

## High-capacity steganography through audio fusion and fission

Namitha Mangikuppe Venkateshaiah, Manjula Govinakovi Rudrappa

Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering,  
Visvesvaraya Technological University, Shivamogga, India

### Article Info

#### Article history:

Received Sep 13, 2023

Revised Sep 15, 2023

Accepted Nov 21, 2023

#### Keywords:

Audio fission

Audio fusion

Audio steganography

Decoding

Discrete wavelet transforms

Encoding

### ABSTRACT

Information security is required for two reasons, either to conceal the information completely or to prevent the misuse of the information by adding watermarks or metadata. Audio steganography uses audio signals to hide secret information. In the proposed audio steganography technique, cover audio files and secret audio files are transformed from time domain to wavelet domain using discrete wavelet transform, the secret audio file is transformed in two levels, leading to secure and high-capacity data hiding. 1% of the 2-level compressed secret is fused to 99% of the 1-level compressed cover. "Peak signal to noise ratio and mean squared error, Pearson's correlation coefficient, spearman's correlation coefficient, perceptual evaluation of speech quality and short-time objective intelligibility" are considered to assess the similarity of cover audio and stego audio and similarity of secret audio embedded, and secret audio retrieved. Results show that the stego audio signal is perceptually indistinguishable from the cover audio signal. The approach also passed the robustness test.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



### Corresponding Author:

Namitha Mangikuppe Venkateshaiah

Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering

Shivamogga, Karnataka, India

Email: namithamv@jnnce.ac.in

## 1. INTRODUCTION

The world is too small as people can communicate from one corner of the world to another corner of the world within an eye blink. There is an implicit demand to develop secure information transmission across networks. Three important approaches considered in information security are cryptography, watermarking, and steganography. The intention of steganography is to reliably transmit secret data covertly by utilizing the features of the cover. Steganography is a subdiscipline of the data communication security domain. Different types of digital data including image, audio, video, text, and internet protocol (IP) datagram can be used as cover [1]. Audio steganography is recommended as audio is a famous communication and entertainment medium and hence does not raise any suspicion to intruders about the existence of data [2]–[10]. As stated in [1] the audio steganography is carried out in 3 major domains namely, i) temporal domain, ii) transform domain, and iii) coded domain.

The proposed work is related to the transform domain as discrete wavelet transform (DWT) is used to perform audio steganography. The motivation for this high-capacity audio steganography was obtained by [11]. The paper deals with image steganography (hiding image inside the image) focusing on increasing the hiding capacity. The hiding capacity is up to 200% of the cover image. Not all image steganography techniques are suitable for audio steganography. The human auditory system (HAS) is different from human vision. In the proposed method the host audio is called a cover, it can hide the secret audio, which is double its size, as we are compressing the secret audio before hiding (taking discrete wavelet transform results in compression of

the signal as well), compression increases security and hiding capacity of the method. The features of DWT such as frequency localization, energy concentration, and efficient representation make it well-suited for compressing audio signals without significantly altering the signal, hence proposed method exploits the HAS to hide secrets inside DWT coefficients and achieves 200% hiding capacity by maintaining good imperceptibility. The approach also sustained compression, change in sampling rate, and white gaussian noise attack to prove its robustness.

The idea of shuffling secret audio before hiding it inside cover media is proposed in [12], the shuffling is done through a 4-D grid multi-wing hyper-chaotic (GMWH) system. Thereby making conventional least significant bit (LSB) strong enough to protect the sensitive data. Hiding images inside audio to successfully transmit secret image covertly is considered in [13]. The secret bits are encrypted and then embedded in random LSBs in the proposed method. The results show stego audio has low signal-to-noise ratio (SNR) and high peak signal-to-noise ratio (PSNR). The method is also resistant towards steganalysis attacks. “A security enhanced robust audio steganography algorithm for image hiding using sample comparison in discrete wavelet transform domain and rivest-shamir-adleman (RSA) encryption” was proposed in [14] The approach is to select detail coefficients of samples through comparison in DWT domain by considering predetermined threshold value T and secret is embedded according to this selected coefficient which results in efficient wavelet masking technique. Experimental results gave good PSNR and high embedding capacity. “A high-SNR steganography for digital audio signal in the wavelet domain” was proposed in [15] the secret message is embedded into DWT lowest-frequency coefficients by the embedding technique along with synchronization codes and watermarks. The technique has achieved High SNR up to 39.8 dB. Hameed [16] high-capacity audio steganography based on contourlet transform (CT) was proposed, in which up to 90% of hiding capacity was achieved. After CT suitable sub bands were identified to hide the secret depending on the energy. Energy was calculated using the (1). The SNR of the stego was 79 dB. However, the analysis was carried out only by considering SNR.

$$E = \sum_i \sum_j |SC(i, j)|^2 \quad (1)$$

where,  $SC(i, j)$ -the value of contourlet coefficient.

