Speech Steganography using DWT and FFT

Valiki Vijayabhaser[1], Shirisha Munasa D[2], Srinivas Gadari[3]
[1,2,3]Associate Professor, [1,2,3]Department of ECE
[1,2,3]Siddhartha Institute of Technology and Sciences, Telangana, India

Abstract--Steganography is art of hiding information into a cover object that can be an image, video or audio. Speech steganography is a process of hiding message data into a cover speech without degrading the quality of cover speech. In this, a novel user interface for spread spectrum representation-based speech steganography using decimated wavelet transform (DWT), which is an extension for Fast fourier transform (FFT) based steganography.Simulation results proved that the proposed algorithm is superior to the conventional algorithms. Also performed good enough simulations with low bit error rate and excellent imperceptibility**.**

***Keywords*---Digital Steganography, Spread Spectrum, Speech steganography, FFT, wavelet analysis and discrete wavelet transform.**

I. INTRODUCTION

Steganography referred to hiding information or any secrete message behind a cover object which might be an image, audio or video. This avails the persons who are authorized recipients can only view the message sent from the source end [1]. To obtain an effective steganography, one must need the following:

- Cover object to hide the secure information.
- Secrete information i.e., message
- Embedding procedure to get a stego information [2].
- Extraction process to reconstruct secrete message at recipient [3].

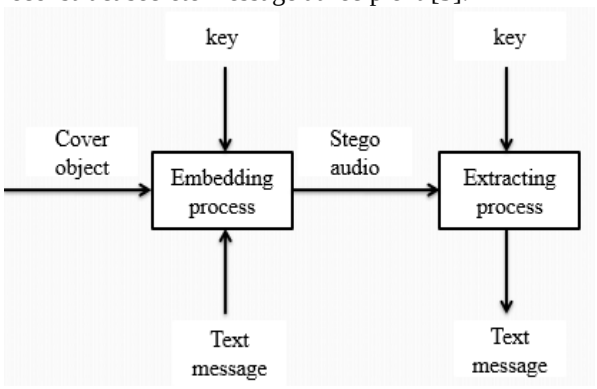

Fig.1 block diagram of digital steganography

This has lots of applications in several fields like multimedia, military, navy and civil etc. In practice, mostof the steganography systems were implemented for images and videos as well. There are very lesser number of research papers published under speech steganography since designing of speech steganography is quite challenging and difficult compared to image and video steganographic systems. Spread spectrum [4] plays a significant role in speech steganographic systems since the speech is a discrete signal information and need to be processed over channel to embed the secrete message and later it must be reconstructed by extracting the accurate message with cover speech separately. Figure 1 depicts the general steganography procedure which consists of cover object, secrete message, secure key, embedding and extraction procedures. First, the secrete message is embedded into a cover audio or speech with the help of key and using an embedding procedure. After embedding the message into a cover speech, stego audio is obtained. Then at the recipient end, the secrete message will get extracted using the symmetric key and an extraction procedure which a reversible for embedding process. Author in [5] addressed least significant bit (LSB) approach for audio steganography, which embeds the message info into the speech data based on LSB approach. However, at the reconstruction end it was difficult to extract message data accurately. Similarly,

there are several speech steganography systems were presented and published [6-10]. Recently, a spread spectrum-based audio steganography is implemented in [11], this approach utilized FFT for embedding the information into the audio signal. This provides enhanced performance over conventional speech steganographic systems. However, it was unable to extract the original embedded message due to the reconstruction issue of FFT algorithm and visibly the stego audio seems far distinct from the original speech signal, which shouldn't happen in practical applications. Therefore, this article proposed wavelet-based speech steganographic system which is an extended version of FFT-based approach [11]. Due to the higher spectral efficiency of decimated wavelet decomposition, there will be a lossless reconstruction of cover speech and hidden message as well.

II. EXISTING METHODOLOGY

This section describes the existing FFT-based spread spectrum representation for speech steganographic system [11]. Primarily, the cover speech is transformed into frequency space using FFT, which computes the discrete fourier transform (DFT) of a speech signal with reduced number of computations. Next, the message info which is to be embedded into the frequency domain signal of a cover speech has converted into binary format by utilizing ASCII codes. Afterwards, this binary message info was spread over the channel using pseudo noise, chip rate as key and an embedding gain factor. Now, this obtained outcome was combined with the FFT signal to get the stego audio. Then, the reconstructed speech signal is obtained by computing inverse FFT to the stego audio. Finally, apply inverse to the embedding procedure to extract the hidden message from the stego audio.

