Real Time Speech Steganography for Secure Data Transmisson

Kalluri Saidatta Subramanya RaviTeja

PG Scholar, Department of Electronics and Communication Engineering, Bharath Institute of Engineering and Technology, Hyderabad, Telangana

Dr.Rajeev Shrivastava

Associate Professor, Department of Electronics and Communication Engineering, Bharath Institute of Engineering and Technology, Hyderabad, Telangana

***Abstract*—Information hiding into a cover object is referred as steganography. Here the cover object might be an image, video, or speech. Speech steganography is a process of hiding message data into a cover speech without degrading the quality of cover speech. This article introduced a spread spectrum representation-based speech steganography using discrete wavelet transform (DWT), which decomposes the cover speech signal into approximated and detail coefficients i.e., low frequency and high frequency. Our proposed speech steganography provides enhanced imperceptibility since DWT reconstructs the decomposed information without degrading the quality of speech. Our proposed approach is an extended version of existing Fast fourier transform (FFT) based steganography, where there is a lack of imperceptibility. 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—steganography; spread spectrum; speech steganography; fast fourier transform; wavelet analysis; decimated wavelet transform.

I. INTRODUCTION

Steganography referred to hiding information or any secret message behind a cover object which might be an image, speech 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.
- Secret information i.e., message.
- Embedding procedure to get a stego information [2].
- Extraction process to reconstruct secret message at recipient [3].

This has lots of applications in several fields like multimedia, military, navy and civil etc. In practice, most of 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 secret message and later it must be reconstructed by extracting the accurate message with cover speech separately.

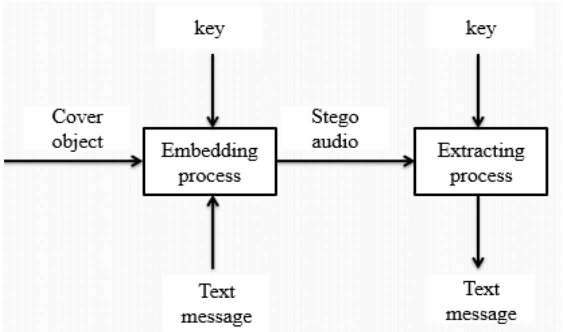


Fig.1 block diagram of digital steganography.

Figure 1 depicts the general steganography procedure which consists of cover object, secret message, secure key, embedding and extraction procedures. First, the secret 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 speech is obtained. Then at the recipient end, the secret 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 speech 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 speech steganography is implemented in [11], this approach utilized FFT for embedding the information into the speech 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 speech 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.

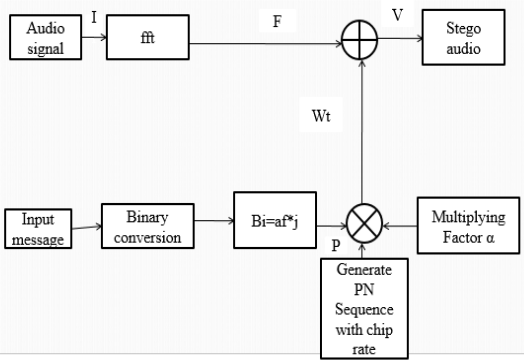


Fig. 2 Speech steganography using FFT approach [11].

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 speech. Then, the reconstructed speech signal is obtained by computing inverse FFT to the stego speech. Finally, apply inverse to the embedding procedure to extract the hidden message from the stego speech.

III. Proposed Methodology

This section explains the proposed DWT based real time speech steganography for military applications, where there is a need for transmission of secret data from person/pc to pc/person with higher imperceptibility. 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.

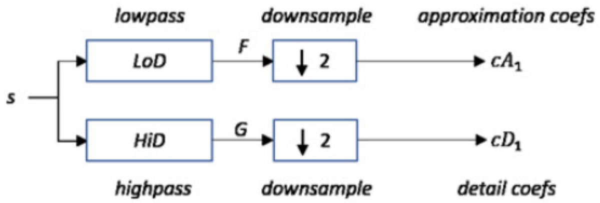


Fig. 3 Decomposition of signal using 1D-DWT.

