**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

# Speech Steganography Using DWT with IPDP-MLE Approach for Defence Applications

**R. Chinna Rao[1*], Padala Vanitha[1], Arunkumar Madupu[1], Pala Mahesh Kumar[2]**

[1]Malla Reddy College of Engineering and Technology, Maisammaguda, Secunderabad, Telangana, India

[2]AI Engineer, SAK INFORMATICS, Pragathi Nagar, Hyderabad, Telangana, India

rayudu.chinnarao@gmail.com, vanithapadala@gmail.com, arunkumar0517@gmail.com, mahesh@sakinformatics.com

**Received:** 2022 March 15; **Revised:** 2022 April 20; **Accepted:** 2022 May 10

---

**Abstract—** In digital media communication, remarkable advancements have observed from the last two decades. As it doesn't require physical media or transport, a range of benefits and business potential have provided. Because of ease of replication, unauthorized use, plasticity, and equivalence of digital works, digital media can produce various large defects for media owners. The digital media can secure through the cover writing which refers to the steganography. Accordingly, we focus on improving the hybrid speech steganography system based on FFT and DWT with IPDP-MLE approach. For embedding the secrete message, cover speech uses in speech steganography. The cover speech may contain many pauses and larger in size but it requires greater usage of power, more storage capacity, and higher computational time. It leads to the degradation of system performance. In order to eliminate the pause from cover speech signal, an intelligent pause detection protocol (IPDP) with maximum likelihood estimation (MLE) technique along with the proposed approach of FFT and DWT with IPDP- MLE. Finally, it reduces the power consumption, storage capacity, and transmission bandwidth compared to the state-of-art approaches.

**Keywords—** quantization index modulation, singular value decomposition, Discontinues transmission, root mean square error.

---

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

## I. INTRODUCTION

In the defence and military secure communication, main important key is the information security. The interception involves between communications of two parties over long distances in real-time. To overcome these issues, researchers have developed various cryptography schemes for secure communication. For making the information unintelligible to ensure that the accessing of authenticated recipients exclusively, the security achieves with the use of cryptography techniques which include the making of signal look garbled for unauthorized people. The progression of a cryptographic communication is existed in the cryptography that allows the eavesdroppers to suspect the valuable data existence. The transmitted message could get intercepted andtried to decode the secret data. This is the limitation of cryptography schemes. To avoid suspicion, steganography involves secrete communication through the masking of the secret signal with another signal i.e., cover signal unlike the cryptography. This feature fosters the researchers to make improvement on the schemes for ensuring resistance against the attackers. According to the Greek terminology, the Steganography word derives: Stego means cover and graphymeans writing. It means cover writing which is the art of hiding writing communications.

## II. LITERATURE SURVEY

The research conducted by Lin et al., (2018) [1], an efficient virtual steganalysis approach is suggested to identifyQIM steganography. A code-word correlation model drawn from RNN and also a feature classification model to differentiate between cover speech and stego speech is presented. Speech Steganography is one of the methods of information hiding for providing security to audio files. Because of the limited perceptual of human audibility, good challenges involve in the information hiding of audio file. This leads to fostering do researches in audio processing based on HAS characteristics like audio compression and audio steganography. However, several criterions affect the audio steganography performance. First criteria involve the fidelity or imperceptibility where the audio is watermarked perceptually as similar to the host audio. The objective quality is used as performance parameter for the criterion that determines based on a formula or model and subjective quality is considered from survey to several persons. In [2], the codebook techniques have used by authors for embedding the watermark data in frequency domain through the log spectrum. Additionally, LDPC and dirty paper codes have added to the technique to increase the robustness. In the embedding process, a

