# Hybrid Speech Steganography System using SS-RDWT with IPDP-MLE approach

**Chinnarao Rayudu** ( rayudu.chinnarao@gmail.com )
Malla Reddy College of Engineering & Technology
**Jayasree P.V.Y.**
GITAM Institute of Technology
**Srinivasa Rao S.**
Malla Reddy College of Engineering & Technology

**Research Article**

**Keywords:** speech steganography, pause detection, spread spectrum, intelligent pause detection protocol, maximum likelihood estimation, redundant discrete wavelet transform

**Posted Date:** March 22nd, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-198593/v1

# Hybrid Speech Steganography System using SS-RDWT with IPDP-MLE approach

[1]R Chinna Rao, Assistant Professor, Department of ECE, Malla Reddy College of Engineering & Technology, Secunderabad
[2]Dr PVY Jayasree, Professor, Department of ECE, GITAM University, Visakhapatnam[2]
[3]Dr S Srinivasa Rao, Professor, Department of ECE, Malla Reddy College of Engineering & Technology, Secunderabad

## Abstract

In the last two decades, it has been observed that there have been remarkable advancements in digital media communication. It has lots of advantages and business potential as it needs no physical media or transport. However, digital media can also create several big problems for media owners due to unauthorized use, ease of replication, plasticity, and equivalence of works in digital form. The possible solution is to secure digital media by cover writing which can also referred as steganography. Therefore, this article focusing on development of hybrid speech steganography system using spread spectrum-based redundant discrete wavelet transform (SS-RDWT). In general, speech steganography utilizes cover speech to embed the secrete message, however, the cover speech might be larger in size and contains many pauses, which requires more storage capacity, higher computational time, and even higher power usage results in system performance degradation. Hence, in addition to the proposed SS-RDWT approach, an intelligent pause detection protocol (IPDP) with maximum likelihood estimation (MLE) technique is employed for removing pause from cover speech signal, which reduces the transmission bandwidth, storage capacity and power consumption as well. Simulation results demonstrate that proposed hybrid steganography using integrated SS-RDWT with IPDP-MLE approach obtained superior performance as compared to state-of-art approaches.

**Keywords:** speech steganography, pause detection, spread spectrum, intelligent pause detection protocol, maximum likelihood estimation, redundant discrete wavelet transform.

## 1. Introduction

One of the most essential forms in human communications is speech. The speech files can be transmitted and received today over the internet because of the rapid evolution in communications technologies. These files could be accessed during communication for malicious purpose by the unauthorized users. To safeguard the speech signals content while transmission throughout any unsecured channel, cryptographic techniques are introduced. Cryptography and steganography are the most basic methods of hiding information. The term steganography refers to “cover writing” and cryptography refers to “secret writing”. Cryptography deals with transmission of messages in a unique form allowing only the intended recipients to remove the hidden information and read the message. These techniques are used to satisfy the need for confidentiality on the web. Steganography also referred as „writing in hiding“ is a method where the data is hidden in a cover media. Recent examples of steganography include usage of special inks to inscribe hidden messages on currency notes, fingerprinting and digital steganography of audio and video for copyright protection. The fast growth of the web in the previous years has rapidly increased the accessibility of digital data in the form of video, audio and images to the community. As we have seen in the previous few years, with the advances in Internet, data sharing has been effortless irrespective of distance. In such cases, since the material is not stored on the server, it becomes difficult for the copyright owner to detect the offending parties. In the music industry, it is estimated that sharing data or files on the Internet and plagiarizing data costs a loss of more than £2.8 billion a year. In this respect, some thoughtful work needs to be done in order to maintain the accessibility of multimedia information. During the past years, digital steganography was most commonly used for still images but during the recent days, they are used more commonly on data related to the communicative media namely video and audio samples. Hiding information in audio records is about examining in light of the fact that the human auditory system (HAS) performs well over a wide powerful range. The HAS spreads over a scope of intensity more prominent than one billion to one and a scope of frequencies more prominent than one thousand to one. Also, it is very hyper sensitive to additional random noise. The derangements in a sound document can be identified as low as one fraction in ten million (80 dB beneath surrounding level). However, few "holes" are also available there. Rather HAS was a huge unique range, it works at a small differentiating range. Therefore, generally loudness in sounds will be covered with the calm audio. Moreover, the HAS could not see absolute phase, rather relative phase only. Ultimately, some environmental perturbations are so common that they cannot be overlooked by the listener in most of the observations. We misuse a significant number of these qualities in the strategies we examine straightaway, while being mindful so as to manage as a main priority the extraordinary sensitivities of the HAS. Replica of multimedia documents can be prepared easily along with modification and can also be transformed and diffused due to their digital nature. For maintaining authentication and integrity of the digital documents, it is important to develop a secured system. When building up information concealing technique for sound, the principal concerns is of the supportive environment the audio signal will go between

stego outcome and reconstructed signal. To meet the current problems in speech steganography, this article contributed as follows:

- For providing more security to the message data, PN sequence-based multiplication is used. This will reduce the attacks presented on the message data while retrieving the data from stego speech.
- A Novel Pause detection and Removal algorithm is developed using the IPDP approach to reduce the bandwidth of the speech signal. This bandwidth reduction results in higher data rates for speech signal.
- Then for performing the speech steganography operation RDWT, Data embedding, and extraction operations are used, it results the maximum efficiency.
- The proposed method is compared with the various states of art approaches such as FFT [7] and DWT [8]. The comparison result shows that the proposed hybrid approaches have robust towards various attacks and noises respectively.

Rest of the paper is contributed as follows: Section 2 deals with the analysis of various existing approaches with their drawbacks. Then, section 3 deals with the detailed operation of proposed Pause detection and removal operation. Then, Section 4 deals with the proposed method with detailed operations of embedding and extraction procedures. Then, section 5 deals with the detailed analysis of simulation results with various parameter and they are compared with the state of art approaches.

## 2. Literature review

Speech steganography is one of information hiding type in audio file as host signal. Information hiding in audio file has good challenges due to the limited perceptual of human audibility. This fact caused much research in audio processing using HAS characteristics, such as audio steganography and audio compression. The performance of audio steganography depends on several criterions. First criteria are imperceptibility or fidelity in which the watermarked audio perceptually same as host audio. The performance parameter for this criterion consists of objective quality that calculated by formula or model and subjective quality that is taken from survey to several persons. In [6] authors used codebook techniques to embed watermark data in frequency domain. He used log spectrum for embedding data. But he added dirty paper codes and LDPC for improving the robustness. They proposed a patchwork method on audio steganography using discrete cosine transform (DCT) in the embedding process. In [7] authors have used FFT transformation bit that is useful to provide audio steganography robustness against attacks de-synchronization attack such as, pitch, time scaling, and jitter. It explains in detail about some of the basic methods of audio steganography, among others: FFT method, phase coding, spread spectrum, echo hiding, SWT, and histogram techniques. And she concludes with a discussion of audio steganography method proposed for performance with excellent durability, capacity, and better imperceptibility. In [8] authors use the merger of three methods of transformation in the process of insertion of data in the audio. Such method is DWT and DCT. Its embedding method is using the quantization of the DWT results. In [9] authors use the sudoku matrix-based was done after FFT process. The proposed embedding will be done at every range of found two Fibonacci number in which its magnitude is increased. The embedding will be stopped when there are no watermark bits to embed. Comparison performance is needed to find out the better transform method than FFT for audio steganography application. In [10] authors use Differential SVD stego outcome with chaotic maps for secure digital audio communication. This audio steganography system involves two levels of security. Cat map or baker map is used to provide the first level of security while optical stego outcome based on Differential SVD is used to provide the second level of security.

In [11] authors used blocks in the process of data embedding utilizing the psychoacoustic models using MDCT, Embedding is done in the frequency domain by first selecting signal in the time domain and the frequency domain using Gamma tone filter bank as well as selecting the watermark area at the time domain by applying a certain threshold. In order to handle synchronization, bit insertion is applied, so that the watermark is resistant to resynchronization. In [12] authors used combination of DCT–SVD method to insert data utilizing the Fibonacci sequence of numbers on a subcarrier signal after converted to the frequency domain with DCT. Embedding is utilizing the DCT techniques and changes in host audio signal into a frequency domain and then the distribution of the frame and embed watermark with a modified spectrum in selected samples using Fibonacci numbers. This paper describes the process for evaluating steganography technology comparing SVD and DCT performance as transform method for embedding watermark. In [13] authors replacing FFT by GBT and finding out the performance difference between FFT and GBT. There are also slightly modifications on framing or segmentation of host audio. This paper proposes framing location before FFT or GBT process. The merger of three methods of transformations such as DWT-GBT-SVD in the process of insertion of data in the audio steganography. In [14] authors used Fourier and Curvelet transform to protect the audio signals during transmission. Permutation is applied to the input audio signal after dividing it to square blocks. Then, shuffling process using Arnold transformation and Fourier map is performed to each permutated block. The key matrix which employed for encrypting the shuffled blocks is generated finally by utilizing a hybrid of Curvelet Cat map

and Fourier map. In [15] authors applied secret keys and transform domains in order to perform the permutation and substitution operations on voice samples. Lifting Wavelet Transform is employed for samples permutation while Wavelet map is employed for key generation which used in the sample‟s substitution.