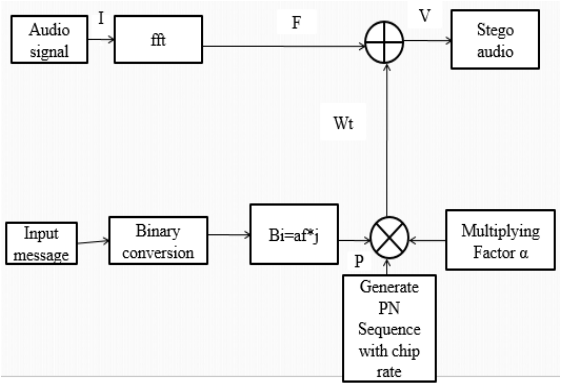

Fig. 2 Speech steganography using FFT approach [11]

III. PROPOSED METHODOLOGY

This section explains the proposed DWT based real time speech steganography for robotic applications. This can be implemented in real time since the cover speech is directly recorded by the use voice as there is an option of speech record in every computer or laptop.

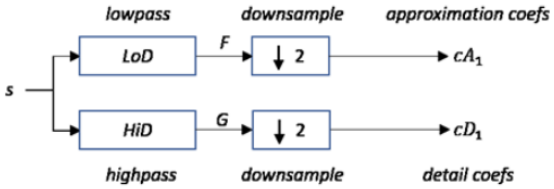

Fig. 3 Decomposition of signal using 1D-DWT

A. DWT

This is an advanced transformation technique compared to fourier transform (FT), short time fourier transform (STFT). It has an adaptive nature of window selection, due to scaling property. As shown in Figure 3, 1D-DWT is applied to a speech signal to decompose the input speech signal into approximate and detail coefficients where the lower frequency subband is referred to approximate layer and higher frequency subband cited to detail layer. Practically, the approximate layer seems like original speech signal.

B. Algorithm

In this section, the steganography scheme will be explained. Cover object used is the raw audio files like WAV files. Suppose we have byte-sequence information that will be inserted into the cover object, the byte-sequence will be converted into bit-sequence information. Then we represent these bits into such signal so that if the bit is 1 then the amplitude of the signal is 1, whereas if the bit is 0 then the amplitude of the signal is -1. As shown as follows:

$$A = \{a_i | a_i \in \{-1,1\}\} \quad (1)$$

Next, open the WAV files and obtain the signal's amplitude data. The amplitude is represented as 16 bit signed integer value with a range of 2^{15}-1 to -2^{15}+1. So, for divide this amplitude with a value of 215-1 in order to obtain the range between 1 to -1. Then, by using FFT, the data will be transformed into frequency domain. Now, a random PN sequence will be generated with 1 or -1. If the chip rate of PN sequence is cr, and there are total of n signals for the information signal, then there must be $cr \times n$ sequences generated. We call the PN sequence P, then

$$P = \{p_i | p_i \in \{-1,1\}\} \quad (2)$$

Modulate each information signal with the PN sequence until cr times, by multiplying the value. It will produce a signal B which is the distributed signal of A and of course with length cr times its original length. Initially, spread the information in A to B as follows:

$$B = \{b_i | b_i = a_i, j \cdot cr \leq i < (j+i) \cdot cr\} \quad (3)$$

Next, modulate B and P and multiply it by a factor α. Then it will be injected into the cover-media. Suppose, the message that is injected is w, the cover is v and the stego object v' i.e., in which we have both the message as well as cover-object. Therefore, it can be formulated as follows:

$$w_i = \alpha \cdot b_i \cdot p_i \quad (4)$$

$$v'_i = v_i + w_i \quad (5)$$

This scheme will generate noise. If the factor of the amplifier is too large, the noise is also large and may damage the cover-object. So, we should be careful in choose of the strength factor and chip-rate. The added signal will be a random signal due to the PN sequence effect which has generated previously. In order for the information to be retrieved, the receiver must generate the same PN sequence. Each cover object signal will be multiplied with the corresponding PN sequence, which can be shown as follows:

$$\sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} p_i v_i' = \sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} p_i v_i + \sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} \alpha b_i p_i^2$$