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

patchwork technique proposes on audio steganography based on discrete cosine transform (DCT). In [3], authors utilize the FFT transformation for providing more robustness towards the de-synchronization attacks in the audio steganography such as jitter, time scaling, and pitch. Some of the fundamental techniques of audio steganography have discussed in detail that include: SWT, echo hiding, spread spectrum, phase coding, FFT, and histogram. The audio steganography technique has proposed to achieve the best results with better imperceptibility, capacity, and excellent durability. In [4], three techniques of transformation integration were discussed for the data insertion process in the audio file. The methods are DCT and DWT. Based on the DWT quantization results, the embedding method has developed. In [5], the authors demonstrate the Sudoku matrix- based technique after processing the FFT. At each range of two Fibonacci number where the magnitude increases, the proposed embedding technique has performed. If in case watermark bits are not embedded, the technique will be stopped. To determine the better transform technique, the proposed technique compares with the FFT for audio steganography. In [6], differential SVD stego outcome with chaotic maps uses for secure digital audio communication. Two different security levels include in the system of audio steganography. For providing the first level of security and the second level of security, baker map or cat map and optical stego outcome based on differential SVD are used respectively. In the data embedding technique, blocks have utilized based on the psychoacoustic models with MDCT [7]. With the implementation of a certain threshold, embedding performs in the frequency domain through the selection of watermark area at the time domain and choosing the signal in both the frequency and time domains with the Gamma tone filter bank. Bit insertion has applied for handling the synchronization. Hence, the watermark resistant towards the resynchronization. The integration of DCT-SVD technique is used in [8] for inserting the data based on the numbers with Fibonacci sequence on a subcarrier signal after completion of converting the frequency domain with DCT. Using embedding method with DCT techniques, the host audio signal transforms into a frequency domain. [9] The frame distribution and watermark embedding with a changed spectrum using Fibonacci numbers in the selected samples are processed. The steganography evaluation has performed by comparing the performance of DCT and SVD as transform technique for watermark embedding.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

## III. EXISTING SPEECH STEGANOGRAPHY METHODS

### *A. FFT Method*

Initially, the cover speech is converted into frequency domain using FFT, because of minimize the complex computations, after whatever secret text message is embedded into cover speech signal and change into binary format by using ASCII codes. After this binary message information insert into over the channel by pseudo noise, chip rate as key and embedding mortification factor, after binary message combine with the FFT signal to generate the stego speech, extraction process is the reverse process of embedding procedure, to get stego speech, in this FFT process the stego speech seems far different the original message from stego speech, this is not except for practical applications. That is the main drawback of FFT technique.