Shahadi *et al.* [17] three levels of integer-to-integer lifting wavelet transform (LWT) was done to the cover audio and adaptive embedding positions were identified. It has achieved hiding capacity to 48% (up to 340 Kbps) along with maintaining good perceptual quality that is above 35 dB SNR. A new method for audio quality assessment was proposed in [18] which states perceptual evaluation of speech quality (PESQ) gives good subjective opinion of the voice quality. The researchers [14]–[17], [19] PESQ and short-time objective intelligibility (STOI) metrics which are more relevant to audio data were not considered for result analysis. Especially for audio it is recommended to consider PESQ and STOI as “mean squared error (MSE) is usually defined in the linear frequency scale, but the human auditory perception follows the mel-frequency scale” [20]. Therefore, in our analysis we have considered PESQ and STOI. Further in this paper, the proposed algorithm of audio steganography using audio fusion and fission will be discussed in section 2. Analysis and experimental results are depicted in section 3. Finally, section 4 draws the conclusions.

## 2. PROPOSED METHOD

The capacity, imperceptibility and robustness makes magic triangle of steganography techniques, here the proposed technique is giving the high capacity and keeps imperceptibility and robustness up right by using audio fusion and fission technique. For compact representation of the methodology abbreviations are used in this paper. Abbreviations are listed with their full form for clarity in Table 1.

Table 1. Abbreviations used and respective full forms

Abbreviation	Full forms
Sca1	Secret audio's approximation coefficients after DWT-1
Scd1	Secret audio's detail coefficients after DWT-1
Sca2	Secret audio's approximation coefficients after DWT-2
Scd2	Secret audio's detail coefficients after DWT-2
Cca1	Cover audio's approximation coefficients after DWT-1
Ccd1	Cover audio's detail coefficients after DWT-1
Stca1	Stego audio's approximation coefficients after DWT-1
Stcd1	Stego audio's detail coefficients after DWT-1

**2.1. Encoding through audio fusion**

Encoding secret audio inside the cover audio through audio fusion is depicted in Figure 1. The sender considers both secret and cover audio signals. Compresses the cover audio signal once and secret audio signal twice using DWT and hence larger secret audio signal can be embedded inside smaller cover audio signal. Compressing secret audio signal twice ensures high capacity and imperceptibility.

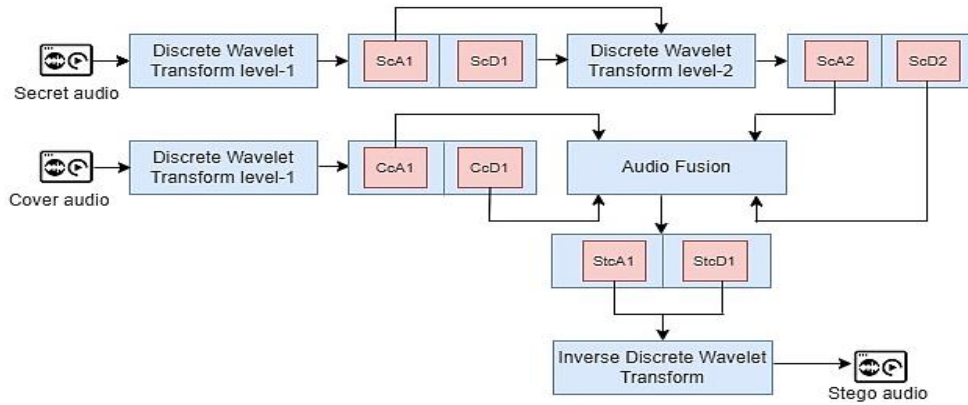


Figure 1. Encoding process at sender

The procedure to encode secret audio inside cover audio is as follows:

- i) Choose the cover audio 'cover\_audio'.
- ii) Apply DWT to the cover audio to generate approximation coefficients 'CcA1' and detail coefficients 'CcD1' accordingly.

$$[CcA1, CcD1] = DWT('cover\_audio', 'haar') \tag{2}$$

- iii) Choose the secret audio 'secret\_audio'.
- iv) Apply DWT to the secret audio to generate approximation coefficients 'ScA1' and detail coefficients 'ScD1' accordingly.

$$[ScA1, ScD1] = DWT('secret\_audio', 'haar') \tag{3}$$

- v) Apply DWT to the approximation coefficients 'ScA1' generated in step 4 to generate approximation coefficients 'ScA2' and detail coefficients 'ScD2' accordingly.

$$[ScA2, ScD2] = DWT('ScA1', 'haar') \tag{4}$$

- vi) Perform audio fusion by considering 99% of cover and 1% of secret.

$$[StcA1] = (0.99 * CcA1) + (0.01 * ScA2) \tag{5}$$

$$[StcD1] = (0.99 * CcD1) + (0.01 * ScD2) \tag{6}$$

- vii) Perform IDWT by considering 'StcA1' and 'StcD1' to produce stego audio.

$$'Stego\_audio' = IDWT('StcA1', 'StcD1', 'haar') \tag{7}$$

the stego audio formed after encoding procedure can be transmitted across the communication channel for covert communication.

