International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

A Robust Digital Speech Watermarking Based on Least Significant Bit

Mritunjay Kumar, Rajeev Kumar, Jainath Yadav

Abstract*: Watermarking is a technique to ensure the original information and to validate the digital content. Watermarking is required because of the rise in the utilization of the internet in one's everyday life. As the usage of digital content is developing quickly, there are numerous occurrences where information is uncertain. Watermarking is a procedure to conceal information for authorization reasons. Watermarking is the ideal approach to make sure about the digital content. Watermarking should be possible through different strategies. Least Significant Bit Watermarking (LSBW) strategy is one of them. Right now, pixel estimations of the image are changed over into binary. The data is covered in the bits of the pixel esteems. Watermarking consistently infers embedding of furtive signal that should be robust and imperceptible within the host data. In order to cater its robustness and imperceptibility, the energy of the watermark is accommodated to the energy of speech components.

Keywords*: Speech watermarking, Least Significant Bit (LSB), Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Minimum Value.

I. INTRODUCTION

In recent decades, there has been a sensational increment of unapproved generation, manipulation, and dissemination of digital multimedia information. It has created a primary issue, i.e., copyright encroachment, which should be handled around the world. While copyright protection is required for different multimedia information, including audio, image, and video, this paper confines its regard, especially for audio and speech information. Digital watermarking is obligating mechanization to deal with the concern of audio information infringement. The ever-developing length of digital evidence is transmitted over the internet. It is like never before essential for productive and reasonable information concealing strategies to be structured to secure protected innovation rights. Digital watermarking systems have verifiably been utilized to guarantee security regarding ownership stability and sealing for a wide assortment of information designs. It incorporates image, sound, video, characteristic language transforming software, social databases.

Revised Manuscript Received on April 30, 2020.
* Correspondence Author
Mritunjay kumar*, Department of Computer Science, Central University Of South Bihar, Gaya, India. Email: mritunjay.19february@gmail.com
Rajeev Kumar, Department of Computer Science, Central University Of South Bihar, Gaya, India. Email: rajeevkr@cusb.ac.in
Dr. Jainath Yadav, Department of Computer Science, Central University Of South Bihar, Gaya, India. Email: jainath@cub.ac.in

This paper centers around speech watermarking. Typically, digital speech watermarking is an innovation of embedding appropriate information inside a host speech signal. The noncognitive nature of the host speech signal ought not to be diminished extensively by the embedding. For various objectives, speech watermarking can be extended into two groups: fragile audio watermarking and robust audio watermarking. Robust audio watermarking is utilized to secure responsibility for digital audio. In differentiate, the motivation behind fragile audio watermarking is digital audio verification, which is to guarantee the rectitude of the digital audio. In the previous decade, scientists have put forth the incredible attempts in creating robust audio watermarking algorithms, which were actualized in either the time domain or transform domain, for example, the DFT, DCT, DWT, cepstrum, SVD [2], and so forth.

Digital watermarking is a procedure of embedding watermarks in unique media, for example, video, sound, and image with the presence that their essence can't be seen. As indicated by the International Federation of the Phonographic Industry (IFPI), digital audio watermarking must have a couple of fundamental attributes. The first is imperceptibility. The watermark ought not to be discernible what estimated by objective and subjective strategies. Moreover, its SNR must be more than 20 dB. Another fundamental essence is the payload, i.e., the number of bits conveyed by an audio watermarked signal ought to be at any rate 20 bps. The third fundamental essence is robustness, i.e., the watermark must be extricated under various signal handling attacks except if the nature of the watermarked signal has been momentously diminished. Another quality of watermark is fragile which is something contrary to robustness, and it is utilized for confirmation intent. Different attributes incorporate preservation, which is interpreted as a watermark location by an approved individual. These attributes are transparency, intricacy, inevitability, and blindness. Transparency implies no distinguishable antiquities or loss of value. Iintricacy commits time for embedding and extraction, particularly progressively framework. Inevitability is the likelihood for the propagation of the unique signal. Blindness is the capacity to separate the watermark without the unique signal.

From early days of image steganography, LSB substitution has been an exceptionally helpful idea where LSB plane of the image is supplanted by the pseudorandom arrangement of classified information. Anyway, it is seen that there happens some auxiliary asymmetry because of the LSB substitution, and these are misused to intensify distinctive steganalytic attacks.