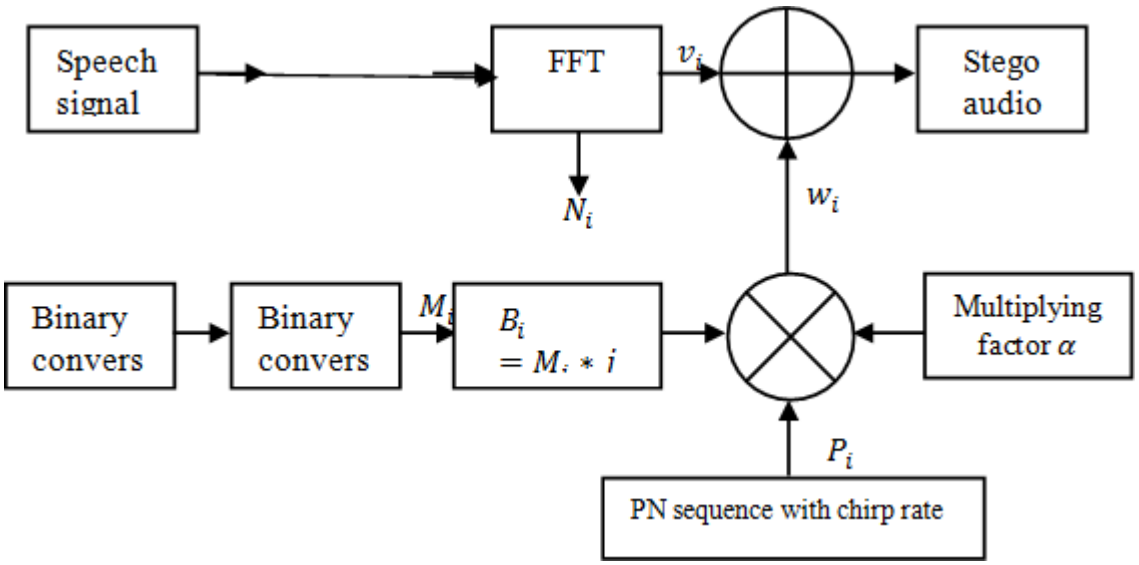


**Figure 1: Embedded processing using FFT method**

### *B. DWT Method*

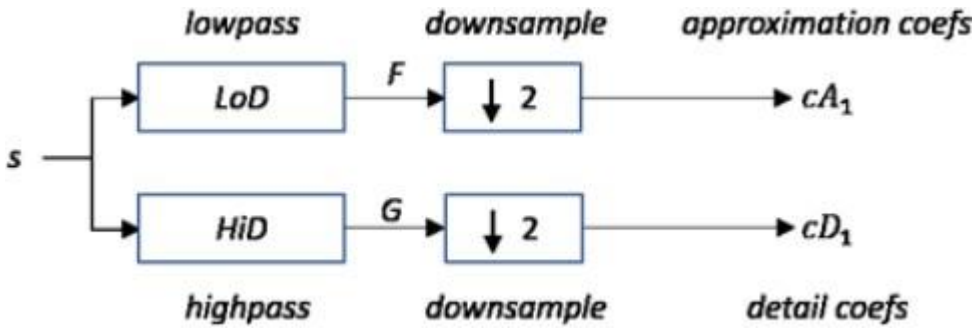


**Figure 2: Embedded processing using FFT method**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

Real time speech Steganography is implemented using decimated wavelet transform, which provides a lossless reconstruction of cover speech and hidden message due to its high spectral efficiency. Here, we considered d a speech in real-time environment, 1D-DWT is applied to a speech signal to decompose in to approximate and detail coefficients where the lower frequency sub band is referred to approximately errand higher frequency sub band cited to detail layer. But the problem presented due to DWT based methods is that it will reduce the samples of speech signal by factor 2. So, it is easy to conduct the Steganography operation, but major drawback is that "the recovered speech signal contains low quality and not immune to attacks".

## IV. PROPOSED METHODOLOGY

This section deals with the implementation of Hybrid speech Steganography with IPDP-MLE based pause removal process. Figure 3 presents the detailed implementation of proposed method. The proposed method deals with the two phases of operation. Initially, the pauses from the cover speech signal will be removed and then the message signal is stored into cover speech signal by applying the DWT approach, respectively.

**IPDP-MLE Pause removal:**

**Phase 1:** the input as human speech voice signal. However, the pitch values of input signal will change continuously over the time. The white Gaussian noise properties contain the speech signal by default. For removal this noise, we estimate the $\mu$ and $\sigma$ values from the pitch levels of speech signal.

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

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

Where, the input speech signal (i) includes 1600 samples and the standard deviation determines to estimate the average distance of speech from mean.

**Phase 2:** In regard to the conventional staircase-based variations, the input speech signal can vary. Better output results can retrieve for the varied speech parameters using MLE. Where, [Ω, x] refers to the likelihood function that requires to define for comparing the incomplete paths of varying length. Based on the likelihood function. P [Ω, x] indicates the PDF which defines as

