. SNR value of the speech signal is tabulated in Table III.

Table III. Results of Weiner filter

Signals	SNR (db)		
	Before filtering	**After filtering**	**Improvement**
Case 1	2.4565	24.8676	22.4111
Case 2	0.7504	6.0803	5.3299
Case 3	7.1191	23.6743	16.5552
Case 4	6.7261	23.6149	16.8888
Case 5	9.4415	23.2383	14.3616

Signal is normalized [15]. To enhance security measures exponential operation is performed. Frame by frame analysis of waveform is analyzed before and after operation. Their energy values are summarized in table IV.

Table IV. Result of Exponential operation

Signals	Frames	Energy	
		Before	**After**
Case 1	Frame 1	0.2671	0.0197
	Frame 2	0.0009	0.0209
	Frame 3	0.0002	0.0219
	Frame4	0.0838	0.0235
	Frame 5	0.0118	0.0255
Case 2	Frame 1	0.4612	0.2543
	Frame 2	0.2386	0.3181
	Frame 3	0.1377	0.3457
	Frame4	0.0029	0.3633
	Frame 5	0.2584	0.4861
Case 3	Frame 1	0.6064	0.4307
	Frame 2	0.0762	0.4619
	Frame 3	0.0017	0.4879
	Frame4	0.0020	0.5106
	Frame 5	0.0722	0.5388
	Frame 6	0.0716	0.5567
Case 4	Frame 1	0.1671	0.0018
	Frame 2	0.0003	0.0018
	Frame 3	0.0785	0.0018
	Frame4	0.0144	0.0019
Case 5	Frame 1	0.2608	0.0117
	Frame 2	0.8220	0.0119
	Frame 3	0.1018	0.0124
	Frame4	0.0839	0.0126
Case 6	Frame 1	0.4131	0.1270
	Frame 2	0.1258	0.1268
	Frame 3	0.2263	0.1271
	Frame4	0.1231	0.1260
Case 7	Frame 1	0.3585	0.0603
	Frame 2	0.0016	0.0633
	Frame 3	0.0010	0.0662
	Frame4	0.0004	0.0688

A non voiced musical sequence [14] is considered as carrier since musical sequences have high energy and high frequency range as compared to speech signal. The low energy watermark can be easily substituted in high frequency components of the host signal providing robustness and imperceptibility. During transmission signals might get contaminated with additive noise. And filtering generally remove low frequency components as a result watermark bits might be removed so embedding watermark in high frequency component will keep watermark safe. The carrier is transformed using FFT. It is observed that frequency sampling of cover is 22050 Hz. The length of the watermark and carrier were estimated. The length of the carrier is observed to be 70641 which is greater than that of watermark length. Thus embedding is possible. Using watermark length as key and frequency masking approach, watermark is embedded in center density of high frequency components. The energy is computed for watermark, cover and reconstructed signal. Results are summarized in table V

Table V. Comparative Energy values of watermark and carrier

Signal	Length of Watermark	Energy		
		Cover	Watermark	Reconstructed Cover (Watermarked)
Case 1	3640	0.6250	0.1214	0.6461
Case 2	3480		0.0114	0.6491
Case 3	4360		0.3626	0.6702
Case 4	2520		0.0019	0.7792
Case 5	3160		0.1712	0.6628
Case 6	3160		0.2162	0.6540
Case 7	2920		0.0688	0.7643