However, the LSB coordinating, which is a analogue of LSB substitution,

doesn't just overwrite the LSB plane. It may be chosen pixel esteems that are expanded or diminished dependent on a pseudorandom key if its LSB does not coordinate the classified message bit to be embedded. By embracing this way, the LSB coordinating strategy dispenses with the issue of fundamental asymmetry, and henceforth it stays imperceptible against the aforementioned attacks.

To ensure this digital audio watermarking is utilized. This procedure protects from awful users and gives copyright, robustness, security to the digital substance. Audio watermarking is the approach of embedding and extraction methods. In the embedding process, the substance like audio, video or picture is embedded into a unique document that is to be made secured. The extraction strategy permits us to remove the substance, however the document is by and large despite everything secured. There are about certain properties that fulfill the requirement for viable watermarking applications. These are robustness, constant bit-rate, fragile, semi-fragile, inaudible, security, and verifiability as discussed below.

Robustness: It is the ability to manage the copyright data of digital works, the embedded watermark can decline to acknowledge the necessary altering process, handling the image and lossy compression. Additionally, after attacks the watermark can't be harmed, and it can be yet recognized to offer confirmation. For instance, editing, noise, cropping, compression, A/D-D/A transformations, geometrical or non-geometrical attacks, and so on.

Constant bit-rate: The measure of watermark data might be safely embedded inside the host signal per unit space or time.

Fragile: Fragile watermarking is utilized for chiefly rectitude assurance which is sensitive to the progressions of the signal. We can decide altered information as per the condition of fragile watermarking.

Semi fragile: It is capable of overseeing changes made to the watermarked image, for example, expansion of lossy compression.

Inaudible: The digital watermark is embedded into audiho information as it ought not to be discernible to the human ear.

Security: A framework is accepted to be ensured if the saltine can't remove the watermark applied without having the information on the embedded algorithm, finder, and structure of the watermark. Just the authorized user can access it.

Verifiability: It tends to be utilized to check the item is secured, for example, copyright-secured and recognize the credibility and control of unlawful replication.

For information security, the watermarking systems are acquainted with giving security of data. As of late, the watermarking systems have been acquainted with center around pictures, and video cuts, yet sound watermarking is progressively entangled that video and image watermarking. Here are two critical reasons so as audio watermarking has gotten arduous. First, to start with, the Human Auditory System (HAS) has bigger affectability than the Human Visual System (HVS) since human ear is fit for distinguishing the amplitude and frequency transformation of the signal. Second, the span and size of the audio signal are exceptionally shorter than video clasps and image records, and this data diminishes the audio signal quality. Least Significant Bit embedding is a straightforward methodology of watermarking. It embeds the information into the spread message with the goal that it can't be identified by visual eyes. This technique works by supplanting bits with the secret message. It is conceivable by changing a few bits with a secret message. It embeds information into the image on any piece plane. It diminishes the varieties in hues that embedding makes. For instance, embedding into the main piece plane transform the incentive by 1. Similarly, for the second piece plane, it changes the incentive by 2. This procedure is followed for all the bits.

II. RELATED WORK

Speech watermarking strategy is one of the utilizations of the biometric watermarking system. Here, the speech signal is considered either as a spreading mechanism or enigma watermark data. Different existing procedures are identified with the proposed method, and they are outlined in this section.

El-Gazar et al. [16] recommended an SVD and DES based speech watermarking method. Right now, the DES algorithm was enforced to the watermark image to produce an encoded watermark image. At that point, this encoded watermark image was embedded into the singular value of speech signal to obtain the watermarked speech signal.

Revathi et al. [2] introduced a DWT based speech watermarking procedure. In this method, they considered the cover of data and watermark speech signals. They were likewise allowing individual recognition techniques dependent on watermarked speech signals utilizing the clustering algorithm.

Nematollahi et al. [4] endorsed a speech watermarking method utilizing the quantization of the LP-standard. Here, every speech signal is separated into two vectors dependent on the odd and even file esteems. The QIM is utilized to embed watermark data into the proportion of LP-standard between these two lists. Decisively, the Lagrange advancement system is utilized to decrease embedding mutilation in the altered signal.