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

follows:

$$\underline{\Omega,}$$

With the multiplication of message data Bi with $cr$ times, the modulation is operated and is generated the spread- spectrum output signal as follows:

$$[\Omega, x] = \frac{1}{2\pi\sigma}$$

$$2^{[2\sigma 2]} \quad (2)$$

By assuming the normal distribution of maximum likelihood function obtains as follows:

$$[\Omega, x] = [x|\Omega] = P(1-P)^{1-x} \quad (3)$$

$$Bi = \{b_i | b = m_i, j \cdot cr \leq i < (j+i) \cdot cr\} \quad (7)$$

Here bi means to the message signal data with the bit level sampling. Here, α is used to increase the robustness and capability of message in the speech signal. Thus, the α is

treated as the embedding strength factor, respectively. Thefinal prototype of secret message signal wi generates using

errors $\varepsilon \sim (0, \sigma)$, the

$$w_i = a \cdot \cdot p_i \quad (8)$$

**Step 5:** To produce the stego speech signal $v'$), the detailed $_i$

**DWT based speech steganography:**

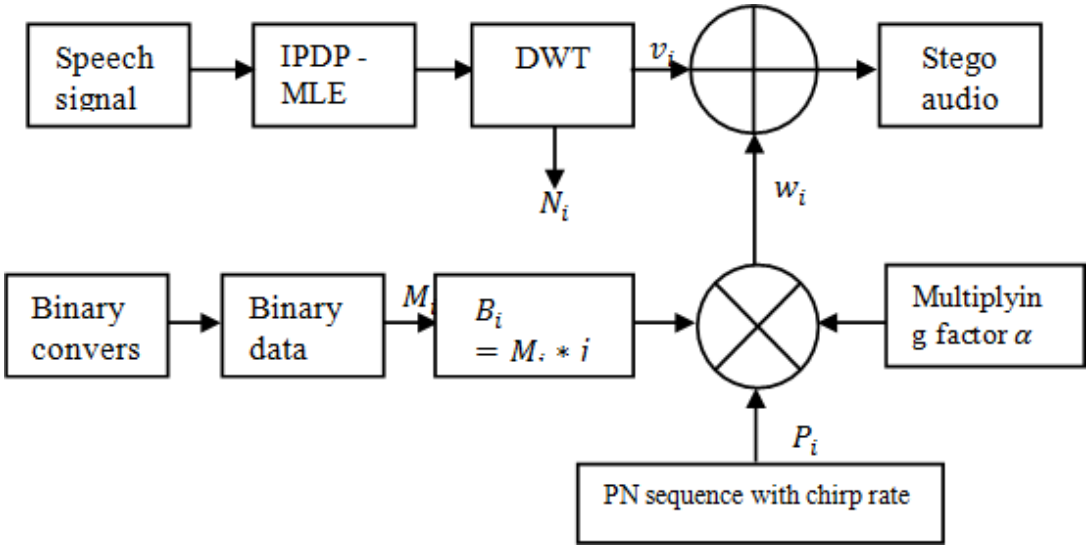


Figure 3: Proposed Stego system

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

**Step 1:** The speech signal is applied to the IPDP-MLE pause removal mechanism. Here, unvoiced speech signals such as pause signals are removed from the cover audio.

**Step 2:** the DWT operation applied on the resultant IPDP- MLE outcome, which divides the speech signal into low-low, low-high, and high-low and high-high bands. Among them, the low-low band considered for further stego process.

**Step 3:** The secret message data will be considered in this step, if any unknown symbols and encrypted text is presented in the message. Then, the message is converted into binary format, respectively. Finally, it reshapes into the 1D vector (Mi) of length $m \times n$.

$$M_i = \{[0, 1], 1 < i < (m, n)\} \quad (4)$$

**Step 4:** The spread spectrum uses in this step to maintain the equilibrium condition between the message samples to the pause removed speech signal samples. It generates the sequence of pseudo noise (PN) with the random data as many numbers of samples of pause removed cover speech. Generally, the PN sequences are in the range of -1 to 1, respectively. Thus, it allows hiding of secret message in the speech at appropriate locations. This scenario is treated as the spread spectrum communication with equal chip rate (cr). The maintaining of chip rate needs, message signal should be in imaginary domain as follows:

$$B_i =* j \quad (5)$$

PN sequence can be given as follows:

$$PN = \{p_i | p_i \in \{-1, 1\}\} \quad (6)$$

output of FFT is combined with Bi final version $w_i$.

$$v' = v_i + w_i$$

(9)

The addition operation has considered in bit wise manner for obtaining the efficient stego speech signal, here it restricts the noise and destines high imperceptibility properties. Its modules with higher chip-rate using an accurate PN sequence and an embedding strength factor.