**2.2. Decoding through audio fission**

The process of decoding secret at the receiver is shown in Figure 2. Decoding is very important as malfunction of decoding unit results in inefficient covert communication. The decoding process at receiver is exactly reverse process of encoding at sender side.

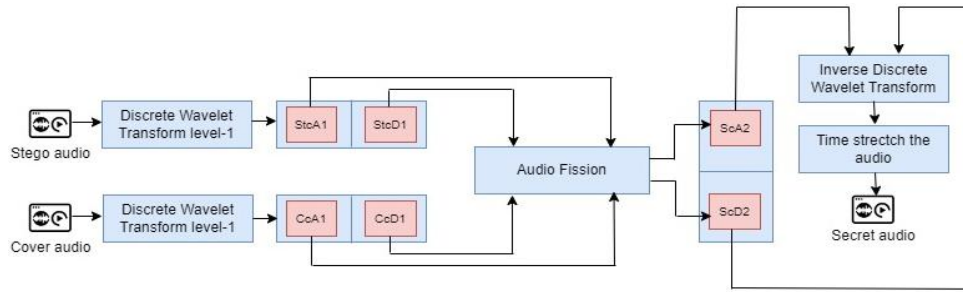


Figure 2. Decoding process at receiver

The procedure to extract secret audio from stego audio is as follows:

- i) Choose the cover audio '*stego\_audio*'.
- ii) Apply DWT to the cover audio to generate approximation coefficients '*CcA1*' and detail coefficients '*CcD1*' accordingly.

$$[CcA1, CcD1] = DWT('cover\_audio', 'haar') \quad (8)$$

- iii) Apply DWT to the stego audio to generate approximation coefficients '*StcA1*' and detail coefficients '*StcD1*' accordingly.

$$[StcA1, StcD1] = DWT('stego\_audio', 'haar') \quad (9)$$

- iv) Perform audio fission by considering 99% of cover and 1% of secret.

$$[ScA2] = (StcA1) - (0.99 * CcA1) \quad (10)$$

$$[ScD2] = (StcD1) - (0.99 * CcD1) \quad (11)$$

- v) Perform IDWT by considering '*ScA1*' and '*ScD1*' to produce stego audio.

$$'Compressed\_secret\_audio' = IDWT('ScA1', 'ScD1', 'haar') \quad (12)$$

- vi) Time stretch the '*compressed\_secret\_audio*' to obtain '*secret\_audio*'.

$$'Secret\_audio' = timestretch('compressed\_secret\_audio', 'required\_rate') \quad (13)$$

The secret audio extracted would reach the intended user covertly through public communication channel. After implementing the above stated method, the results obtained were analyzed to measure imperceptibility of stego by measuring similarity between cover and stego. Robustness of stego signal for various attacks. The results are discussed in section 3 in a detailed manner.

### 3. RESULTS AND DISCUSSION

#### 3.1. Analysis of imperceptibility

Parameters namely MSE, PSNR, pearson's correlation (PC), spearman's correlation (SC), PESQ, and STOI are considered to assess the similarity of cover audio and stego audio to measure the imperceptibility and to measure similarity of secret audio embedded and secret audio retrieved to ensure safe transmission of secret. Unlike image steganography, only MSE and PSNR are not enough for audio steganography. Covers and secrets are considered from two sources [21], [22] for the proposed method.

##### 3.1.1. Mean square error and peak signal to noise ratio

"The distortion of the 'stego' with respect to 'cover' audio is measured by using MSE". The value 1 indicates no distortion whereas 0 indicates high distortion. MSE is measured using (14).

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (14)$$

where,  $N$  is the number of signal samples,  $x_i$  is the value of the  $i^{\text{th}}$  sample in  $x$ ,  $y_i$  is the value of the  $i^{\text{th}}$  sample in  $y$ .

“The term PSNR is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation”. PSNR is measured using (15). PSNR is measured in dB and it’s a good measure for comparing restoration result for the same audio signal.

$$PSNR(dB) = 10 \log_{10} \left( \frac{R^2}{MSE(x,y)} \right) \tag{15}$$

where,  $R$  is the peak signal value that exists in the original audio signal,  $MSE$  is mean squared error.