Lu et al. [6] consolidated watermarking with the CELP (Code Excited Linear Prediction) speech coding process for validation of compacted speech by CELP-composed coders. This verification plot is pertinent just to constrict speech. The work displayed here is the extended work, and it targets giving validation of speech signal that is powerful against amplitude scaling. The fragility of the watermark can be constrained by determining the detection edge as per the normal SNR.

Faundez-Zanuy et al. [8][9][10] effectively consolidated watermarking based speech validation into telephonic

chronicle utilized in standardized savings observing and security upgraded speaker confirmation or speaker recognition applications. Another lucrative application and the unambiguous pattern is adopted in air traffic regulation,

where the quality of VHF (very high frequency) radio channel is completely investigated when structuring the verification algorithms.

Saraswathi [7] changed the Mel-frequency cepstrum (MFCC) in speech portions having flat force for speech authentication. However, the adjusted MFCCs should be sent to the acceptor, which makes this plan a non-blind one. Non-blind watermark location has restricted application in speech validation since we have to accumulate and disseminate side data. A unique model in the watermark signal can likewise be planned and used to distinguish the sort of the attacks or in part recoup the speech content.

Yong et al. proposed the SS-based audio watermarking approach which has much predominant embedding capacity and satisfactory imperceptibility and robustness. The high embedding capacity is practiced through a lot of systems. Embedding various watermark bits in a single audio fragment, packing host signal mediation on watermark extraction, and adaptively accommodating PN arrangement sufficiency in watermark embedding is dependent on the value of audio portions. The view of SS-based audio watermarking contains a watermark bit that is inserted into a host audio segment by utilizing an extensive sequence. At the decoder, the inserted watermarks are extricated by relating the watermarked signal with the spreading sequence. The watermark extraction system utilized the result to have signal interference. Huge host signal impedance could fundamentally debase the precision of watermark extraction, and accordingly, decrease the power of the audio watermarking technique.

Li and Yu proposed phase modulation as the other watermarking method moves the period of speech and marmalade the power range with no changes. As the first and watermarked signals have a similar power range, the signal isn't distorted. Phase modification and phase coding are two renowned strategies for phase balance. Phase modification uses various groups for watermarking, while phase coding utilizes one edge for the entire watermark information. Embedding watermark data in the cepstrum coefficients of a log spectral domain is a robust and imperceptible strategy for watermarking into the speech signal.

Thanki et al. [5] introduced a curvelet change and compressive sensing (CS) encryption-based audio watermarking strategy. Right now, the scrambled speech signal is utilized for proprietorship distinguishing proof of sound signals.

Nematollahi et al. [3] recommended a speech watermarking strategy utilizing the hybridization of DWT + SVD. This current strategy was the use of the Bhat method (Bhat et al. 2010) for the assurance of speech signal. Inamdar and Rege (2014) asserted different watermarks based strategy for insurance of facial appearance and speech signal.

The fragile watermarking approach recognizes altering and find adaptation in the signals. A characteristic that wants notwithstanding distinguish altered locales is to revitalize them, and with this thought, the self-recovery pattern was introduced. The plan proposed by Fridrich and Goljan first presented embedding an image into itself to restore the altered regions. Self-recovery schemes ascertain a compressed rendition of its signal or significant attributes and embed this compressed portrayal into the signal itself, this compressed data close by a fragile watermark permit the detection of altered regions and the reclamation of modifications to the signal.

III. CONTRIBUTIONS OF THE PROPOSED WORK AND MATHEMATICAL PRELIMINARIES

A. LSB

The LSB alteration is one of the least sophisticated audio steganography methods providing high capacity. The technique is being covered up in the least significant bit(s) of audio examples. The weightage of LSBs in examination with the joined weightage of the entire case is very small. However, changing the LSBs will initiate some noise, but as long as the commotion instigated is below the recognizable threshold, sound steganography is conceivable. Expanding the number of changed LSBs will launch more noise. If noise increments over the threshold and it gets perceivable through any of the steganalysis strategies, then audio steganography procedure comes up short. Utilizing more LSBs per test builds the limit and diminishes the transparency. Then again, utilizing fewer LSBs per sample will reduce capacity and increase transparency. Along these lines, there is continuously a tradeoff between both these parameters.