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

This section describes the proposed real time implementation of speech steganography, where the speech or speech is utilized as a cover object. In general, speech signals will be in .MP3 format most of the times. These MP3 speech signals have an information in byte sequences which later exchanged to bit sequence and subsequently represented in the range of -1 to 1 since spread spectrum methodology is utilized for steganography. Following formula shows the conversion of speech signal:

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

Now, FFT is used to transform the cover speech into frequency domain, which now has both real and complex information in it. After that, a pseudo noise (PN) sequence is generated in the range of 1 and -1 same as cover speech as given in equation (1). The generation of PN sequence is functioned via chip rate denoted with cr. If there are n number of signals, then there must be a generation of $cr \times n$ sequences. The generation of PN sequence P is as follows:

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

Now, utilize PN sequence form equation (2) and modulate every signal information by multiplying cr times, which then results in distributed signal of equation (1) and denoted with B as follows:

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

Afterwards, modulate both equation (2) and equation (3) with a multiplication factor α, which is referred as embedding strength factor. Later, it is included into the cover speech. For instance, w denotes the message included, cover speech referred as v and v' denotes stego speech i.e., where it consists both secret message and cover speech. Hence, it is formulated as:

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

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

This approach doesn't produce any kind of noise as existing algorithm [11], that generates higher noise when there is a large amplification and results in cover speech impairment. So, it is not necessary to be careful in choosing the strength factor and chip-rate as happened in [11]. The added signal will be a random signal due to the PN sequence effect which has generated previously. There must be similar PN sequence generation of receiver to retrieve the information. 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:

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

For majority of samples, zero will be the value for these terms, which is due to the sequence of PN that causes addition of signal reaching 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 above term has some noteworthy properties since the value of PN sequence is 1 or -1 which led the outcome of p_i^2 is 1.

Thus, the term can be simplified into:

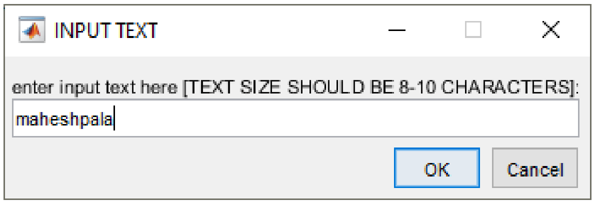
$$\sum_{i=j\cdot cr}^{(j+1)\cdot cr-1} \alpha b_i$$

Since the value of bi is 1 or -1 which usually decides that when the value of term exceeds 0 then it is considered that the retrieved info is 1 and similarly when it is <0, retrieved info is 0. This is the reason for domain selection of B and P.It is clear that, earlier description concludes that for an accurate extraction of information which was embedded, the value of $\propto b_i$ must exceed a certain threshold value.

IV. Results and Discussion

This section describes the experimental analysis of proposed speech steganography with comparison to the FFT-based approach presented in [11]. All the simulations have been done in MATLAB 2018a environment. We tested the proposed and existing methods for various speech samples of different kind of persons like male, female, child and old age with a chip rate $cr = 400$. First, it will ask the user to speak anything for a period of 5sec (this can be varied according to user interest). Then user must enter the secret message (might be 8 to 10 characters) to be embedded into the cover speech as shown in Fig.4. Figure 5 discloses the secret message entered in input text box as shown in figure 4.

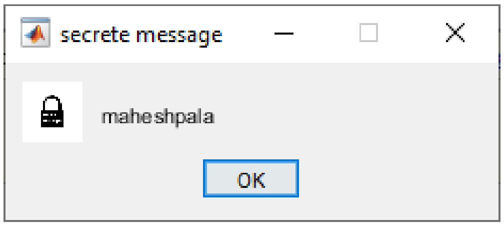
Performance of FFT-based speech steganography described in [11] shown in Figure 6, where the stego-speech lacks the imperceptibility as it is quite dissimilar to the cover speech that means there is a chance for unauthorized parties to identify something must be exists in stego-speech. In addition, bit error rate (BER) values also quite larger as shown in Table 1. Figure 7 demonstrates that performance of proposed real-time speech steganography, which disclosed the similarity between cover and stego speech and results in higher imperceptibility over existing FFT-based speech steganography.



enter input text here [TEXT SIZE SHOULD BE 8-10 CHARACTERS]:

maheshpala

Fig. 4 Entering secret message to be embedded.



maheshpala

Fig. 5 Embedded message.