**3.1.2. Pearson’s correlation and spearman’s correlation**

“Correlation is the degree to which two variables are linearly related. This is an important step in bi-variate data analysis. In the broadest sense correlation is any statistical relationship, whether causal or not, between two random variables in bivariate data. The values range between -1.0 and 1.0. A correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. A correlation of 0.0 shows no linear relationship between the movements of the two variables”.

In statistics, “the pearson correlation coefficient (PCC) also referred to as pearson’s  $r$  or the bivariate correlation is a statistic that measures the linear correlation between two variables  $X$  and  $Y$ ”. The formula for PC is given by (16);

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{16}$$

where,  $x_i$  is values of the  $x$ -variable in a sample,  
 $\bar{x}$  is mean of the values of the  $x$ -variable,  
 $y_i$  is values of the  $y$ -variable in a sample,  
 $\bar{y}$  is mean of the values of the  $y$ -variable.

In statistics, “spearman’s rank correlation coefficient (SCC), named after charles spearman, is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables)”. It assesses how well the relationship between two variables can be described using a monotonic function. The formula for spearman’s correlation is given by (17);

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \tag{17}$$

where,  $d_i$  is the difference between two rankings,  $n$  is the number of observations.

**3.1.3. PESQ and STOI**

Most of the researchers use MSE as loss function to compare stego and cover. Human auditory perception is not matched by MSE. STOI and PESQ are closely related to human auditory perception and widely used in speech separation research evaluation criteria. Therefore, STOI and PESQ may be better choices for the loss function.

The PESQ algorithm is designed to predict subjective opinion scores of a degraded audio sample. PESQ returns a score from 4.5 to -0.5 (sometimes 1 to 5), with higher scores indicating better quality. The STOI metric is based on a correlation coefficient between the temporal envelopes of the time-aligned reference and processed speech signal in short-time overlapped segments. STOI is relatively preferred metric as stated in [23] because of the consistency between training and evaluation procedures, it also considers human intelligibility to the evaluation metric. The MSE, PSNR, PCC, SCC, PESQ, and STOI for stego with respect to cover for the proposed method is listed in Table 2. The average PSNR, PESQ, and STOI between stego and cover are listed and compared with [24] in Table 3.

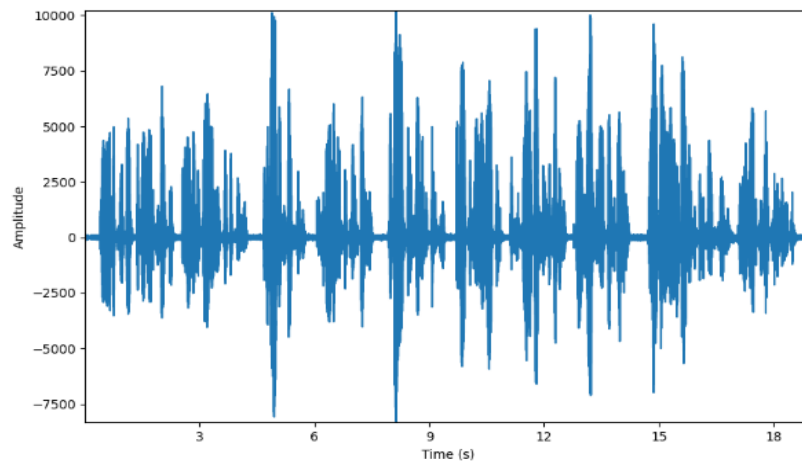
The proposed technique exhibits good PSNR and PESQ and having slightly less STOI when compared to the [24]. Figure 3 is showcasing signal plots for audio signals before embedding after embedding the secret. Specifically Figures 3(a) and 3(b) depicts waveform of cover and stego respectively, the perceptual difference is hard to notice in the waveform. By looking at the plots of cover audio and stego audio it is evident that both audios are highly similar and do not exhibit any differences of amplitudes with respect to time after embedding secret into it.