B. Watermarking procedure

Digital watermarking is a procedure of embedding supplementary furtive data into digital signals, for example, audio, picture or video signals. If the watermarked signals are replicated, at that point, the implanted data is likewise conveyed in duplicate. The inserted data or watermark is generally in double configuration, and the nature of unique signals ought not to be influenced by the embedding watermark. The watermark bits are not clear to any unauthorized user, and they can be utilized to demonstrate the ownership responsibility for watermarked signals. The watermark ought to be robust and handily extricated from watermarked signals much after some incidental and purposeful attacks, for example, expansion of noise, re-sampling, and MP3 compression.

C. Embedding Process

In the embedding system of LSB technique, the edge parameters ought to be converted into multi-ary numbers first separately. The interpretations are autonomous from every parameter. Subsequently, the secret message is converted into various numerations correspondingly. In Fig. 1, binary bits are firstly embedded by LSB substitution strategy. At that point, various assigned bits are taken out for ternary embedding after a binary to-ternary interpretation. And embedding activity is done similarly, from that point forward, all stego parameters are interpreted binary numeration for speech transmission. More subtleties of embedding procedure are described as in the accompanying steps:

1) Read the original speech signal.
2) Segment the speech signal with lengths of 20–30 ms into frames.
3) For each frame, select the minimum sample value.
4) Read the watermark image.

5) Find minimum value from each row (watermark signal must be less than or equal to frame size).
6) The minimum value of each frame is divided by the mean of mean to normalize the data.
7) Apply the LSB method for both signals.
8) Watermark is embedded on the first-bit plane.

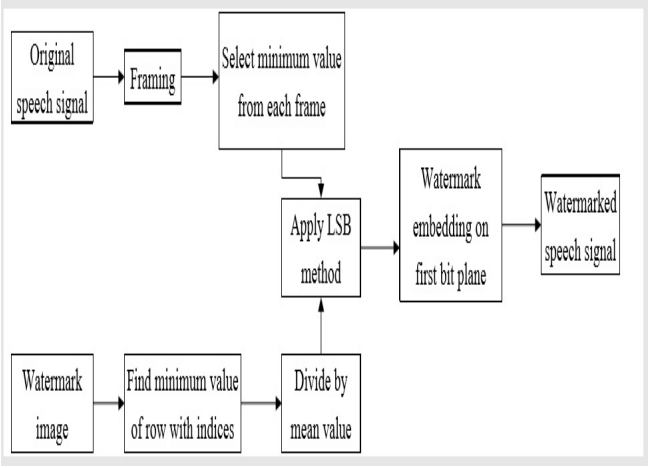

Fig.1: The block diagram of the proposed mechanism of watermark embedding

D. Extraction Process

The extracting technique is generally straightforward; we just concentrate the LSBs of the stego outline parameters as indicated by the numeration received in the embedding algorithm. And afterward, we transform it into binary bits. Fig. 2 shows the flow diagram for extracting the watermark. The process to extract watermark is written as follow:

1) Read a watermarked speech signal.
2) The watermarked speech signal is segmented into frames with the same length as during embedding.
3) Find watermarked location value.
4) Multiply the signal by the mean value.
5) Replace the bit with the original watermark bit by bit.
6) Extract the binary watermarked image.

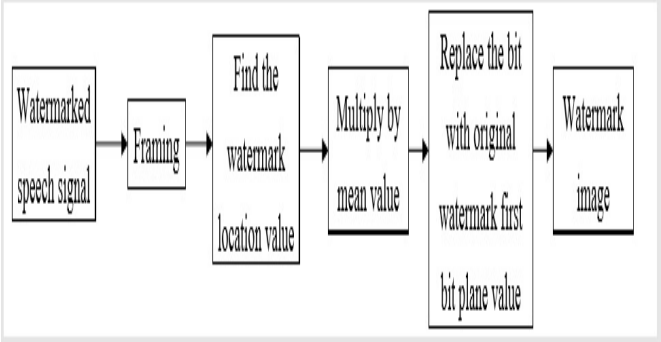

Fig. 2: The block diagram of the proposed watermark extraction process

E. Quality evaluation parameters

The accomplishment of the introduced model dependent on the LSB watermark technique is explored for their imperceptibility and robustness across the various assortment of attacks. There are two quality evaluation methods for the watermarked signal, subjective and objective methods. One of the target precedent to assess the detectable quality is PSNR, which essentially demonstrates the connection between the host signal and watermarked speech signal. Another fundamental specification for any watermarking framework is the robustness. The robustness can be assessed by ascertaining the normalized correlation coefficient (NCC). The following equations compute PSNR and MSE:

$$PSNR = 20\ log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \qquad (1)$$

$$MSE = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[X(i,j) - w(i,j)\right]^2 \qquad (2)$$

Where m*n indicates the size of the host signal, X(i,j) is cover speech signal, and w(i,j) is the watermarked speech signal.