In [16] authors contrived a method for speech signals ciphering relies on chaotic shift keying. The original speech signal is divided to four layers. Each layer is permutated then using four different chaotic maps. The mechanism of chaotic shift keying assigns Logistic map, Tent map, Quadratic map, and Bernoulli‟s map to shuffle the speech samples for the first, second, third and last layers respectively. The final permutation process is performed by utilizing Chen map for further security. In [17] authors connoted a new system that depends on quantization index modulation coding, hybrid chaotic shift transforms and chaotic maps for encrypting the audio signals. Two dimensional modified Henon map is used in this system to implement HCST whereas DNA technology is used to improve the system security. In [18] authors proposed a semi blind speech steganography scheme using DCT and SVD and tested the algorithm by taking several videos and applying different attacks for robustness and imperceptibility. In this paper, a digital video steganography technique is proposed using DWT and DCT +SVD. The data chosen as watermark is inserted or embedded in each frame of the video and later extracted from the video. The watermarked video is subjected to a set of attacks and the watermark is extracted. In [19-20] authors suggested an algorithm that consists of two security layers: RSA and DNA in order to cipher speech signals. DNA technique is used to add more layer of security over RSA. The analysis of security for the introduced algorithm indicates moderate results.

## 3. Proposed hybrid speech steganography

### IPDP-MLE approach

Figure 1 demonstrate the IPDP-MLE approach for pause detection and removal from cover speech signal. This method mainly consists of three phases of operation such as mean and root mean square error (RMS) calculation phase, MLE based continuous speech analysis with pause occurrence estimation phase and DTX algorithm-based pause removal phase. The operation of each phase as follows:

**Phase 1:** Human speech voice signal is considered as the input to the system. The speech signal pitch values will be changed continuously with the time. The speech signal contains by default the white Gaussian noise properties. To remove this noise, mean (μ) and standard deviation (σ) calculated from the speech signal pitch levels.

$$\mu = \frac{1}{1600} \sum_{i=1}^{1600} (i) \quad (1)$$

$$\sigma = \sqrt{\frac{1}{1600} \sum_{i=1}^{1600} (x(i) - \mu)^2} \quad (2)$$

Here, the input speech signal $(i)$ consisting of 1600 samples, respectively. Here, the standard deviation is calculated for the estimation of average distance of speech from mean. So, by using this standard deviation average levels of pauses occurrence will be easily identified. These mean and standard deviation are also used to differentiate pause from speech. Because the frame with pause signal contains the low standard deviation compared to the speech signal standard deviation levels and repeat the procedure for entire speech signal.

**Phase 2:** The speech input can be varied with respect to conventional staircase-based variations. However, better results can be obtained by varying the speech parameters based on the MLE. MLP is a unique approach depending on two aspects namely Stimulus selection policy and MLE. In this case, the mean and standard deviation levels parameters are hypothesized so as to arrive at an efficient result. The mean and standard deviation levels are applied as input to the MLE algorithm. The MLE is majorly responsible for calculation of pauses with their occurrences. The Speech signal relied on a generic MLE based adaptive N-altered forced choice detection approach for the estimation of absolute pause occurrence threshold.

$$[\Omega, x] = [\Omega, x] \quad (3)$$

Here, $[\Omega, x]$ is likelihood function, to carry out this kind of search, a likelihood function need to be defined which allows us to compare incomplete paths of varying length. The likelihood used for incomplete paths during the search. A search based on this likelihood function is easily implemented by having a stack in which entries of the form are stored.

P [Ω, x] is probability density function (PDF); it is defined as the

$$P\,[\Omega, x] = \frac{1}{2\pi\sigma}_{2}\ {}^{2\sigma^2}[\frac{\Omega, x}{}] \quad (4)$$

Furthermore, the probability density function in above equation can be transformed into $y$ to obtain likelihood function. Here, we use Jacobi transformation method to find pdf of $y$. $\boldsymbol{\Omega}$ is an estimator which is a function of $x$. The estimator equilibrium vector y is given as follows.

$$y = (1 - \mu)^{-1}(X\sigma + \varepsilon) \quad (5)$$

We assumed that errors are normally distributed, $\varepsilon \sim (0, \sigma)$. For these errors are normally distributed. Finally, by solving the Maximum Likelihood Function is obtained as follows:

$$[\Omega, x] = [x|\Omega] = P^x(1 - P)^{1-x} \quad (6)$$

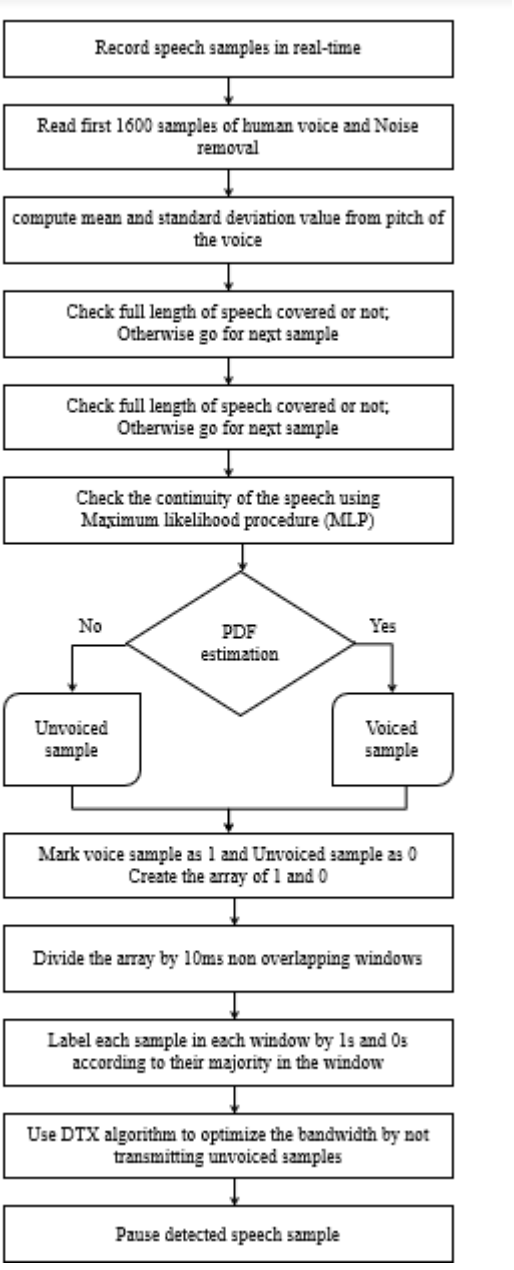



Fig. 1: Procedure of pause detection using IPDP technique.

**Phase 3**: The Vocoders are bandwidth is optimized by using Discontinuous transmission (DTX) algorithm as shown in figure 2 to eliminate the pauses effectively in mobile applications. This is an addition to voice activity detection (VAD)/VBR operation, and it is a method that suspends transmission in case a pause in the normal flow of conversation when is detected in the device and background noise is stationary. DTX, also known as silence indicator frame, is composed of the VAD and comfort noise generator (CNG) and zero crossing rate (ZCR) algorithms. It is used to reduce the transmission rate during silence periods of speech. The purpose of VAD is to detect whether the audio being encoded is speech. Two situations are possible:

- Presence of CNG algorithm: VAD conveys the proper information to the CNG algorithm.
- Absence of CNG algorithm: non-speech periods are encoded with enough bits to reproduce the background noise.

VAD is always implicitly activated when encoding in VBR. CNG algorithm allows the insertion of an artificial noise during silent intervals of speech. The purpose of the CNG algorithm is to create a noise that matches the actual background noise with a global transmission cost as low as possible.

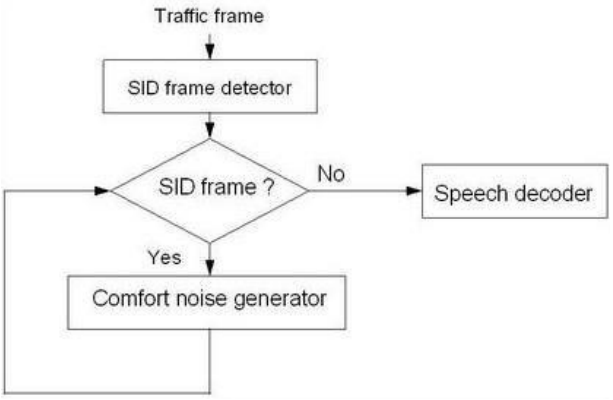



Fig. 2: pause removal using DTX operation.

Packet loss can be burst in nature. The average packet loss rate for a call may be low; however, these periods of high loss rate can cause noticeable degradation in speech quality. PLC is a technique used to mask the effects of lost or discarded packets. PLC algorithms involve either replaying the last packet received or some more sophisticated algorithm that uses previous speech samples to generate speech. PLC is effective only for small numbers of consecutive lost packets and for low packet loss rates. Perceptual enhancement is a set of optional post processing which can attempt to enhance the quality of the signal and to reduce the perception of the artifacts produced by the coding/decoding process. An example of such processing is bandwidth extension.

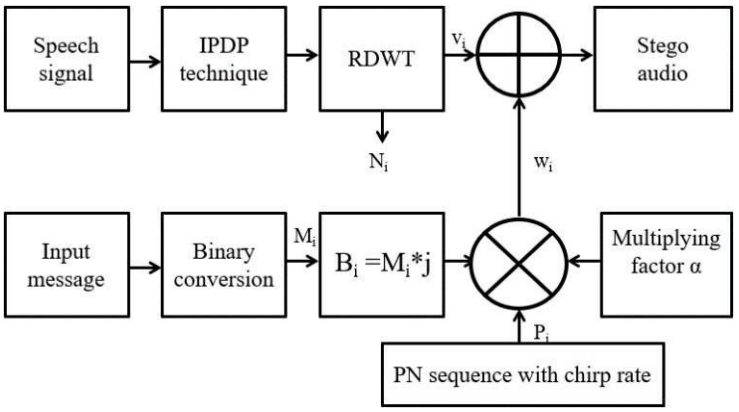
**SS-RDWT approach**

This section describes the proposed SS-RDWT approach for hybrid speech steganography system, which utilizes the pause detected audio signal as a cover speech obtained from IPDP-MLE approach. Figure 3 illustrates the block diagram of proposed hybrid speech steganography using SS-RDWT approach, where embedding is shown in Fig. 3(a) and extraction process if demonstrated in Fig. 3(b), respectively.