## V. RESULTS AND DISCUSSION

All the existing techniques of speech Steganography DWT- based approaches. Additionally, the method DWT with IPDP-MLE hybrid method compare to determine the pause removal system effectiveness in the applications of speech Steganography.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

**Table 1: pauses removed sample 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 |

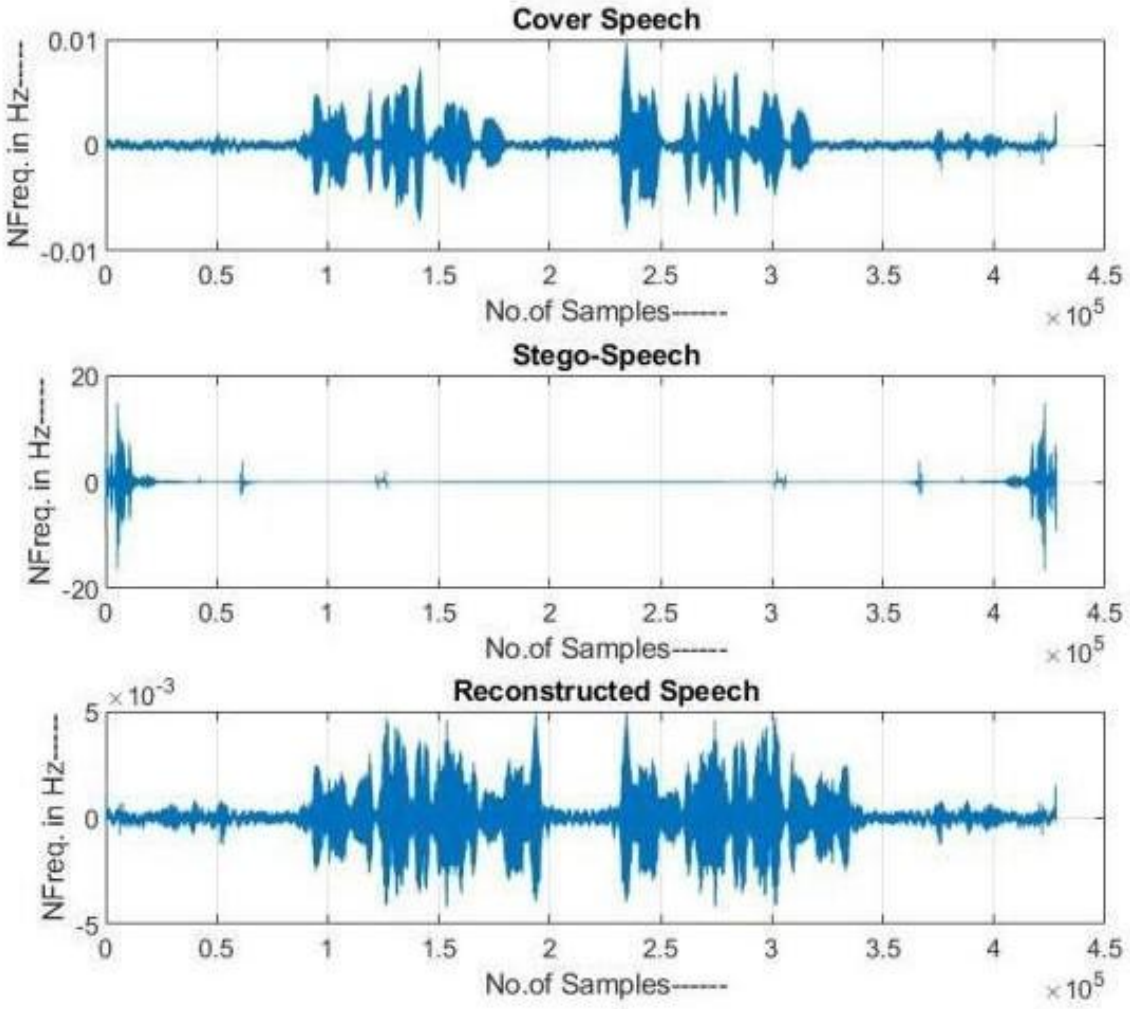


**Figure 4: FFT**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

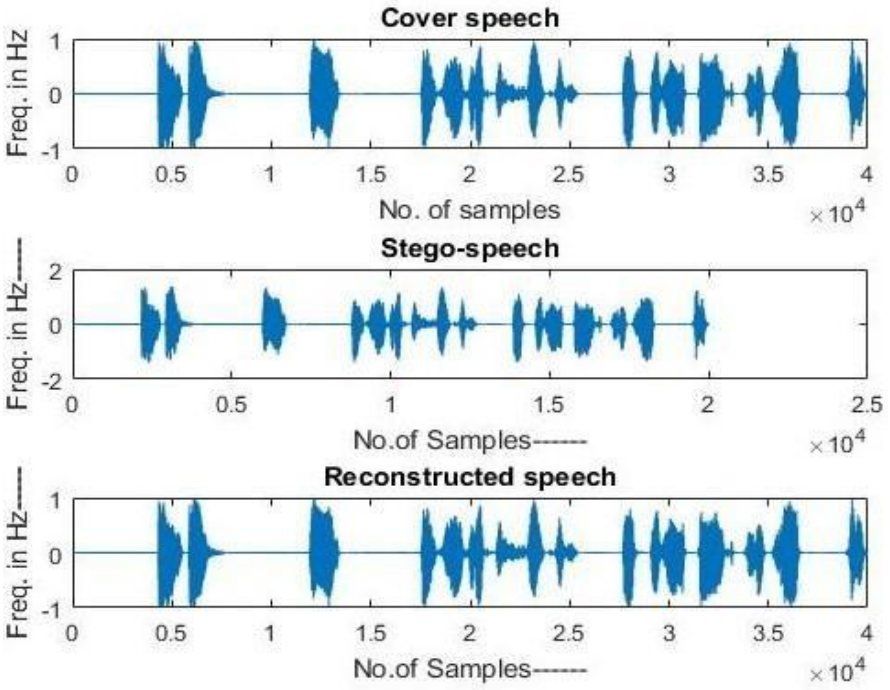


**Figure 5: DWT**

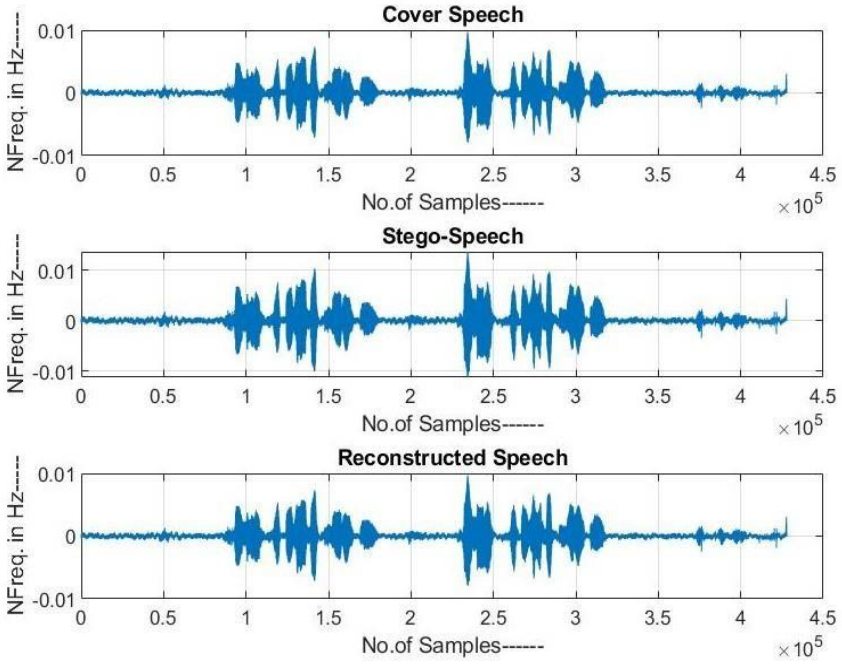


**Figure 6: Proposed IPDP-MLE DWT**

**Fig 4 (a) FFT-based method. (b) DWT-based approach. (c)Proposed DWT with IPDP-MLE approach.**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

**Table 2: memory size of cover, stego, and reconstructionspeech with IPDP-MLE and without IPDP-MLE**

| | 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 |

Table 3 The Difference of CPU running time demonstrates inbased on the approaches of FFT, and DWT. proposed DWT with IPDP-MLE approach by comparing the existing techniques of speech Steganography

Table 3: Comparison of CPU speed

| | CPU time (in sec) |
|---|---|
| FFT-based method [7] | 22.6333 |
| DWT-based approach [8] | 19.7939 |
| DWT with IPDP-MLE approach | 2.021 |

The proposed DWT with IPDP MLE approach technique discloses the less computational complexity. Figure 6 illustrates the secrete and extracted message information withthe proposed FFT with IPDP approach. The proposed method's performance depicts using DWT with IPDP for sample 1 in Figure 7. Here, sample contains the similar sizes of reconstructed speech, stego speech, and cover speech. The higher robustness and imperceptibility achieve than the FFT and DWT without IPDP-MLE approach. Table 4 shows the results of BER and CC of proposed DWT with IPDP which results 0.9998. It is more than 5 times lesser than the FFT and DWT with IPDP Table 4 demonstrates the results of that means, the lesser storage bandwidth requires for the proposed hybrid steganography that achieves high-speed communication.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