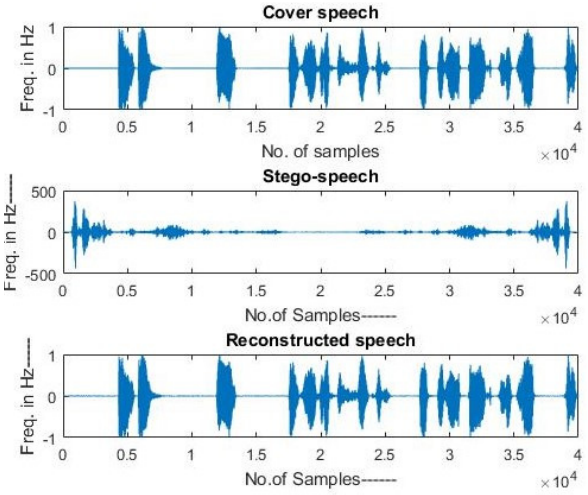


Fig. 6 Performance of FFT-based speech steganography [11].

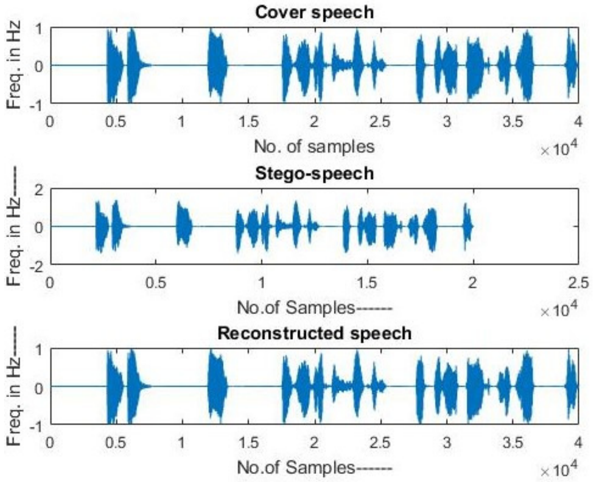
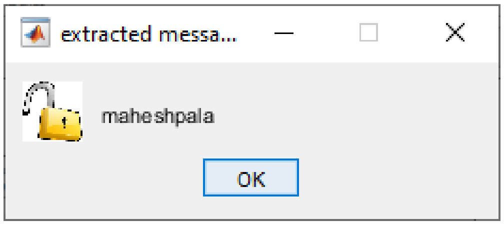


Fig. 7 Performance of proposed speech steganography.



maheshpala

Fig. 8 Extracted message with proposed implementation.

Figure 8 displays the extracted secret message from stego-speech after separation of reconstructed speech from it. Figure 9 show the outcome of proposed speech steganography when there is a noise attack at in the stego-speech. Figure 10 describes the extracted messages from the noisy stego-audio where figure 10(a) show the extracted message using [11] and figure 10(b) show extracted message of proposed speech steganography. This concludes that proposed speech steganography lessens the BER when there is a noisy attack. In addition, it produced accurate message at the receiver end without degrading the perceptual quality of secret message.