Table 2. Measuring average MSE, PSNR, PCC, SCC, PESQ, and STOI for stego in the proposed method

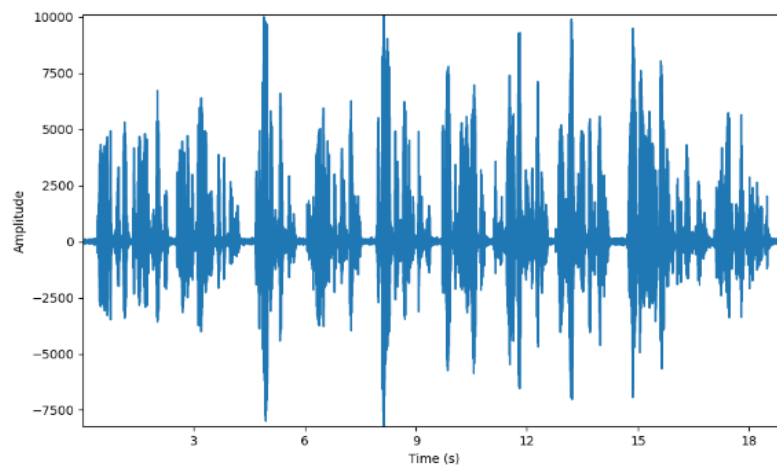
Cover	Secret	MSE	PSNR	PCC	SCC	PESQ	STOI
PinkPanther30.wav	BabyElephantWalk60.wav	2.64E-05	75.43	0.9997	0.9994	4.1175	0.9731
	CantinaBand60.wav	1.60E-05	77.61	0.9998	0.9997	4.2575	0.9700
	Fanfare60.wav	1.33E-05	78.41	0.9998	0.9997	4.2948	0.9824
	ImperialMarch60.wav	5.41E-05	72.31	0.9993	0.9987	4.0961	0.9433
	StarWars60.wav	1.23E-04	68.74	0.9985	0.9973	3.7821	0.9256
gettysburg10.wav	BabyElephantWalk60.wav	7.30E-06	75.67	0.9997	0.9963	3.8621	0.9568
	CantinaBand60.wav	4.59E-06	77.69	0.9998	0.9973	4.0852	0.9566
	Fanfare60.wav	5.92E-06	76.59	0.9998	0.9968	4.2898	0.9878
	ImperialMarch60.wav	2.49E-05	70.35	0.9990	0.9833	4.2756	0.9787
	StarWars60.wav	3.69E-05	68.63	0.9985	0.9778	4.1681	0.9643
preamble19.wav	BabyElephantWalk60.wav	1.45E-05	73.45	0.9995	0.9912	4.2052	0.9675
	CantinaBand60.wav	7.90E-06	76.08	0.9997	0.9944	3.8569	0.9534
	Fanfare60.wav	6.94E-06	76.64	0.9998	0.9971	4.1564	0.9907
	ImperialMarch60.wav	4.30E-05	68.72	0.9985	0.9737	4.0437	0.9367
	StarWars60.wav	6.47E-05	66.95	0.9977	0.9649	4.0916	0.9432
Average		3.00E-05	73.56	0.9992	0.9911	4.1055	0.9620

Table 3. Average PSNR, PESQ, and STOI between stego and cover audio for [24] and proposed method

Metric	Amiri and Naderi [24]	Audio fusion and fission (proposed)
PSNR	52.29	73.56
PESQ	4.063	4.1055
STOI	0.9749	0.9620



(a)



(b)

Figure 3. Signal plot of (a) sample cover audio and (b) sample stego audio

**3.2. Analysis of hiding capacity**

The proposed technique gives a good amount of hiding capacity when compared to the state-of-the-art techniques. In the Table 4 comparison of hiding capacity with other techniques are listed. From the Table 4 its evident that the proposed method is having very good hiding capacity.

Table 4. Hiding capacity comparison with other two techniques

Technique	Hiding capacity (up to)
Shahadi <i>et al.</i> [17]	48%
Ali <i>et al.</i> [25]	100%
Bharti <i>et al.</i> [26]	100%
Chaharlang <i>et al.</i> [27]	1 qubit/sample
Hameed [16]	90%
Audio fusion and fission(proposed)	200%

**3.3. Analysis of robustness**

Banik and Bandyopadhyay [28] developed attack resistant audio steganography method. But the chances of a covert information or channel getting attacked cannot be denied. We performed three attacks to measure the strength of the proposed technique against attacks, namely changing the sampling rate, adding the noise to the stego and compression attack.

**3.3.1. Re-sampling attack**

Re-Sampling attack is a very common attack to hamper the quality of the secret communicated in covert communication. It is very important to measure the effect of this attack on the stego. Authors considered the stego, changed the sampling rate and changed back to original sampling rate and measured all the metrics to assess the impact of re-sampling attack.

**3.3.2. Additive gaussian noise attack**

Additive gaussian noise is one more possible attack to hamper the hidden secret in the stego. It is required to test the capability of the approach in withstanding the same. Authors considered the stego and added noise to it and measured all the metrics to assess the impact of attack.

**3.3.3. Compression attack**

Authors considered the stego and compressed it and again uncompressed and measured all the metrics to assess the impact of attack. The average results of all three attacks are depicted in Table 5. The results show that the proposed technique withstands the attacks and hence proves to be robust.