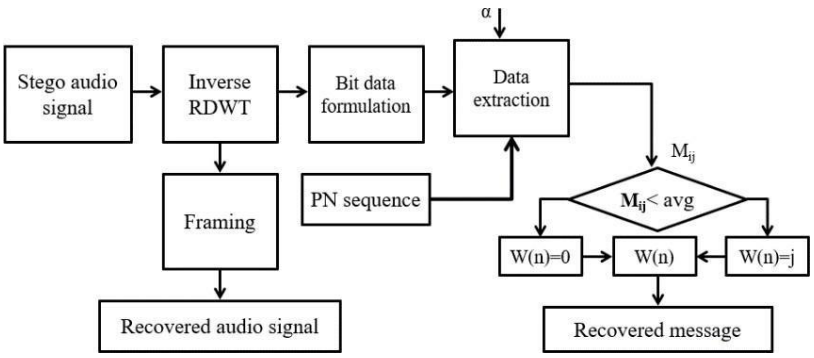
**Embedding process**

**Step 1:** First read a cover speech signal i.e., pause removed speech signal obtained from IPDP-MLE approach.

**Step 2:** Then apply RDWT (shown in Fig. 4) to decompose the cover speech into approximate and detail coefficients.




(a)




(b)

Figure 3: Block diagram of proposed speech steganography. (a) embedding process. (b) extraction process.

**Step 3:** Input the secrete text message which is to be embedded into the cover speech. Next, convert it into binary data and it is reshaped to the 1D Vector ($M_i$) of length $m \times n$.

$$= \{[0,1], 1 < i < (m,n)\} \quad (7)$$

Usually, the message signal is purely in the real domain, however the decomposed speech signal of RDWT contains both real and imaginary data. So, it is complicated to perform the steganography operation between these different types of data. Hence, it is better to hide the message information in imaginary part of speech signal to avoid the errors and losses and to provide the better imperceptibility properties.

**Step 4:** Now use the concept of spread spectrum, where the pseudo noise (PN) sequence is generated with the same properties of cover speech signal. Thus, the message information will be hidden into the speech in appropriate locations identified by the PN sequence, respectively. Here, PN sequence considered in the range of -1 to 1 like the speech signal. Then, it generates the chip rate ($cr$) for converting the message to the speech rate which can also be referred as spread spectrum nature.

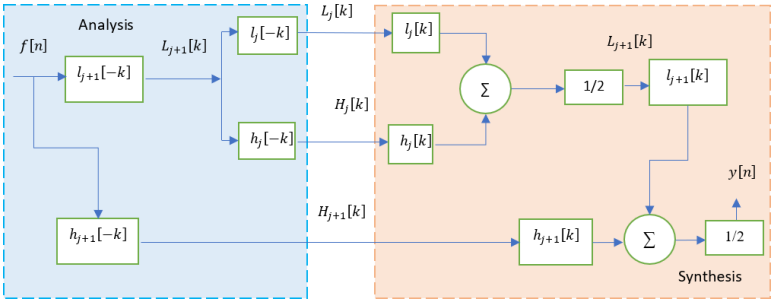



Fig. 4: Decomposition and reconstruction operations using RDWT process.

Thus, the message signal is converted into imaginary domain as follows.

$$= M * j \quad (8)$$

The PN sequence is given by.

$$= \{p_i | p_i \in \{-1,1\}\} \quad (9)$$

It performs the modulation operation by multiplying the message data $B_i$ with $cr$ times and generates the spread-spectrum output signal as

$$= \{b_i | b_i = m_i, j \cdot cr \leq < (j+i) \cdot cr\} \quad (10)$$

Here, $b_i$is the bit level sampled information of message signal.

Further, multiplication factor α is used to increase the embedding strength levels, thus it is treated as the embedding strength factor, respectively and generates the final version of message signal $w_i$ as follows:

$$= \alpha \cdot {}_i \cdot p_i \quad (11)$$

**Step 5:** Finally, the RDWT detailed output ( ) added with the final version of message signal $w_i$ to generate the stego speech signal ( $'_i$) respectively.

$$v' = v_i + w_i \quad (12)$$

Here, the addition operation takes place in bit wise manner to obtain the efficient stego signal, which is immune to noise and shows high imperceptibility properties, as it is modulated with the accurate PN sequence with higher chip-rate and the embedding strength factor.

### 3.2.2 Extraction process

Figure 3(b) illustrates the detailed procedure of proposed extraction process using SS-RDWT approach from the stego speech signal, where the inverse RDWT operation is applied on stego speech signal obtained from embedding process. This process generates the sampled speech signal that does not contain any pauses. To recover the exact speech signal, apply the framing operation and add the manual pauses (delays) between each frame to generate the final recovered speech signal. RDWT generates the output as detailed and approximate coefficients. Then, perform the matrix formulation operation by using the data extraction procedure. It contains both message and speech signal properties. To separate the message and speech based singular values, perform the data extraction operation by generating the same chirp rate $cr$ and multiplying factor $\alpha$.

$$M_{ij} = \frac{v'_i - v_i}{cr.\alpha} \quad (13)$$

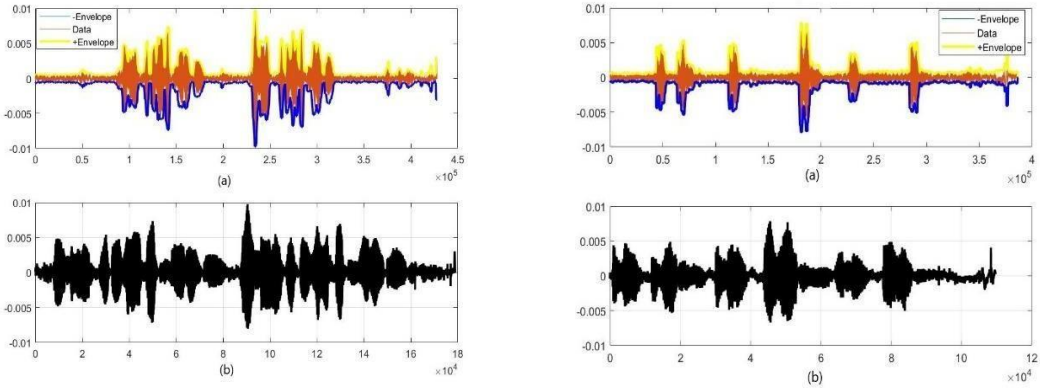
To extract the information bits from each frame, examine non-diagonal values of the matrix. It has been experimentally noticed that there are two groups of non-diagonal values that are extremely distinct, meaning that the values at the positions where bit-0 has been embedded tend to be much smaller than those values at the positions where bit-1 has been embedded. Thus, to determine the message bit $(n)$, the average of non-diagonal values is first computed; named it as $M_{avg}$, then for each non-diagonal value $M$ , $w(n)$ is extracted according to the following formula:

$$(n) = \begin{cases} 0, M_{ij} < M_{avg} \\ , \quad M_{ij} \geq M_{avg} \end{cases} \quad (14)$$

The generated message signal contains both real and imaginary data, by performing the imaginary to real conversions original data will be generated.

## 5. Results and discussion

This section describes the performance analysis of proposed hybrid speech steganography using SS-RDWT with IPDP-MLE approach. Several speech samples from various age group of male and female versions are tested with proposed SS-RDWT with IPDP-MLE approach and disclosed the superior performance as compared to existing speech steganography methods like FFT-based approach [7] and DWT-based approach [8]. In addition, it is also compared that SS-RDWT and SS-RDWT with IPDP-MLE approach for demonstrating the effectiveness of pause removal system in speech steganography applications.



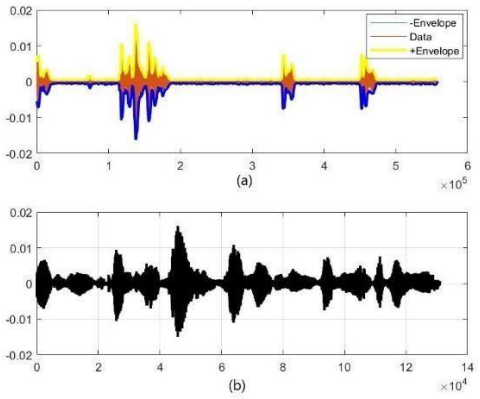


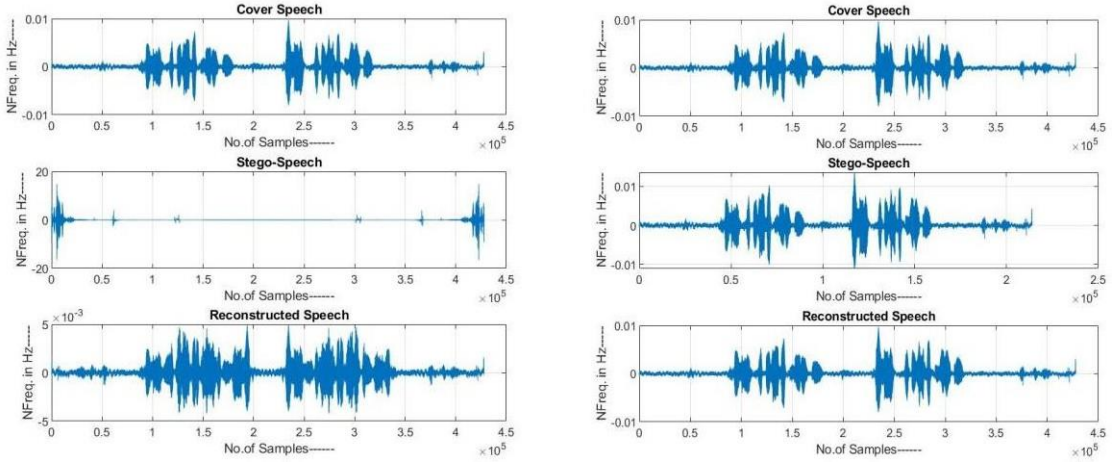

Fig. 5: Pause removal using IPDP-MLE. (a) original speech. (b) output speech after removal of pause.