IV. SIMULATION RESULT

In the inclusion of imperceptibility, it is also recommended to conspire accomplishes higher robustness. In simulations, imperceptibility is determined to utilize the normalized correlation (NC) value between the watermark and extracted watermark speech signal. Correspondingly, for testing robustness, the watermarked speech signal was presented to the various condition of attacks in the interim of transmission. These attacks are speech preparing attacks, for example, noise expansion attacks, de-noising attacks, compression attacks, and geometrical attacks such as rotation, scaling and translation, etc. Three distinct noises with various densities were supplemented to the watermarked speech signal in our analysis. Table 1 comprises quality parameters like PSNR, MSE, NC, and its simulated result. The proposed speech watermarking technique is evaluated on the basis of six different speech signal of length 4secs. Each frame is calculated using the duration of 25ms. In this simulation, we found 171 frames. Fig. 3 shows the original speech signal in the form of the number of frames and amplitude. Fig. 5 represents the watermark image of Babbon.

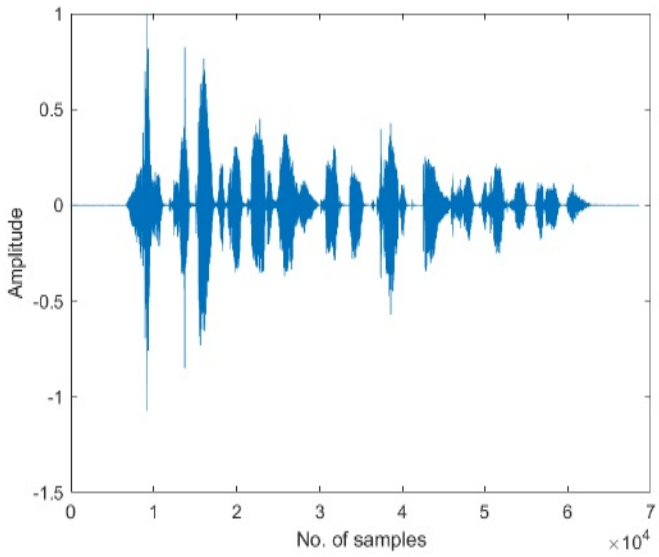

Fig. 3. Original speech signal

Collection of frames of length 400 samples of each frame are depicted in Fig. 4. Using the proposed approach, the watermark is embedded into the selected minimum value of speech signal, and finally get the watermarked speech signal, which is demonstrated in Fig. 6. After the extraction process of the watermarked speech signal, we obtain the watermark image, as shown in Fig. 7.

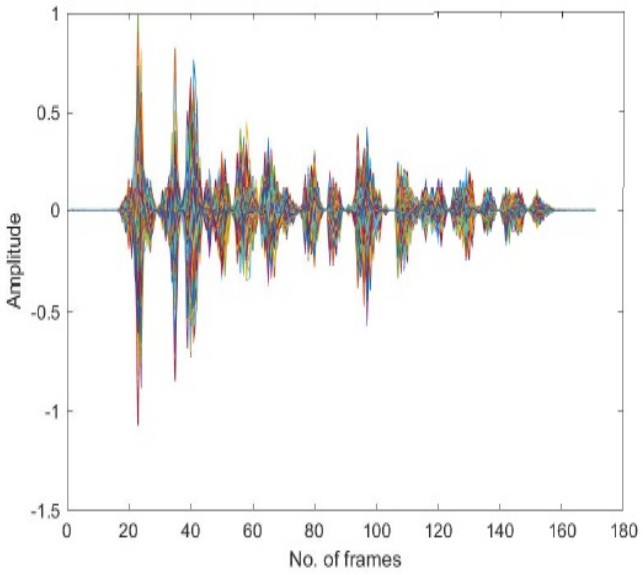

Fig. 4. Total number of frames of normalised data

The total number of frames of normalized data of speech signal as depicted in Fig. 4 that comprises 171 numbers of frame in which each frame contains 400 samples. We have taken various speech signal data like emotional speech signal, anger speech signal, natural speech signal and so on. The extensive outcome of numerous sample of speech signal based on certain parameters such as PSNR, MSE and NC are corroborated in table 1. The simulation outcome provides an almost similar speech signal after the extraction of watermark image that shows in NC values.