Table 5. Average MSE, PSNR, PCC, SCC, PESQ, and STOI between secret audio and secret audio received after attack

Attack	MSE	PSNR	PCC	SCC	PESQ	STOI
Re-sampling	0.095	49.96	0.9701	0.9625	2.9038	0.8510
Additive gaussian noise attack	0.099	49.01	0.9566	0.9399	2.5204	0.7644
Compression	0.073	50.47	0.9745	0.9632	2.8164	0.8723

**3.4. Analysis of quality of secret received**

Much of the literature focuses on imperceptibility, hiding capacity and robustness. It is highly required to analyze the quality of secrets received at the sender side to justify the goodness of the technique. Table 6 lists the average of MSE, PSNR, PCC, SCC, PESQ, and STOI between secret sent and secret received.

Figure 4 depicting the signal plots of the secret audio communicated and received. Figure 4(a) plotting audio signal sent as secret, and Figure 4(b) plotting audio signal retrieved as secret. Since we are time stretching the audio received as secret after enhancing it to 100% from 1%, it is amplifying the secret audio signal a bit, However the quality of the secret audio received is enough to full fill the purpose. Since the hiding capacity is double the size of cover, this approach is promising when hiding capacity is crucial requirement.

Table 6. Average MSE, PSNR, PCC, SCC, PESQ, and STOI between secret audio and secret received audio

Technique	MSE	PSNR	PCC	SCC	PESQ	STOI
Audio fusion and fission (proposed)	0.060	59.27	0.9800	0.9702	3.679	0.8981

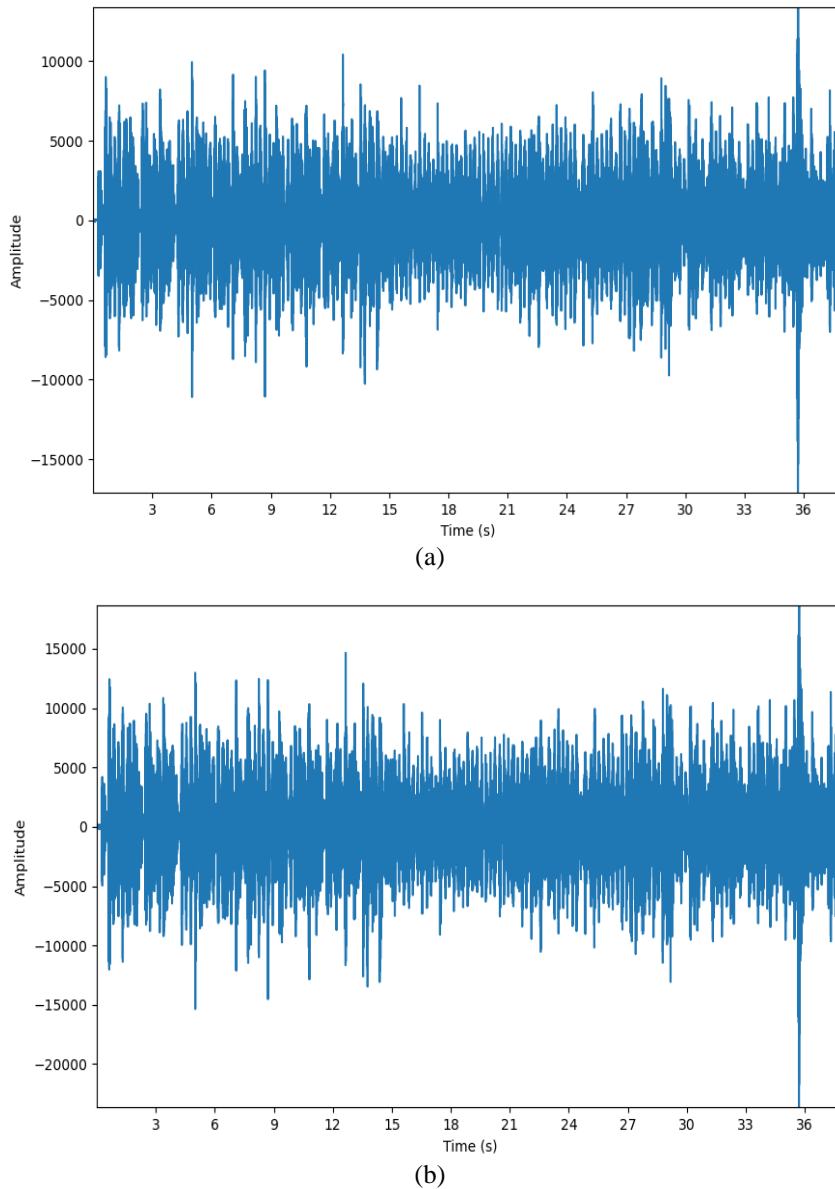


Figure 4. Signal plot of (a) secret audio sent and (b) secret audio received

#### 4. CONCLUSION

This work presents a model for the steganography of audio signal by the audio fusion and fission technique in wavelet domain. Experimental results verify that the embedded audio has good imperceptibility, and the hidden secret message has better robustness against signal processing attacks, such as re-sampling, additive gaussian noise, and compression. Further the same technique can be extended to intelligent fusion where in the imperceptibility increases further.