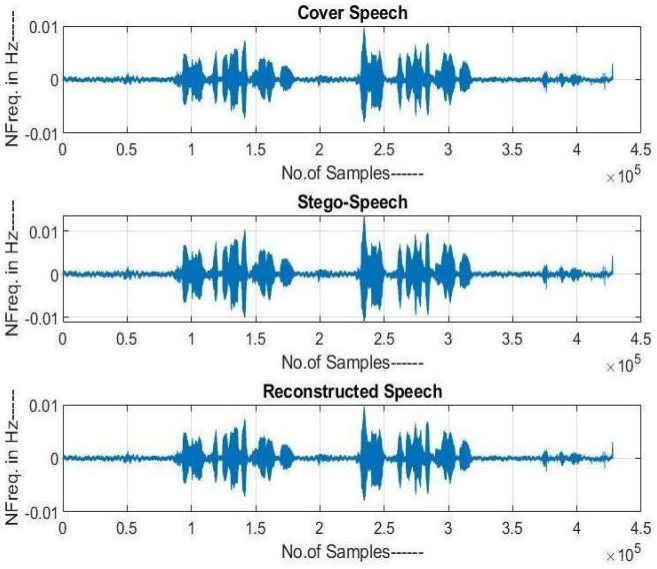
Figure 5 discloses the obtained pause detection and removal speech signals using proposed IPDP-MLE technique, where three different speech samples are tested and disclosed good enough results in terms of pause elimination. Positive and negative amplitudes of original speech samples are demonstrated in yellow and blue color font. Table 1 listed with the obtained values of eliminated pauses using IPDP-MLE technique. In addition, it is also given the values of lower and higher pitch with mean and STD of MLE, respectively.

Table 1. Obtained values of pauses removed speech using IPDP-MLE.

| | Pauses eliminated | Low pitch | High pitch | MLE of mean | MLE of STD |
|---|---|---|---|---|---|
| Sample 1 | 248977 | -8.022826e-03 | 9.780349e-03 | 5.549054e-04 | 9.072066e-04 |
| Sample 2 | 277372 | -6.708403e-03 | 7.854997e-03 | 3.555205e-04 | 5.598328e-04 |
| Sample 3 | 428050 | -1.504372e-02 | 1.606212e-02 | 4.041609e-04 | 9.053830e-04 |

Figure 6 represents the existing speech steganography performance using FFT-based and DWT-based approaches, where it is clearly visible from Fig. 6(a) that FFT-based method has failed to produce higher imperceptibility as the stego-speech completely dissimilar to the cover speech and even the reconstructed speech also not so like the cover speech. From fig. 6(b), DWT-based approach obtained better outcome as compared to FFT-based method in terms of both imperceptibility and robustness properties as the stego-speech and reconstructed speech looks quite similar with few amplitude errors. Additionally, it reduces the original size of stego-speech to half of its actual size due to the decimation operation of DWT. The performance of proposed SS-RDWT is shown in Fig. 6(c), where the cover speech, stego-speech and reconstructed speech are of equal size and looks remarkably similar which results in higher imperceptibility and robustness properties as compared to both FFT-based and DWT-based speech steganography approaches.





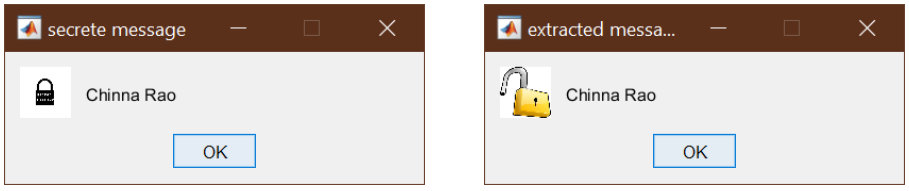

(c)

Fig. 6: Performance of existing speech steganography methods for sample 1. (a) FFT-based method [7]. (b) DWT-based approach [8]. (c) proposed SS-RDWT approach.

Table 2. CPU running time of existing and proposed speech steganography systems.

| | CPU time (in sec) |
|---|---|
| FFT-based method [7] | 22.6333 |
| DWT-based approach [8] | 19.7939 |
| SS-RDWT approach | 2.6524 |
| SS-RDWT with IPDP-MLE approach | 0.5844 |

Table 2 demonstrate the comparison of CPU running time using existing FFT, and DWT and proposed SS-RDWT approaches, which discloses that the computational complexity of proposed SS-RDWT approach is quite less as compared to existing speech steganography approaches. Figure 7 discloses the secrete and extracted message information using proposed SS-RDWT with IPDP-MLE approach. Figure 8 depicts the performance of proposed hybrid speech steganography using SS-RDWT with IPDP-MLE approach for sample 1, where the cover speech, stego-speech and reconstructed speech are of same size and had higher imperceptibility and robustness as compared to SS-RDWT approach without IPDP-MLE technique. For instance, Table 2 listed with the CPU running time of proposed SS-RDWT with IPDP-MLE approach, where the results are obtained in just 0.5844 sec, which is more than 4 times lesser to SS-RDWT without IPDP-MLE technique. Further, Table 3 demonstrates the size and storage memory of cover speech, stego-speech, and reconstructed speech signals with and without IPDP-MLE technique. As given in Table 3, both the size and storage memory obtained using IPDP-MLE technique is quite lesser than without IPDP-MLE, which means that proposed hybrid speech steganography using SS-RDWT with IPDP-MLE approach requires lesser storage bandwidth and achieves high-speed communication as the storage memory is very less.




(a) (b)

Fig. 7: Secrete message. (a) embedded message. (b) extracted message.

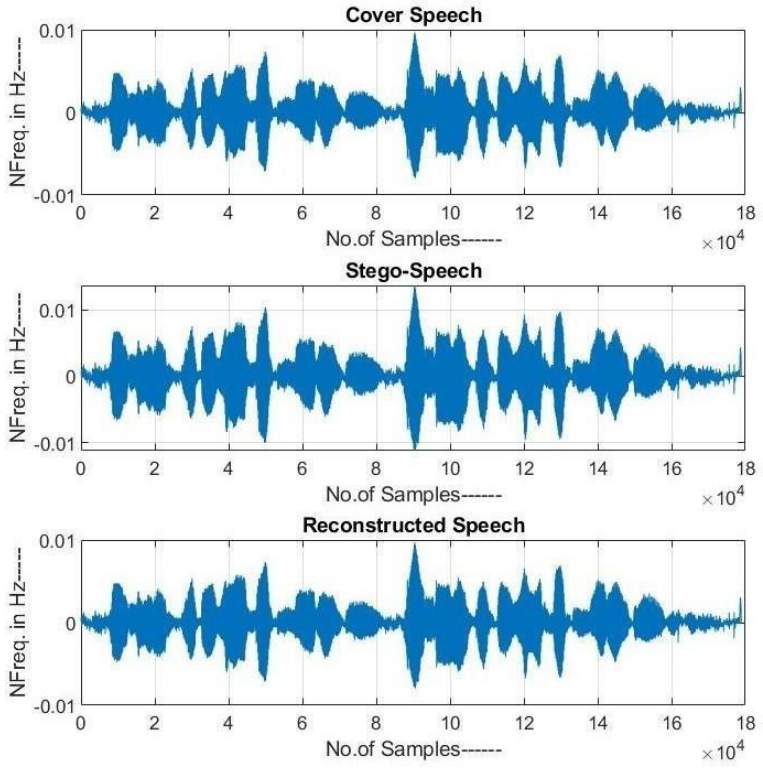


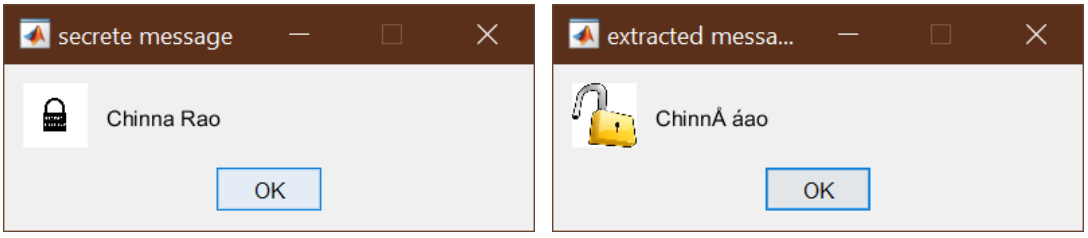

Fig. 8: Performance of proposed hybrid speech steganography using SS-RDWT with IPDP-MLE approach for sample 1.