If we look at the following terms:

$$\sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} p_i v_i$$

The value of these terms will be close to 0 for a large number of samples (large chip rate). This is because the random value of PN sequence causes the sum of the signal approaching 0 or a certain threshold value.
While the second term:

$$\sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} \alpha b_i p_i^2$$

The second term has interesting properties. Because the PN sequence has value 1 or -1, then the result of p_i^2 is 1.
Thus, the term can be simplified into:

$$\sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} \alpha b_i$$

Because we have defined Bi has a value 1 or -1, then we simply conclude that if the term exceeds the value of zero, we assume that the information retrieved is 1 and if the value is less than zero, we assume that the information retrieved is 0. This is the reason we choose the domain of B and P. From the previous explanation, we can conclude that the value of $\propto b_i$ must exceed a certain threshold value in order for a clear information retrieval.

IV. EXPERIMENTAL ANALYSIS

Simulation results have been done in MATLAB 2016b. We tested the proposed and existing methods for various audio samples with different cr values. Fig. 4 discloses the audio steganography of FFT based implementation,

where Fig.4 (a) is a cover audio, Fig. 4(b) is stego audio which comprises of both cover audio and the secret message demonstrated in Fig. 5. Reconstructed audio is shown in Fig. 4(c) and the extracted text message is displayed in Fig. 7. As shown in Fig. 4(b), stego audio seems different from the cover audio signal which results in lower imperceptibility. Our proposed audio steganography outcome is disclosed in Fig. 8, where it seems both cover audio and stego audio are very similar, which results in higher imperceptibility and its quite hard to identify or detect the message embedded into it.

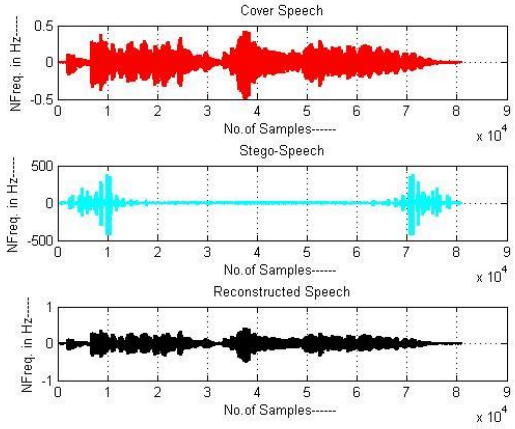

Fig. 4. FFT based embedding secret message into an audio signal (a) Cover audio. (b) Stego audio. (c) Reconstructed audio

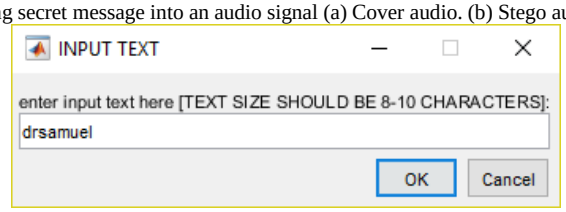

Fig. 5. Dialog box for entering a text to be embedded into an audio signal

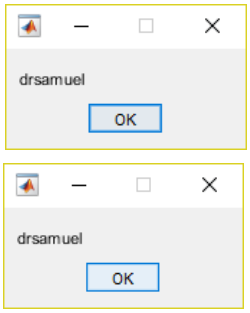

Fig. 7 Embedded and Extracted messages

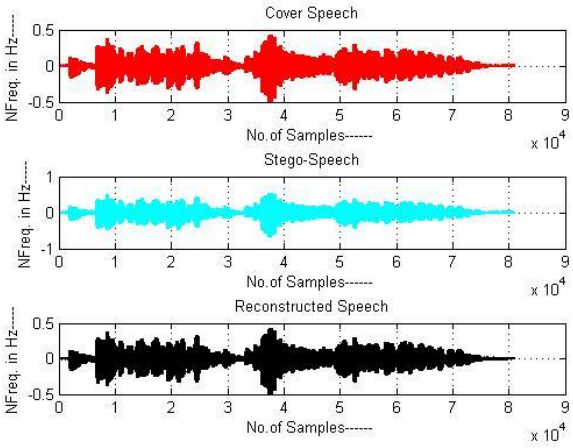

Fig. 8. Proposed audio steganography (a) cover audio. (b) Stego audio. (c) Reconstructed audio.