#### ACKNOWLEDGEMENTS




This project is funded by vision group on science and technology (VGST), Karnataka science and technology promotion society (KSTePS), Government of Karnataka and the authors would like to express gratitude to staff and management of JNN College of Engineering for their support.






## REFERENCES

- [1] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative study of digital audio steganography techniques," *Eurasip Journal on Audio, Speech, and Music Processing*, vol. 2012, no. 1, p. 25, Oct. 2012, doi: 10.1186/1687-4722-2012-25.
- [2] S. Y. Tang, Y. J. Jiang, L. P. Zhang, and Z. B. Zhou, "Audio steganography with AES for real-time covert voice over internet protocol communications," *Science China Information Sciences*, vol. 57, no. 3, pp. 1–14, Feb. 2014, doi: 10.1007/s11432-014-5063-2.
- [3] Z. Wu, R. Li, and C. Li, "Adaptive speech information hiding method based on K-means," *IEEE Access*, vol. 8, pp. 23308–23316, 2020, doi: 10.1109/ACCESS.2020.2970194.
- [4] G. Xin, Y. Liu, T. Yang, and Y. Cao, "An adaptive audio steganography for covert wireless communication," *Security and Communication Networks*, vol. 2018, pp. 1–10, Aug. 2018, doi: 10.1155/2018/7096271.
- [5] Z. Yang, X. Du, Y. Tan, Y. Huang, and Y.-J. Zhang, "AAG-stega: automatic audio generation-based steganography," *arXiv e-prints*, 2018, doi: 10.48550/arXiv.1809.03463.
- [6] D. Ye, S. Jiang, and J. Huang, "Heard more than heard: an audio steganography method based on GAN," *arXiv preprint*, 2019, doi: 10.48550/arXiv.1907.04986.
- [7] D. C. Kar and C. J. Mulkey, "A multi-threshold based audio steganography scheme," *Journal of Information Security and Applications*, vol. 23, pp. 54–67, Aug. 2015, doi: 10.1016/j.jisa.2015.02.001.
- [8] S. Jiang, D. Ye, J. Huang, Y. Shang, and Z. Zheng, "SmartSteganography: light-weight generative audio steganography model for smart embedding application," *Journal of Network and Computer Applications*, vol. 165, p. 102689, Sep. 2020, doi: 10.1016/j.jnca.2020.102689.
- [9] A. A. Alsabhany, A. H. Ali, F. Ridzuan, A. H. Azni, and M. R. Mokhtar, "Digital audio steganography: systematic review, classification, and analysis of the current state of the art," *Computer Science Review*, vol. 38, p. 100316, Nov. 2020, doi: 10.1016/j.cosrev.2020.100316.
- [10] A. A. Alsabhany, F. Ridzuan, and A. H. Azni, "The adaptive multi-level phase coding method in audio steganography," *IEEE Access*, vol. 7, pp. 129291–129306, 2019, doi: 10.1109/ACCESS.2019.2940640.
- [11] A. Danti and G. R. Manjula, "Secured data hiding of invariant sized secret image based on Discrete and Hybrid Wavelet transform," in *2012 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC*, Dec. 2012, pp. 1–6, doi: 10.1109/ICCIC.2012.6510181.
- [12] H. A. Abdulkadhim and J. N. Shehab, "Audio steganography based on least significant bits algorithm with 4D grid multi-wing hyper-chaotic system," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 320–330, Feb. 2022, doi: 10.11591/ijece.v12i1.pp320-330.
- [13] S. T. Abdulrazaq, M. M. Siddeq, and M. A. Rodrigues, "A novel steganography approach for audio files," *SN Computer Science*, vol. 1, no. 2, p. 97, Mar. 2020, doi: 10.1007/s42979-020-0080-2.
- [14] S. E. El-Khamy, N. O. Korany, and M. H. El-Sherif, "A security enhanced robust audio steganography algorithm for image hiding using sample comparison in discrete wavelet transform domain and RSA encryption," *Multimedia Tools and Applications*, vol. 76, no. 22, pp. 24091–24106, Nov. 2017, doi: 10.1007/s11042-016-4113-8.
- [15] S. T. Chen, T. W. Huang, and C. T. Yang, "High-SNR steganography for digital audio signal in the wavelet domain," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 9597–9614, Nov. 2021, doi: 10.1007/s11042-020-09980-6.
- [16] A. S. Hameed, "High capacity audio steganography based on contourlet transform," *Tikrit Journal of Engineering Sciences*, vol. 25, no. 1, pp. 1–7, Feb. 2022, doi: 10.25130/tjes.25.1.01.