Table 3. Comparison of storage memory for cover, stego and reconstructed speech signals with and without IPDP-MLE technique.

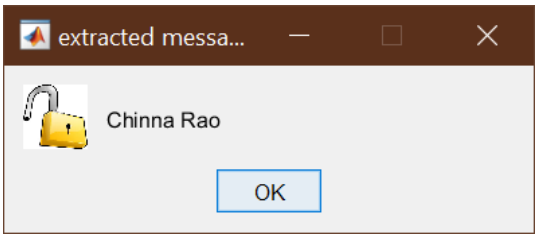
| | Without IPDP-MLE | | With IPDP-MLE | |
|---|---|---|---|---|
| | Size | Bytes | Size | Bytes |
| Cover speech | 427990x1 | 3423920 | 179014x1 | 1432112 |
| Stego-speech | 1x427990 | 6847840 | 1x179014 | 2864224 |
| Reconstructed speech | 1x427990 | 3423920 | 179014x1 | 1432112 |

**Robustness against noise attack**

The noise effect is an important issue that needs to be considered to assure the speech steganography system efficiency. Therefore, the speech steganography system performance is assessed in terms of extracted secrete message by adding random impulse noise to the stego-speech. Figure 9 depicts the extracted secrete messages against noise attack, Fig. 9(b) shows the extracted message using DWT-based approach which has higher error rate whereas the proposed SS-RDWT with IPDP-MLE approach extracted secrete message without any error rate. In addition, bit error rate and correlation coefficient is computed for both existing and proposed hybrid speech steganography approaches and demonstrated in Table 4.




(a) (b)




(c)

Fig. 9: Extraction of secrete message against noise attack. (a) embedded message. (b) extracted message using DWT-based approach. (c) extracted message using proposed SS-RDWT with IPDP-MLE approach.

Table 4. Comparison of correlation coefficient and BER values against noise attack.

| Method/parameter | BER | | CC |
|---|---|---|---|
| | Without noise | With noise | |
| FFT-based method [7] | 0.000452 | 4.25 | 0.917 |
| DWT-based approach [8] | 0.0000001 | 0.00145 | 0.972 |
| SS-RDWT with IPDP-MLE approach | 0 | 0.0000054 | 0.9998 |

## 5. Conclusion

This article focused on development of hybrid speech steganography system using SS-RDWT with IPDP-MLE approach, where the IPDP-MLE technique employed to reduce storage capacity, computational time, and even power usage by removing pauses from cover speech signal. In addition, it is also provided the scenario of speech steganography performance with and without pause removal. Further, CPU running time is computed to disclose the effectiveness of proposed pause removal-based speech steganography system. Furthermore, proposed SS-RDWT with IPDP-MLE approach tested against noise attack and performed superior as compared with existing speech steganography approaches in terms of both BER and CC metrics. Finally, extensive simulation results demonstrated that proposed hybrid steganography using integrated SS-RDWT with IPDP-MLE approach obtained enhanced performance as compared to state-of-art approaches. The proposed system can be employed in applications such as secured telephone communication, narrowband radio systems, real time speech stego systems as well as secure transferring of confidential data throughout the Internet.

**Informed consent:** Informed consent was obtained from all individual participants included in the research.

**Conflict of interests:** The authors declare that they do not have any conflict of interests.

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of authors.

**Funding Details:** No funding details

## AUTHORS CONTRIBUTION

**MANUSCRIPT TITLE: - Hybrid Speech Steganography System using SS-RDWT with IPDP-MLE approach**

All the persons who meet authorship criteria are listed as authors and all author certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept ,design, analysis, writing, or revision of the manuscript furthermore, each author certifies has this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the speech processing and Steganography Applications.

**Authorship contribution:**

**Category-I**

Concepts and design of study: [1]R Chinna Rao

Acquisition of data: [1]R Chinna Rao

Analysis and/or interpretation of data: [1]R Chinna Rao

**Category-II**

Drafting the manuscript: [1]R Chinna Rao

Revising the manuscript critically for the important intellectual content: [1]R Chinna Rao P .V.Y Jayasree

**Category-III**

Approval of the version of the manuscript to be published (the name of the all authors listed)

[1]R Chinna Rao P .V.Y Jayasree[3]S. Srinivasa Rao

Acknowledgment:
All the who have made substantial contribution to the work reported in the manuscript (eg: technical help ,writing and editing assistant support but who do not meet the criteria for the authorship ,are named in the acknowledgement, then that indicates that we have not include an acknowledgements, the that indicates that we have not received contribution from non- authors.
This statement is given by all authors:

[1]R Chinna Rao, [2]P.V.Y Jayasree, [3]S. Srinivasa Rao
[1]*Research scholar Department of Electronics and communication Engineering, GITAM university, Andhra Pradesh, India.*
*E-mail: rayudu.chinnarao@gmail.com*
[2]*Professor &HOD, Department of Electronics and communication Engineering, GITAM university, Andhra Pradesh, India.*
*E-mail:jpappu@gitam.edu*
[3]*Professor, principal, Department of Electronics and communication Engineering, Malla Reddy College of Engineering & Technology, Secunderabad, India..*
*E-mail: ssrao.atri@gmail.com*

**References**

[1]. Dutta, Hrishikesh, et al. "An overview of digital audio steganography." IETE Technical Review (2019): 1-19.
[2]. Rekik, Siwar, et al. "Speech steganography using wavelet and Fourier transforms." *EURASIP Journal on Audio, Speech, and Music Processing* 2012.1 (2012): 20.
[3]. Jayaram, P., H. R. Ranganatha, and H. S. Anupama. "Information hiding using audio steganography–a survey." *The International Journal of Multimedia & Its Applications (IJMA) Vol* 3 (2011): 86-96.
[4]. Tian, Hui, Jin Liu, and Songbin Li. "Improving security of quantization-index-modulation steganography in low bit-rate speech streams." *Multimedia systems* 20.2 (2014): 143-154.
[5]. Kreuk, Felix, et al. "Hide and speak: Deep neural networks for speech steganography." *arXiv preprint arXiv:1902.03083* (2019).
[6]. Zhijun, Wu, and Sha Yongpeng. "An implementation of speech steganography for iLBC by using fixed codebook." *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*. IEEE, 2016.
[7]. 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.
[8]. Kumar, Pala Mahesh, and Kalyanapu Srinivas. "Real Time Implementation of Speech Steganography." *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2019.
[9]. Yang, Zhongliang, Xueshun Peng, and Yongfeng Huang. "A sudoku matrix-based method of pitch period steganography in low-rate speech coding." *International Conference on Security and Privacy in Communication Systems*. Springer, Cham, 2017.
[10]. Xue, Y., Mu, K., Wang, Y., Chen, Y., Zhong, P., & Wen, J. (2019). Robust Speech Steganography Using Differential SVD. *IEEE Access*, *7*, 153724-153733.
[11]. Wen, Juan, et al. "An SVD-based adaptive robust speech steganography using MDCT coefficient." *Multimedia Tools and Applications* (2020): 1-20.
[12]. Kanhe, Aniruddha, and Gnanasekaran Aghila. "A DCT–SVD-based speech steganography in voiced frames." *Circuits, Systems, and Signal Processing* 37.11 (2018): 5049-5068.
[13]. Amiri, Noshin, and Iman Naderi. "DWT-GBT-SVD-based Robust Speech Steganography." *arXiv preprint arXiv:2004.12569* (2020).
[14]. Singh, Manwinder, Navdeep Kaur Jhajj, and Anudeep Gandam. "Fourier and Curvelet transform based speech steganography."
[15]. Kathum, Ahlam Majead, and Saad Najim Al-Saad. "Speech Steganography System Using Lifting Wavelet Transform." *International Information Institute (Tokyo). Information* 19.10B (2016): 4633.
[16]. Gupta Banik, Barnali, and Samir Kumar Bandyopadhyay. "Novel text steganography using natural language processing and part-of-speech tagging." *IETE Journal of Research* 66.3 (2020): 384-395.
[17]. Huang, YongFeng, et al. "Steganography in low bit-rate speech streams based on quantization index modulation controlled by keys." *Science China Technological Sciences* 60.10 (2017): 1585-1596.
[18]. Kanhe, Aniruddha, and Gnanasekaran Aghila. "DCT based audio steganography in voiced and un-voiced frames." *Proceedings of the International Conference on Informatics and Analytics*. 2016.
[19]. Liu, Peng, Songbin Li, and Haiqiang Wang. "Steganography integrated into linear predictive coding for low bit-rate speech codec." *Multimedia Tools and Applications* 76.2 (2017): 2837-2859.
[20]. Tian, Hui, et al. "Detecting steganography of adaptive multirate speech with unknown embedding rate." *Mobile Information Systems* 2017 (2017).\

**About the Authors:**

**Mr R. Chinna Rao,** received his B.Tech degree in Electronics &communication from JNT University. M.Tech from Malla Reddy College of Engineering & Technology He is currently working asAssistantProfessor, Dept. of ECE in Malla reddy College of Engineering and Technology, Secunderabad, India. PhD Scholar of GITAM, Visakhapatnam,A.P.

**Dr.PVYJayasree**received her B.E degree in Electronics &Communication Engineering from College of Engineering, GITAM affiliated to Andhra University in 1989.M.E in Electronics Communication Engineering from Andhra University, Visakhapatnam in 1999.Ph.D inElectromagnetic interference and Compatibility from JNTUK, Kakinada in 2010.Presently working as a professor & HOD, Department of Electronics and communication Engineering ,GITAM University, A.P India, her Interests are Electromagnetic interference and Compatibility, Antennas and Microwaves.