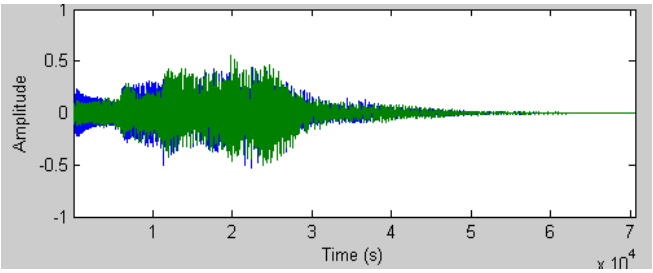

Fig. 6 Plot of cover signal before embedding watermark

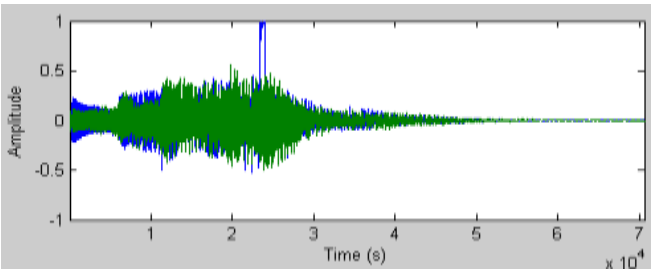

Fig.7 Plot of cover signal with watermark (Case 1)

The length of the constructed signal and original host signal is observed to be same, that is 70641 and also the size of both the wave file is 275KB.The energy difference between cover and reconstructed signal confirms that the watermark is embedded within cover signal. Fig 6 and Fig 7 shows the plot of signal before and after embedding process respectively for case 1. Almost similar results were obtained for remaining cases.

B. Extraction Process

Reverse procedure is followed to extract watermark. Extraction of watermark does not require the original cover signal. The watermark length as a key is used to retrieve the embedded data. Based on frame by frame analysis, table VI summarizes the energy of the embedded watermark and extracted watermark.

Table VI. Comparative Result of Watermark Frame by Frame

Signal		Energy	
		Embedded Watermark	Extracted Watermark
Case 1	Frame 1	0.2671	0.2671
	Frame 2	0.0009	0.0009
	Frame 3	0.0002	0.0002
	Frame4	0.0837	0.0838
	Frame 5	0.0118	0.0118
Case 2	Frame 1	0.2543	0.2543
	Frame 2	0.3181	0.3181
	Frame 3	0.3457	0.3457
	Frame4	0.3633	0.3633
	Frame 5	0.4861	0.4860
Case 3	Frame 1	0.4307	0.4307
	Frame 2	0.4619	0.4620
	Frame 3	0.4880	0.4879
	Frame4	0.5106	0.5106
	Frame 5	0.5388	0.5389
	Frame 6	0.5567	0.5567
Case 4	Frame 1	0.0018	0.0018
	Frame 2	0.0018	0.0018
	Frame 3	0.0018	0.0018
	Frame4	0.0019	0.0019
Case 5	Frame 1	0.0117	0.0117
	Frame 2	0.0119	0.0119
	Frame 3	0.0124	0.0124
	Frame4	0.0126	0.0125
Case 6	Frame 1	0.1270	0.1270
	Frame 2	0.1268	0.1268
	Frame 3	0.1271	0.1270
	Frame4	0.1260	0.1260
Case 7	Frame 1	0.0603	0.0603
	Frame 2	0.0633	00.631
	Frame 3	0.0662	0.0660
	Frame4	0.0688	0.0687

Both embedded and extracted watermarks are compared with respect to energy and pitch parameter and their comparative values are summarized in table VII for all seven cases.

Table VII. Comparative Result of Watermark

Signal	Energy		Pitch	
	Embedded Watermark	Extracted Watermark	Embedded Watermark	Extracted Watermark
Case 1	0.3636	0.3637	210.53	210.53
Case 2	0.5891	0.5890	181.8212	181.8212
Case 3	0.7281	0.7281	216.216	216.216
Case 4	0.0019	0.0019	177.7778	177.7778
Case 5	0.5286	0.5286	242.42	242.42
Case 6	0.2263	0.2260	163.2653	162.2550
Case 7	0.0688	0.0688	186.0465	186.0400

Both the signals have almost same energy and pitch value.

Subjective listening test confirms that the embedded signal and retrieved speech signals are same.