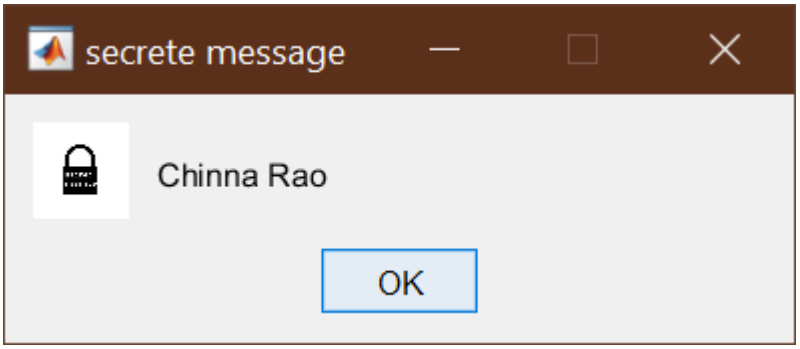


**Figure 7: Original message signal**

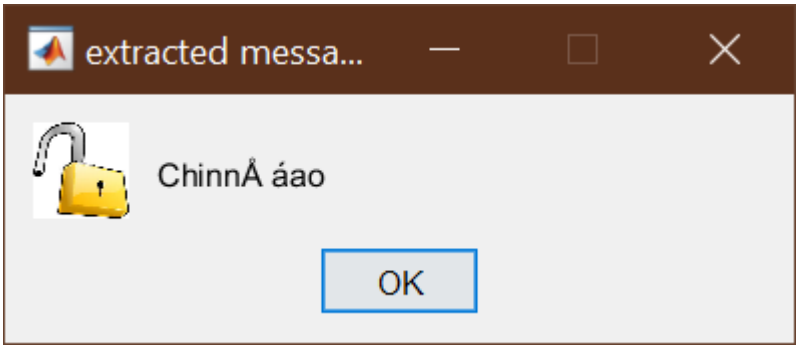


**Fig. 8: Extracted message using FFT method**

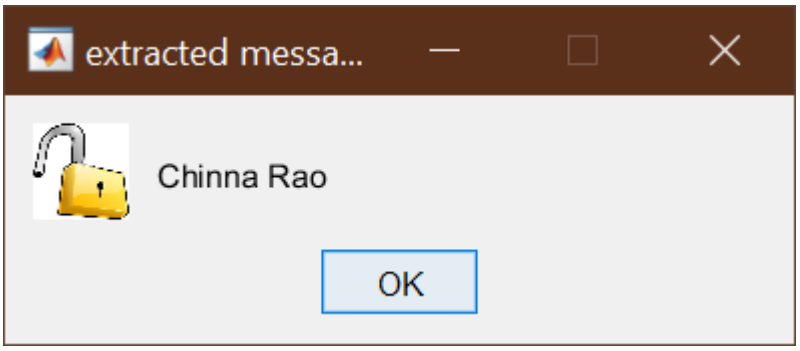


**Fig. 9: Extracted message using Proposed IPDP-MLE DWTmethod**

**Table 1: Comparing BER values and correlationcoefficient 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] | 0.0000221 | 0.00145 | 0.972 |
| DWT with IPDP-MLE approach | 0.00011 | 0.0000054 | 0.9998 |

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 3, 2022, p. 4805-4817
https://publishoa.com
ISSN: 1309-3452

## VI. CONCLUSION

The hybrid speech steganography system has improved basedon DWT with IPDP approach, in which IPDP-MLE is used to reduce the computational time, storage capacity, and utilization of power through the elimination of pauses from cover speech signal. Additionally, to analyze the proposed pause-removal speech steganography system's effective performance, CPU running time is calculated. By comparing with the previous approaches of speech steganography, the proposed system of steganography using DWT with IPDP approach shows superior performance in terms of CC and BER parameters against noise attacks. Based on the simulation results, the proposed method based on integrated approach DWT with IPDP-MLE provides better outputs than the state-of-the-art methods. The proposed method can be used in various applications involving real-time speech stego systems, narrowband radio systems, secured telephone communication, and secure transferring of confidential information over the Internet.

## REFERENCES

[1] Lin, Zinan, Yongfeng Huang, and Jilong Wang. "RNN-SM: Fast steganalysis of VoIP streams using recurrent neural network." *IEEE Transactions on Information Forensics and Security* 13, no. 7 (2018): 1854-1868.