**Dr. S.Srinivasa Rao**, received the B.Tech degree from Madras Institute of Technology, Anna University, and the M.Tech and Ph.D from JNTU Hyderabad,Telangana, India. Presently working as Professor and Principal of Malla Reddy College of Engineering and Technology, Secunderabad. He has 24 years of experience in the field of teaching. He is a member of professional bodies like IEEE, ISTE and IETE.

# Figures

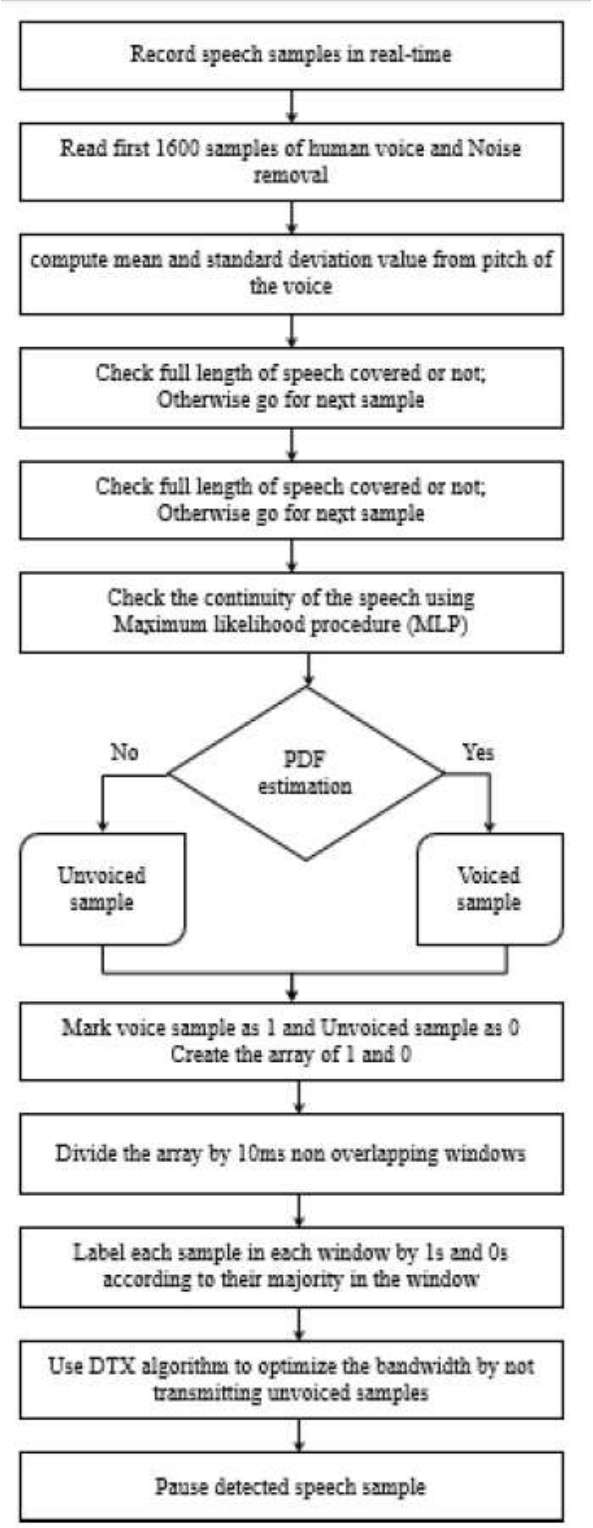



**Figure 1**

Procedure of pause detection using IPDP technique

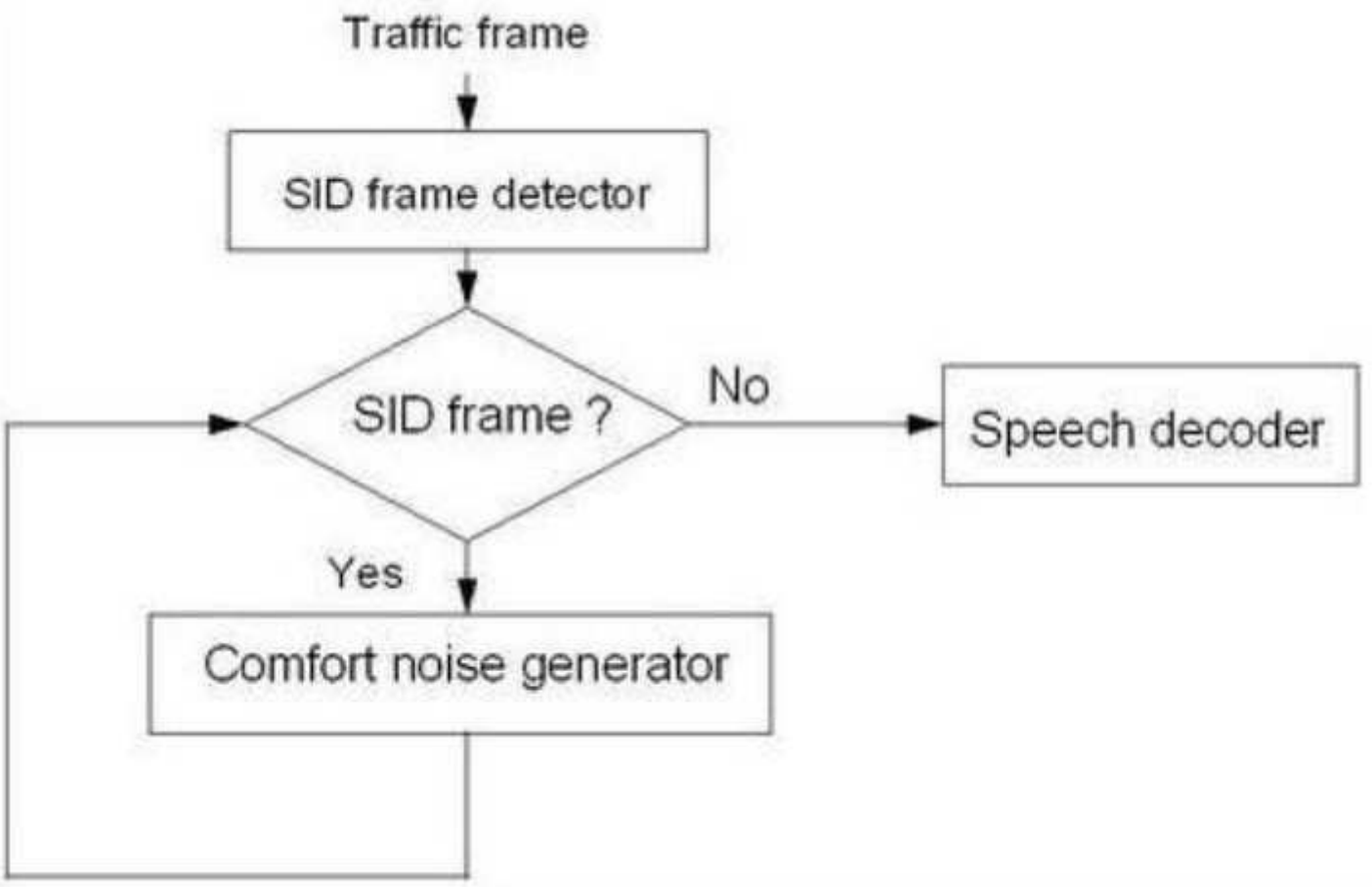



**Figure 2**

pause removal using DTX operation.

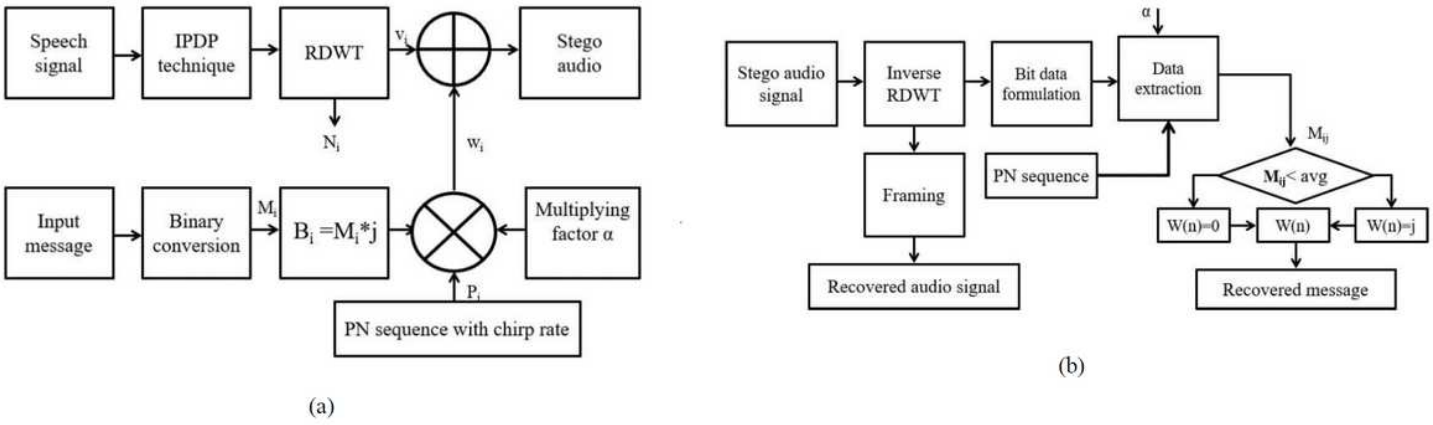



**Figure 3**

Block diagram of proposed speech steganography. (a) embedding process. (b) extraction process.