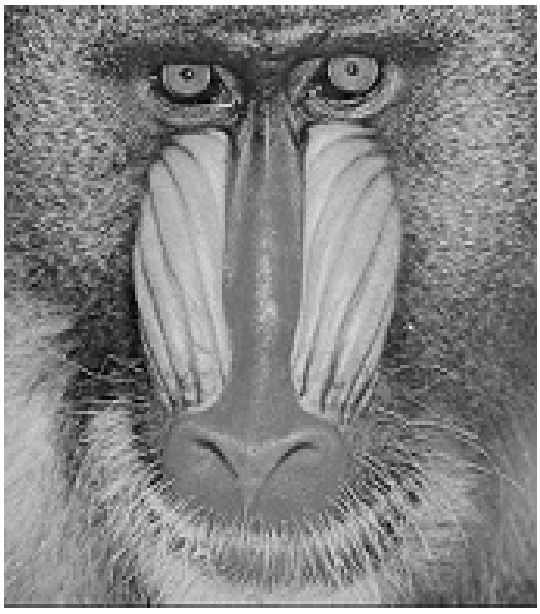

Fig. 5 Watermark image

TABLE I
PSNR, MSE AND NC VALUES AFTER EXTRACTION OF WATERMARK

Speech signal	*PSNR (dB)*	*MSE*	*NC*
Emotional.wav	63.523	0.0289	0.995
Natural.wav	65.667	0.0176	1.000
Anger.wav	55.641	0.2234	0.989
Story.wav	64.043	0.0256	0.997
Happy.wav	66.445	0.0147	0.998
Sad.wav	63.050	0.0322	0.999

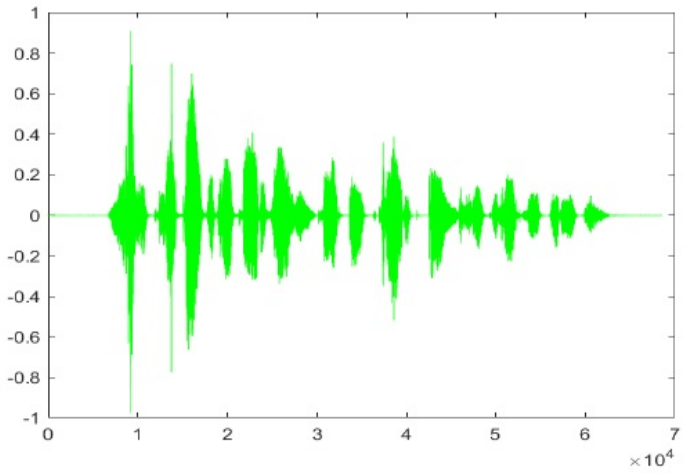

Fig. 6. Watermarked speech signal

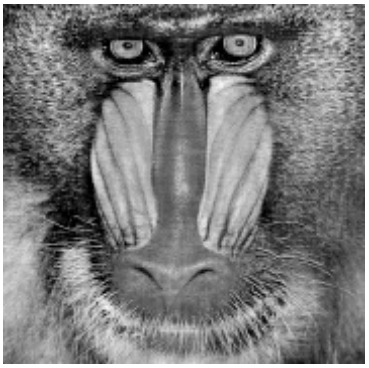

Fig. 7 Extracted watermark image

V. CONCLUSION

In this paper, we introduced the LSB based watermarking procedure for validation. We have built up a robust non-blind watermarking strategy dependent on minimum value LSB procedures against a few attacks. The prospective method is tested against different attacks such as geometric, purposeful, inadvertent, filtering, pivot, editing, scaling, compression, gamma revision, alongside various noise irritation viz. Gaussian noise at various density levels. Besides, the test consequence checks that the proposed method provides high imperceptibility and robustness. Likewise, the recommended scheme additionally has the unrivaled exhibition of robustness regarding NC over the existing plan alongside higher defiance for brute force attack.

Due to having superior imperceptibility, security, and robustness, the proposed plan might be a convenience for speech transformation and watermarking applications.