- [17] H. I. Shahadi, R. Jidin, and W. H. Way, "A novel and high capacity audio steganography algorithm based on adaptive data embedding positions," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 7, no. 11, pp. 2311–2323, Mar. 2014, doi: 10.19026/rjaset.7.531.
- [18] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2001, vol. 2, pp. 749–752, doi: 10.1109/icassp.2001.941023.
- [19] R. Shanthakumari, E. M. R. Devi, R. Rajadevi, and B. Bharaneeshwar, "Information hiding in audio steganography using LSB matching revisited," *Journal of Physics: Conference Series*, vol. 1911, no. 1, p. 12027, May 2021, doi: 10.1088/1742-6596/1911/1/012027.
- [20] H. Zhang, X. Zhang, and G. Gao, "Training supervised speech separation system to improve STOI and PESQ directly," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Apr. 2018, vol. 2018-April, pp. 5374–5378, doi: 10.1109/ICASSP.2018.8461965.
- [21] D. Ellis, "Sound examples," [www.ee.columbia.edu](http://www.ee.columbia.edu), 2003. <https://www.ee.columbia.edu/~dpwe/sounds/> (accessed Sep. 06, 2023).
- [22] "CS 101 - sample sound files," [www2.cs.uic.edu](http://www2.cs.uic.edu). <https://www2.cs.uic.edu/~i101/SoundFiles/> (accessed Sep. 06, 2023).
- [23] S. W. Fu, T. W. Wang, Y. Tsao, X. Lu, and H. Kawai, "End-to-end waveform utterance enhancement for direct evaluation metrics optimization by fully convolutional neural networks," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 26, no. 9, pp. 1570–1584, Sep. 2018, doi: 10.1109/TASLP.2018.2821903.
- [24] N. Amiri and I. Naderi, "DWT-GBT-SVD-based robust speech steganography," *arXiv preprint*, 2020, doi: 10.48550/arXiv.2004.12569.
- [25] A. H. Ali, L. E. George, A. A. Zaidan, and M. R. Mokhtar, "High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain," *Multimedia Tools and Applications*, vol. 77, no. 23, pp. 31487–31516, Jun. 2018, doi: 10.1007/s11042-018-6213-0.
- [26] S. S. Bharti, M. Gupta, and S. Agarwal, "A novel approach for audio steganography by processing of amplitudes and signs of secret audio separately," *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 23179–23201, Apr. 2019, doi: 10.1007/s11042-019-7630-4.
- [27] J. Chaharlang, M. Mosleh, and S. Rasouli-Heikalabad, "A novel quantum steganography-steganalysis system for audio signals," *Multimedia Tools and Applications*, vol. 79, no. 25–26, pp. 17551–17577, Feb. 2020, doi: 10.1007/s11042-020-08694-z.
- [28] B. G. Banik and S. K. Bandyopadhyay, "Blind key based attack resistant audio steganography using cocktail party effect," *Security and Communication Networks*, vol. 2018, pp. 1–21, 2018, doi: 10.1155/2018/1781384.

**BIOGRAPHIES OF AUTHORS**

**Namitha Mangikuppe Venkateshaiah**    is Assistant Professor at the Department of Computer Science and Engineering, JNN College of Engineering, Shimoga, Karnataka, India, and a research scholar at Visvesvaraya Technological University (VTU). She has 12 years of teaching experience and 1 year of industry experience. She has published more than 8 publications, in both international conferences and journals. Her research areas are steganography, information security and machine learning. She can be contacted at email: namithamv@jnnce.ac.in.



**Manjula Govinakovi Rudrappa**    is Professor at the Department of Computer Science and Engineering, JNN College of Engineering, Shimoga, Karnataka, India. She has received a Ph.D. degree in Computer Science from the Kuvempu University, Shankaraghatta, Karnataka, India. She has supervised and co-supervised more than 25 masters and she is guiding 5 Ph.D. students. She has authored or co-authored more than 28 publications. She is also a permanent member of LMISTE and MIETE. Her research interests include steganography, soft computing, machine learning, and intelligent systems. She can be contacted at email: grmanjula@jnnce.ac.in.