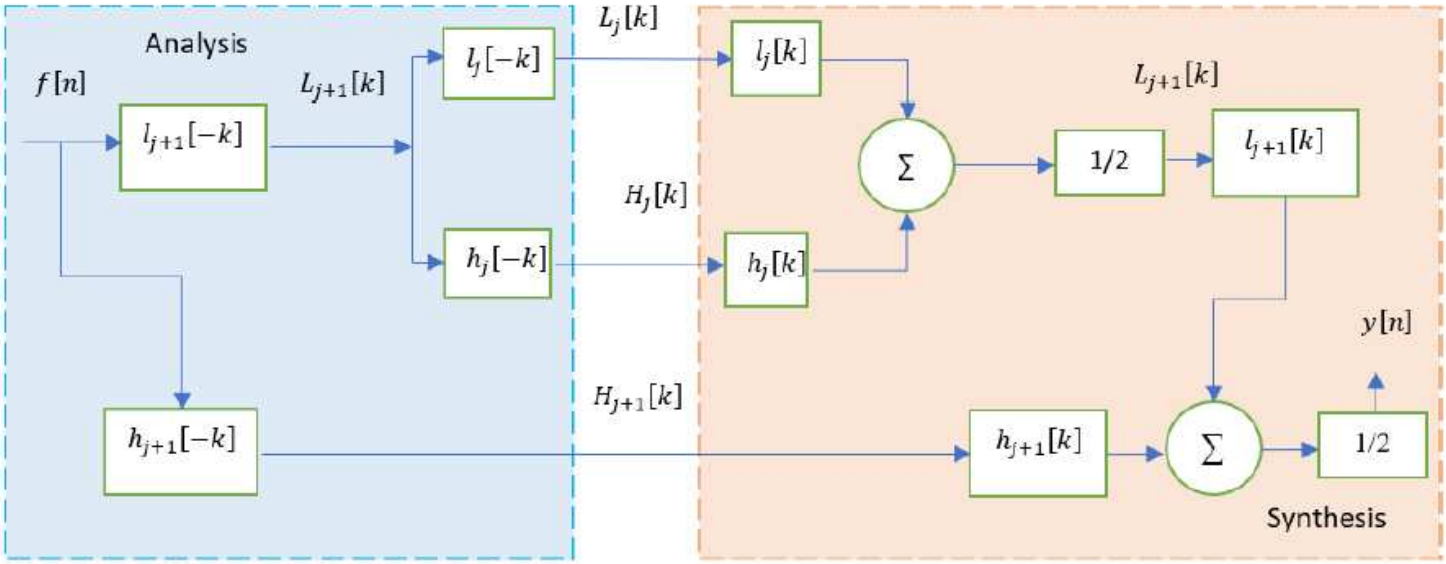
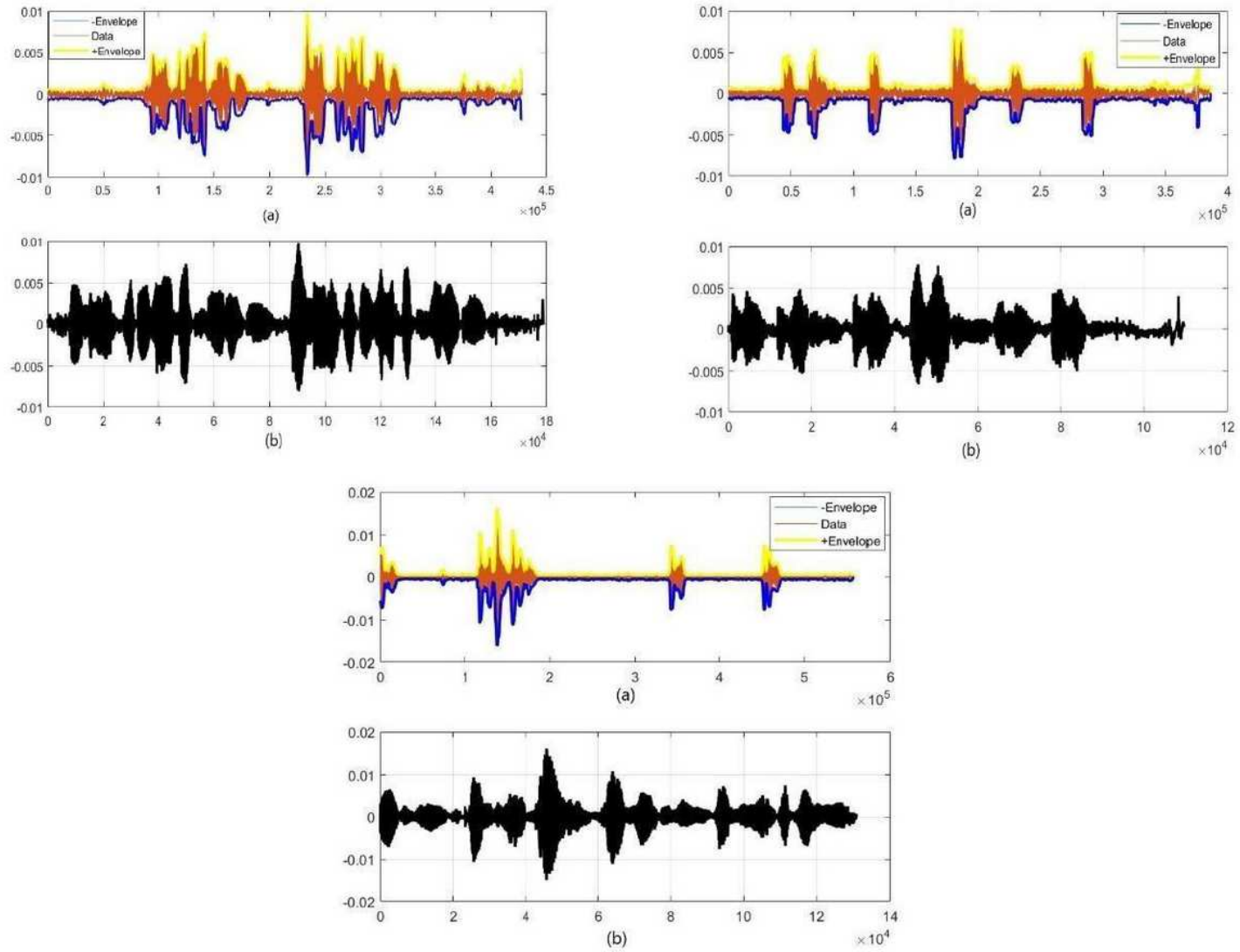



**Figure 4**

Decomposition and reconstruction operations using RDWT process.




## Figure 5

Pause removal using IPDP-MLE. (a) original speech. (b) output speech after removal of pause.

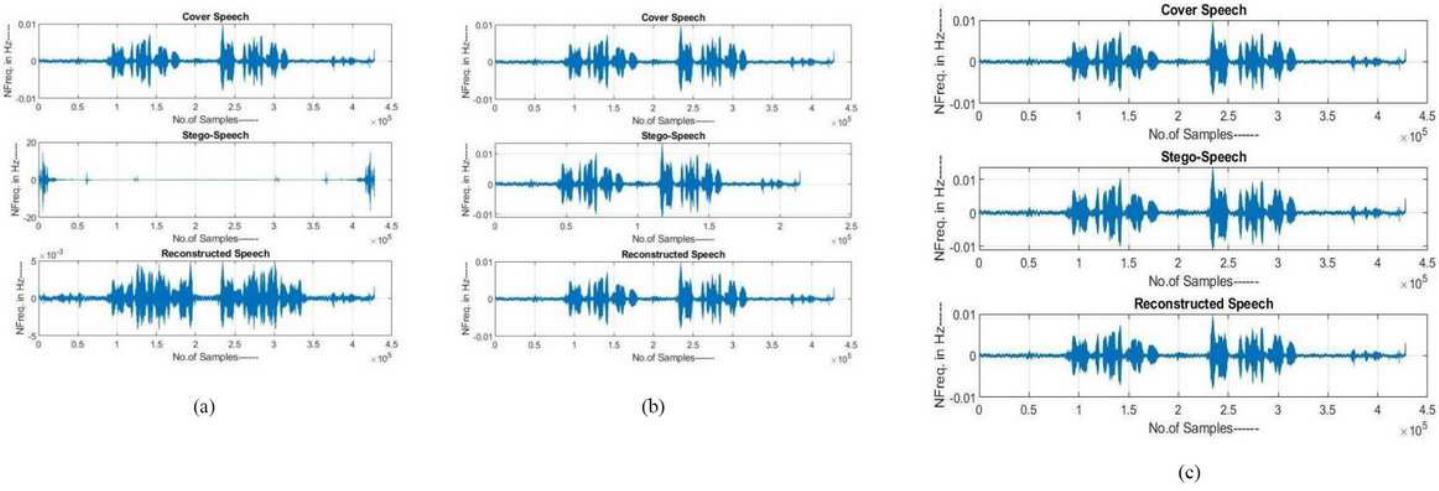


**Figure 6**

Performance of existing speech steganography methods for sample 1. (a) FFT-based method [7]. (b) DWT-based approach [8]. (c) proposed SS-RDWT approach.