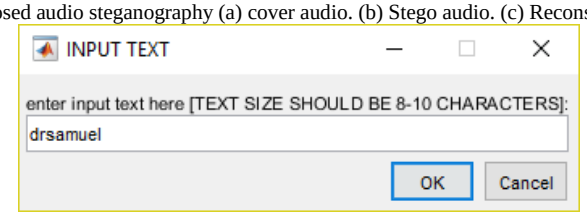

Fig. 9. Dialog box for entering a text to be embedded into an audio signal

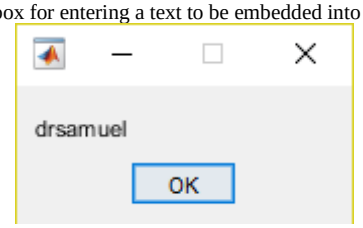

Fig. 10. Embedded message

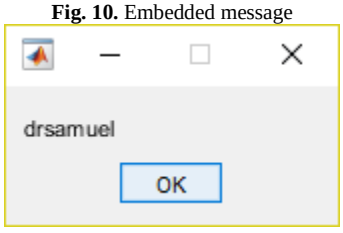

Fig. 11. Extracted message

V. CONCLUSIONS

Implementation of novel user interface for spread spectrum representation-based speech steganography using translation invariant wavelet transform is done successfully. Obtained simulations proven that the proposed speech steganography got superior performance over conventional FFT-based steganography algorithm. This method proved that it is very robust against audio manipulation and very safe with the resulting noise is quite small. Also it reduces number of computations and does not use any complex equations. It is very simple and easy method to implement even in real time environment.

REFERENCES

1. Shouyuan Yang, Zanjie Song and Jong Hyuk Park "High capacity CDMA Watermarking Scheme based on orthogonal Pseudo random subspace projection". International Conference on Multimedia and Ubiquitous Engineering, June 2011
2. Lionel Fillatre "Adaptive Steganalysis of Least Significant Bit Replacement in Grayscale Natural Images" IEEE Transactions on Signal Processing ,Vol. 60, No. 2, February 2012
3. R.R.Ahirwal, Deep chand Ahirwal and Jpgendar jain "A High Capacitive and Confidentiality based Image Steganography using Private Stego key" International coference on Information Science and applications, Feb 2010.

4. Rikzy M. Naguraha "Implementation of Direct sequence Spread Spectrum on Audio Data" International Conference on Informatics Engineering, June 2011.
5. A. Binny and M. Koilakuntla, "Hiding Secret Information using LSB based Audio steganography", In Proc. of International Conference on Soft Computing and Machine Intelligence, New Delhi, India, Apr. 2015.
6. N. Kundu and A. Kaur, "Ä Secure approach to audion steganography", International Journal of Engineering Trends and Technology, vol. 44, no. 1, pp. 1-7, 2017.
7. N. Kumari and B. Yadav, "Audio steganography", International Journal of Innovative Research in Technology, vol. 3, no. 12, pp. 115-119, May 2017
8. B. G. Banik and S. K. Bandyopadhyaay, "Blind Key Based Attack Resistant Audio Steganography Using Cocktail Party Effect", Security and Communication Networks, vol. 2018, Article ID 1781384, 21 pages, 2018.
9. R. Joshi and P.S. Venugopala, "Improved security in audio steganography using packet forger at third level", International Conference on Smart Technologies For Smart Nation, IEEE, Bangalore, India, May 2018.
10. M. H. A. Al-Hooti and T. Ahmad, "Reversible Data Hiding in Audio based on Discrete Cosine Transform and Location Maps", International Journal of Intelligent Engineering and Systems, vol. 12, no. 3, pp. 41-49, 2019.
11. A. S. Patil and G. Sundari, "An embedding of secret message in audio signal", 3rd IEEE International Conference on Convergence Technology, Pune, India, Nov. 2018.