REFERENCES

1. Almeida LB(1994) The fractional Fourier transform and time-frequency representations.IEEE Trans Signal Process 42(11):3084–3091.
2. Revathi A, Sasikaladevi N, Jeyalakshmi C (2018) Digital speech watermarking to enhance the security using speech as a biometric for person authentication. Int J Speech Technol 21(4):1021–1031.
3. Nematollahi M, Al-Haddad S, Zarafshan F (2015) Blind digital speech watermarking based on eigen-value quantization in DWT. J King Saud Univ Comput Inf Sci 27(1):58–67.
4. Nematollahi MA, Vorakulpipat C, Gamboa Rosales H (2017b) Optimization of a blind speech watermarking technique against amplitude scaling. Secure Commun Netw 2017:1–13.
5. Thanki R, Borisagar K (2017) Watermarking scheme with CS encryption for security and piracy of digital audio signals. Int J Inf Syst Model Des (IJISMD) 8(4):38–60.
6. Lu, Z. M., Yan, B. and Sun, S. H.: Watermarking Combined with CELP Speech Coding for Authentication. IEICE Transactions on Information and systems. E88D No.2. (2005) 330-334.
7. Saraswathi S (2010) Speech authentication based on audio watermarking. Int J Inf Technol 16(1):34–43.
8. Faundez-Zanuy M (2010) Digital watermarking: new speech and image applications. In: SolCasals J, Zaiats V (eds) Advances in nonlinear speech processing. Lecture notes in computer science, vol 5933. Springer, Berlin, pp 84–89.
9. Faundez-ZanuyM, Hagmüller M, Kubin G (2006) Speaker verification security improvement by means of speech watermarking. Speech Commun 48(12):1608–1619.
10. Faundez-Zanuy M, Lucena-Molina JJ, Hagmüller M (2010) Speech watermarking: an approach for the forensic analysis of digital telephonic recordings. J Forensic Sci 55:1080–1087.
11. J. Fridrich, M. Goljan, Protection of digital images using self-embedding, Proceedings of the Symposium on Content Security and Data Hiding in Digital Media, Newark, NJ, USA, 1999.
12. Campisi P, Kundur D, Neri A (2004) Robust digital watermarking in the Ridglet domain. IEEE Signal Process Lett 11(10):826–30.
13. Akhaee, M. A., Khademi Kalantari, N., and Marvasti, F. (2010). Robust audio and speech watermarking using Gaussian and Laplacian modeling. Signal Processing, 90(8), 2487–2497.
14. Blamey, P., Dowell, R., Clark, G. M., and Seligman, P. (1987). Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant. The Journal of the Acoustical Society of America, 82, 38.
15. Chen, O.-C., and Liu, C.-H. (2007). Content-dependent watermarking scheme in compressed speech with identifying manner and location of attacks. IEEE Transactions on Audio, Speech, and Language Processing, 15(5), 1605–1616.
16. El-Gazar S, Abbas AM, El-Dolil S, El-Dokany IM, Dessouky MI, El-Rabaie ESM, El-Samie FEA (2018) Efficient SVD speech watermarking with encrypted images. Int J Speech Technol 21(4):953–965.
17. Hofbauer K, Kubin G, Kleijn WB (2009) Speech watermarking for analog flat-fading bandpass channels. Audio Speech Lang Process IEEE Trans 17(8):1624–1637.
18. Dhar PK, Shimamura T (2013) A DWT–DCT–based audio watermarking method using singular value decomposition and quantization. J Signal Process 17(3):69–79.

AUTHORS PROFILE

Mritunjay Kumar is pursuing M.Tech Computer science from central university of south Bihar. He has completed B.Tech from Rajasthan Technical University. His area of research is signal processing and speech watermarking.

Rajeev Kumar is pursuing Ph.D from central university of south Bihar. He has completed B.Tech from GITA, Bhubaneswar. He has completed M.Tech degree from Central University of Punjab, Bathinda. His area of research is signal processing, image compression, video, image, and speech watermarking. He has published several papers in the reputed conferences and journals.

Dr. Jainath Yadav is a Assistant Professor in the department of Computer Science, Central University of South Bihar. He has completed M. Tech and PhD from IIT Kharagpur. He has contributed several research papers in the reputed journals like IEEE Transactions on Audio Speech and Language Processing, IEEE Signal Processing Letters, Speech Communication, etc. He has published and presented several papers in the reputed international conferences.