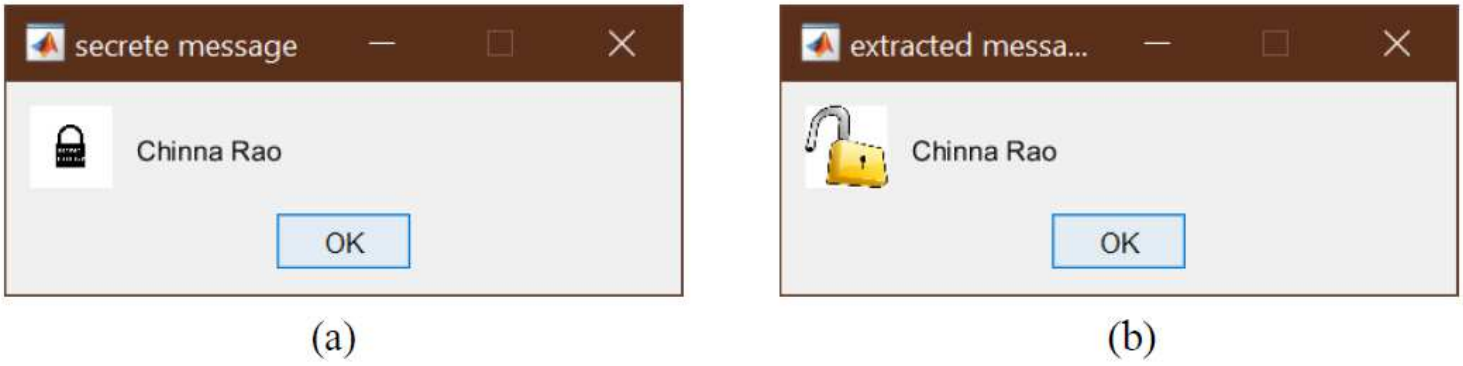



**Figure 7**

Secrete message. (a) embedded message. (b) extracted message.

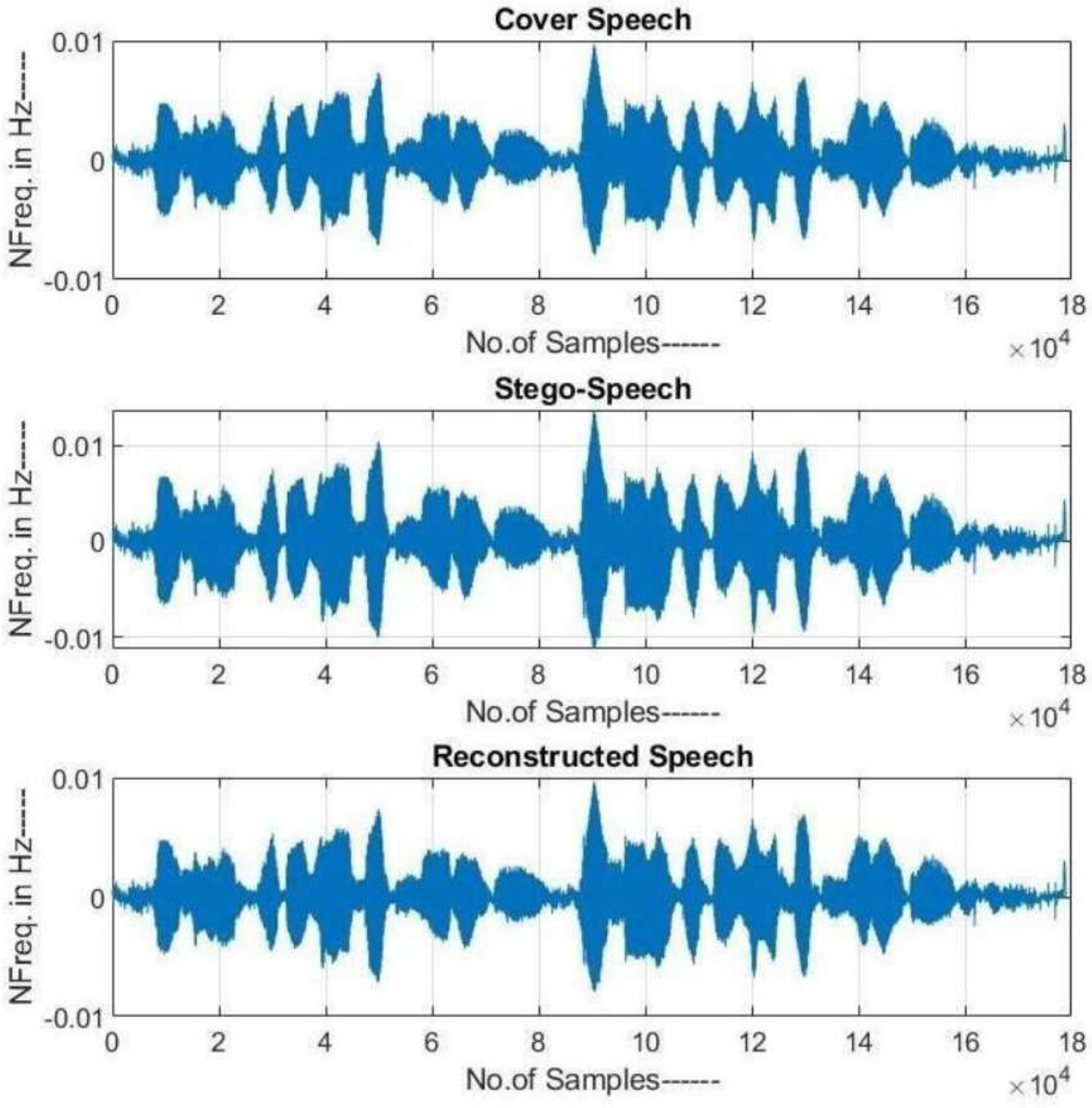



**Figure 8**

Performance of proposed hybrid speech steganography using SS-RDWT with IPDP-MLE approach for sample 1.

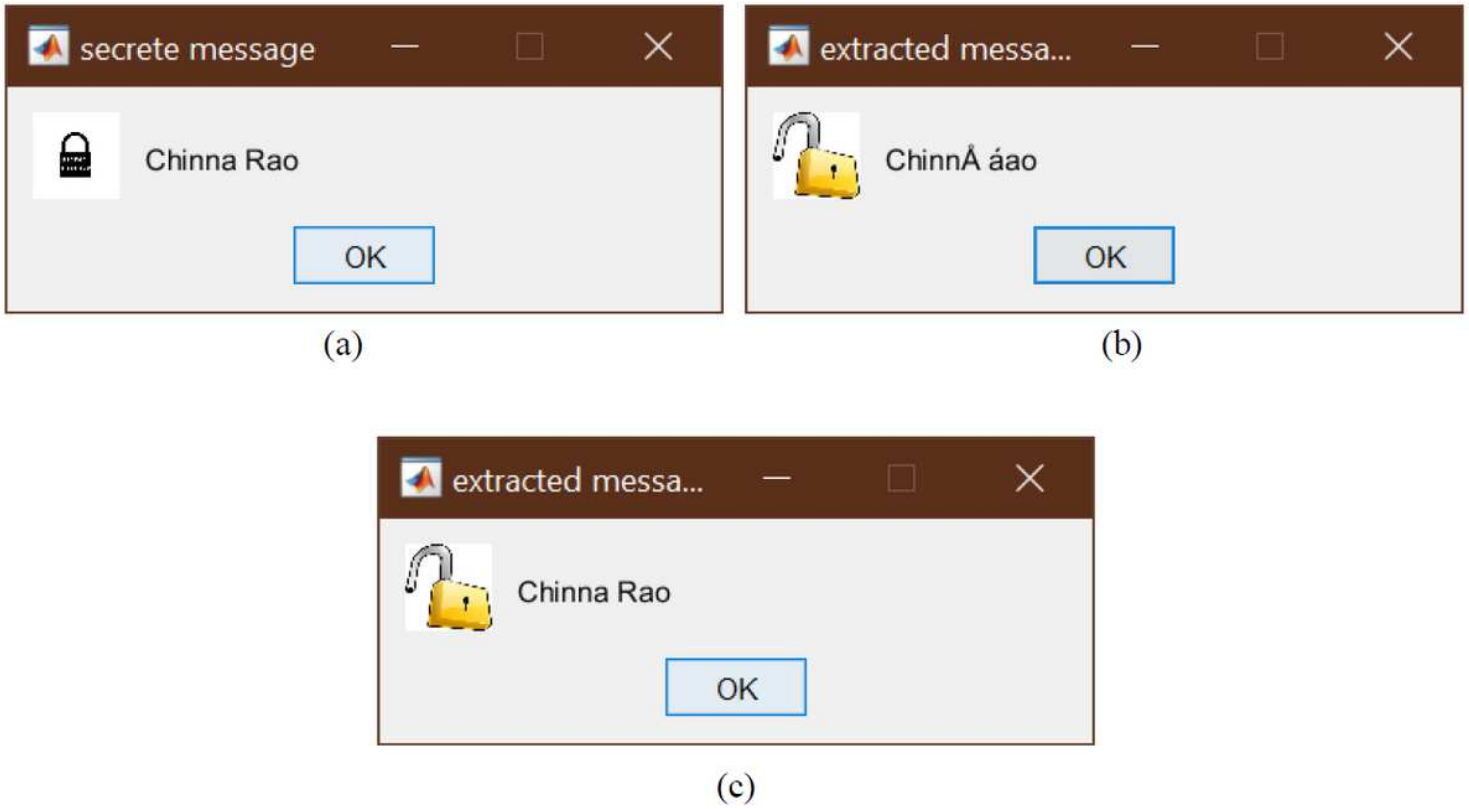



**Figure 9**

Extraction of secrete message against noise attack. (a) embedded message. (b) extracted message using DWT-based approach. (c) extracted message using proposed SS-RDWT with IPDP-MLE approach.