[2] 38. 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.

[3] 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.

[4] Kumar, Pala Mahesh, and Kalyanapu Srinivas. "Real Time Implementation of Speech Steganography." *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2019.

[5] 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

[6] Xue, Y., Mu, K., Wang, Y., Chen, Y., Zhong, P., & Wen, J. (2019). Robust Speech Steganography Using Differential SVD. *IEEE Access*, *7*, 153724-153733.

[7] Wen, Juan, et al. "An SVD-based

adaptive robust speech steganography using MDCT coefficient." *Multimedia Tools and Applications* (2020): 1-20.

[8] Kanhe, Aniruddha, and Gnanasekaran Aghila. "A DCT–SVD- based speech steganography in voiced frames." *Circuits, Systems, and Signal Processing* 37.11 (2018): 5049-5068.

[9] Amiri, Noshin, and Iman Naderi. "DWT-GBT-SVD-based Robust Speech Steganography." *arXiv preprint arXiv: 2004.12569* (2020).

[10] Singh, Man winder, Navdeep Kaur Jhajj, and Anudeep Gandam. "Fourier and Curvelet transform based speech steganography."

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

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

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

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

[15] 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.

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

[17] 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.

[18] Kumar, Pala Mahesh, and Kalyanapu Srinivas. "Real Time Implementation of Speech Steganography." *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2019.