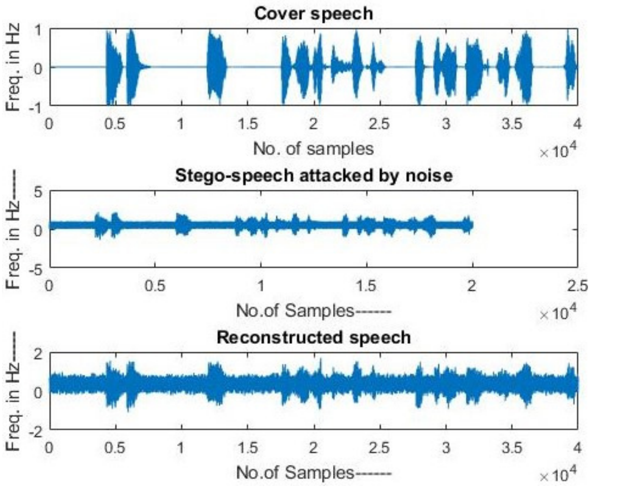
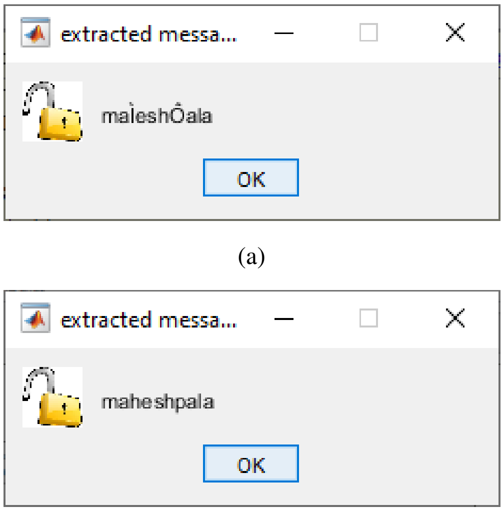


Fig. 9 Obtained results of proposed speech steganography with noise attack.

Table 1. Obtained BER values of existing and proposed speech steganography.

Parameter	FFT-based speech steganography [11]	Proposed speech steganography
BER without noise	0.00145	0.0000001
BER with noisy attack	4.25	0.000452



maÌeshÔala

(a)

maheshpala

(b)

Fig. 10 Extracted message from noisy stego-speech. (a) FFT-based speech steganography. (b) proposed speech steganography.

V. Conclusions

This article addressed the implementation of spread spectrum representation-based speech steganography using DWT. 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 any sort of noise attacks. It also reduced the computational complexity since it doesn't utilize any complex equations. Further, it is very easy and simple to implement in real-time.

References

[1] S. Yang, Z. Song and J. H. Park, "High capacity CDMA Watermarking Scheme based on orthogonal Pseudo random subspace projection," International Conference on Multimedia and Ubiquitous Engineering, Jun. 2011.

[2] L. Fillatre, "Adaptive Steganalysis of Least Significant Bit Replacement in Grayscale Natural Images," IEEE Transactions on Signal Processing, vol. 60, no. 2, Feb. 2012.

[3] R. R. Ahirwal, D. C. Ahirwal and J. Jain, "A High Capacitive and Confidentiality based Image Steganography using Private Stego key," International coference on Information Science and applications, Feb. 2010.

[4] R. M. Naguraha, "Implementation of Direct sequence Spread Spectrum on Audio Data," International Conference on Informatics Engineering, Jun. 2011.

[5] S. Rekik, D. Guerchi,H. Hamam and S.-A. Selouani, "Audio Steganography Coding Using the Discrete Wavelet Transforms,"International Journal of Computer Science and Security, vol. 6, no. 1, 2012.

[6] A. Kaushal and V. Chaudary, "Secure image steganography using different transform domains," Int. J. Comp. App., vol. 77, no. 2, pp. 24-28, 2013.

[7] P. P. Balgurgi and S. K. Jagtap, "Audio steganography usde for secure data transmission," In Proc. of Int. Conf. Adv. Comp., vol. 174, Springer, New Delhi, pp. 699-706, 2013.

[8] S. Wang and M. Unoki, "Speech watermarking method based on for-mant tuning," IEICE Trans. Inf. Syst., vol. E98-D, no. 1, pp. 29–37, Jan. 2015.

[9] K. Bhowal et al., "A steganographic approach to hide infromation in audio signal using wavelet transform," Int. J. Adv. Comp. Sci., vol. 7, no. 4, pp. 42-47, 2016.

[10] V. V. Priyanka and Y. V. A. Sathyanarayana, "A New user interface for spreadspectrum representation based speech steganography using translation invariant wavelet transform," Int. J. Elec. Comp. Comm. Eng., vol. 4, no. 3, pp. 147-151, Jun. 2019.

[11] J. Chen and J. Carlos, "A Spread Spectrum Representation Based FFT Domain Speech Steganography Method," IEEE Transaction on Audio, Speech and Language letters, vol. 23, no. 1, 2015.