V. CONCLUSION

Human Auditory System properties have been used by proposed technique to embed the watermark in the high frequency components of the cover signal without degrading the quality of carrier, achieving high robustness and imperceptibility. The cover signal is transformed in frequency domain and then instead of embedding sensitive data in low frequency components of the cover signal, normalized watermark is embedded into center density of high frequency components of the cover signal. This ensures that the data will remain intact even after filtering process. The watermark bits were not embedded directly within the frequency coefficients of cover but rather watermark bits after exponential operation are embedded into center density frequency components of cover signal. Human ears are less sensitive to high frequency sounds such as musical sequence, so embedding sensitive data within this sequence achieves robustness and imperceptibility.

The performance of the proposed algorithm is provided by evaluating the parameters such as signal to noise ratio, energy and pitch. The robustness of the watermarking methods has been quantitatively measured by comparing the extracted watermark with the original watermark. To extract the watermark, approach does not require access to the original cover signal. Subjective listening and objective test results confirms that there is perceptual transparency between original and watermarked signals. The objective of securely transmitting and recovering confidential information is accomplished.

REFERENCES

[1] Rupa Patel, Urmila Shrawankar, V. M.Thakre,2011, Secure Transmission of Password using Speech Watermarking, IJCST Vol. 2, Issue 3, September 2011.

[2] Dr. Vipula Singh."Digital Watermarking - A Tutorial", Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunicationss, (JSAT),January edition 2011, pp 10-21.

[3] Nedeljko Cvejic,"Algorithms for audio watermarking and steganography", OULU 2004

[4] Mehmet Celik, G. Sharma, and A. M. Tekalp, "Pitch and duration modification for speech watermarking," in *Proc. IEEE ICASSP*, Mar. 2005, pp. II,17–20.

[5] Zwicker E and Fastl H Psychoacoustics. Springer Verlag, Berlin, Germany,1999

[6] Michael Arnold,Martin Schmuker,Stephen D. Wolthusen, "Techniques and Applications of Digital Watermarking and Content Protection", Boston, London: Artech House, INC, 2003.

[7] Rajkiran Ravula. "Audio Watermarking Using Transformation Techniques", M.S. thesis, Dept. Elect. And Computer Eng.,Louisiana State University, December, 2010

[8] Hyoung Joong Kim,"Audio Watermarking Techniques", Pacific Rim Workshop on Digital Steganography, Kyushu Institute of Technology, Kitakyushu, Japan, July 3-4, 2003.

[9] Poulami Dutta,Debnath Bhattacharyya and Tai hoon Kim,"Data Hiding in Audio Signal : A Review",international Journal of Database Theory and Application,Vol.2,No. 2,June 2009

[10] L. R. Rabiner, R.W. Schafer **,** *Digital processing of speech signals* ,Prentice-Hall, Inc., Reprint 2009

[11] Urmila Shrawankar, Dr. V M Thakare, "Voice Activity Detector and Noise Trackers for Speech Recognition System in Noisy Environment", *International Journal of Advancements in Computing Technology (IJACT),* 2010, ISSN: 2005-8039

[12] Urmila Shrawankar, Dr. V M Thakare, "Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment", *Springer-IIP2010*, Manchester , UK , October 13-16, 2010

[13] Md. Jahangir Alam, Md. Faqrul Alam Chowdhury, Md. Fasiul Alam, "Comparative Study of a Priori Signal - To- Noise Ratio(SNR) Estimation Approaches for Speech Enhancement", *Journal of electrical and electronics engineering,* 2009,vol 1

[14] Konard Hofbauer, et.al. " Speech Watermarking for Analog Flat Fading Bandpass Channels" , IEEE Transactions on Audio, Speech, and Language Processing, Vol 17,Issue 8,2009, pp 1624-1637

[15] Rupa Patel, Urmila Shrawankar, Dr. V M Thakare," Normalization Techniques for Hiding Speakers Identity ",International Conference on Data Science & Engineering, July 2012, pp 75-80.