I. J. Computer Network and Information Security, **2025, 1, 57-67**
Published Online on February 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2025.01.05

Vowel based Speech Watermarking Techniques using FFT and Min Algorithm

Rajeev Kumar
Faculty of Information Technology, Gopal Narayan Singh University, Jamuhar, Bihar, India
E-mail: rajeev6136@gmail.com
ORCID iD: https://orcid.org/0000-0001-6036-3581

Kshitiz Singh
Department of Computer science, Central University of South Bihar, Gaya, Bihar, India
E-mail: kshitiz@cusb.ac.in
ORCID iD: https://orcid.org/0000-0003-4332-7604

Jainath Yadav*
Department of Computer science, Central University of South Bihar, Gaya, Bihar, India
E-mail: jainath@cusb.ac.in
ORCID iD: https://orcid.org/0000-0002-1117-4175
*Corresponding Author

Ajay Kumar
Department of Computer science and informatics, Central University of Himachal Pradesh, HP, India
E-mail: ajaykr.bhu@hpcu.ac.in
ORCID iD: https://orcid.org/0000-0002-3991-9757

Indranath Chatterjee
Department of Computer engineering, Tongmyong University, Busan, South Korea
E-mail: indranth@tu.ac.kr
ORCID iD: https://orcid.org/0000-0001-9242-8888

Received: 29 April 2023; Revised: 19 October 2023; Accepted: 28 December 2023; Published: 08 February 2025

Abstract: The critical challenge with the continuously increasing number of Internet users is copying and duplication, which has caused content integrity and protection. To manage and secure the signals from unauthorized consumers of digital content, we require certain procedures. Digital watermarking scheme on vowel-based approach can address these problems. Thus, we can provide a robust and secure method that solves the issues of copyright, illegal intentional or unintentional modification. In this paper, we have proposed vowel-based speech watermarking techniques using the FFT method with the help of the Min algorithm. We observe that the proposed FFT-based watermarking scheme provides better results in comparison to the existing methods.

Index Terms: Arnold Transform, Vowel, FFT, Min Algorithm, PSNR, SSIM.

1. Introduction

As we know that in recent decades the content sharing/communicating through the Internet has been increased rapidly. In this case, our contents must need to be secure from the outside world. Digital speech watermarking is one of the most popular schemes to protect our contents from the real world. The main role of the digital watermarking scheme is to embed the information optimally in the original content [1, 2].

The digital speech watermarking scheme is partitioned into two parts: the first one is watermark embedding and the second part is the extracting process. The initial step should prevent our materials from unauthorized duplication while preserving the quality of the original signal. The second process is used to recover our watermark data from watermarked signals and later, it is used for evidence purposes. For these, watermarking methods should meet basic

digital requirements [3-7].

The speech watermarking techniques require the algorithm for embedding and extricating the watermark, sampling rate/frequency, selection of watermark location, and parameters for finding the imperceptibility and robustness. In general, speech signals are used for visible [8] and dual [9] scheme instead of invisible [10] watermarking. The working methods decide how to embed, where we can insert, how much we can embed, and in which form, i.e., time or frequency or in either domain or encrypt the watermark before insertion.

The watermarking extraction process is determined based on three different algorithms, i.e., blind, semi-blind, and non- blind methods. These algorithms navigate us in which way we can extract our watermark. In general scenario, common applications have been used for watermarking techniques that incorporate copy control, fingerprinting, broadcast monitoring, speech authentication, copyright protection [11-13].

In this paper, our first contribution is to encrypt the watermark data using Arnold transform method, which provides both facilities encryption and decryption. That's why Arnold transform provide us more secure watermark data. Second, originally our signals are in time domain so we have to transform into the frequency domain using FFT method. The benefit of this is to increase the security and reduce the time complexity. The third main advantage is the Min algorithm, which has been developed for embedding and extracting the watermark.

The contents of the paper are organizing as follows: The first section discusses introduction about speech watermarking. Section 2 reviews existing methods related to speech and audio watermarking schemes. Section 3 presents the details of the proposed FFT method. In section 4, quality and robustness parameters have been discussed. Section 5 discusses simulation results on both methods and comparison with existing methods. Lastly, we have presented the conclusion in section 6.

2. Review Work Related to Speech Watermarking

In the literature, different working models are used for the watermarking process. Vowel onset and offset events based speech watermarking for copyright protection and authentication has been proposed by Kumar and Jainath, [14]. Wu et al., [15] have described the spectrum distribution method for audio signals. They worked on synchronization mechanism for solving the large signals converted into small sampling of the signals. Digital audio watermarking technique was based on voiced and unvoiced frame using time domain transformation methods. The high energy of the audio signal (voiced and unvoiced) has been selected for embedding the watermark by Kanhe and Gnanasekaran, [16]. The multi-level and multiple watermark images were used for audio watermarking by Singla et al., [17]. They distribute the watermark in the complete audio signals.

Liu et al., [18] has described a strategy for dealing with various threats issues based on time and frequency modification approaches and an upgraded patchwork algorithm. BER, ODG, and SNR metrics were used to assess the existing method. Also, they carried out a number of attacks, which improved copyright protection and forensics tracking. Charfeddine [19] has proposed a neural network and spectral based architecture for audio signals. They used neural network for watermark insertion and extraction. Additionally, they worked on speech emotion and speakers using LPA method.

Alshathri et al., [20] have described watermarking system on health care based on medical Internet of things. This system provides multilevel secure system during communication. They achieved high security using scrambled the medical image. They are mainly focused on fusion of two watermark images. The fusion based scheme improves the digital requirements. The median filtering algorithm has been used to improve the result of the watermarking scheme proposed by Boateng et al., [21]. Two different methods were implemented to improve the reliability and robustness. For the reliability, they used DCT and DWT methods. For robustness, they used to optimize the result using optimization method. PSO is suitable to find the scaling factors proposed by Run et al., [22]. The watermark image is twice inserted into the host signal's DWT-DST (discrete sine transform) domain in the proposed approach for audio signals. The outcomes of the simulation demonstrate that the suggested watermarking system retains excellent quality and robustness against different attacks. Sometimes, this method is not appropriate for watermark extraction without knowing the keys generated during the insertion process.

We studied the several existing papers on speech and audio signals and analysed different evaluation parameters. These existing methods are useful to finding the gaps for novel concept.

3. Proposed Method for Speech Watermarking

The proposed method explored the novel concept for speech watermarking techniques. The vowel based speech watermarking methods have been presented in the frequency domain with the help of the Min algorithm and Arnold transform. The vowel region based FFT method in the frequency domain has been discussed in subsection 3.1.

Arnold transform is used to encrypt the watermark image, and it offers a different stage of encryption. Henceforth, the encrypted watermark data is selected for embedding purposes. For the watermark embedding process, the Min algorithm is useful for finding the minimum values from each frame and embedding the minimum value of the compressed form of watermark data. In this way, watermark data is inserted throughout the speech signals in a random manner. Due to this reason, it is difficult to detect watermark information from the watermarked speech signal and it

also provides higher security. Fig. 1 shows the procedure for extraction of vowel regions from speech utterances. Fig. 1(a) represents the manual places of the vowel regions in the given speech utterance. Fig. 1(b) represents the detected positions of vowel starting transition points in the illustrated signal. Fig. 1(c) represents the detected the positions of vowel end transition points in the illustrated signal. Finally, in fig. 1(d), vowel regions are extracted from speech signals [14]. A vowel region contains high energy values, and it is useful to embed the watermark based on the FFT and min algorithms.

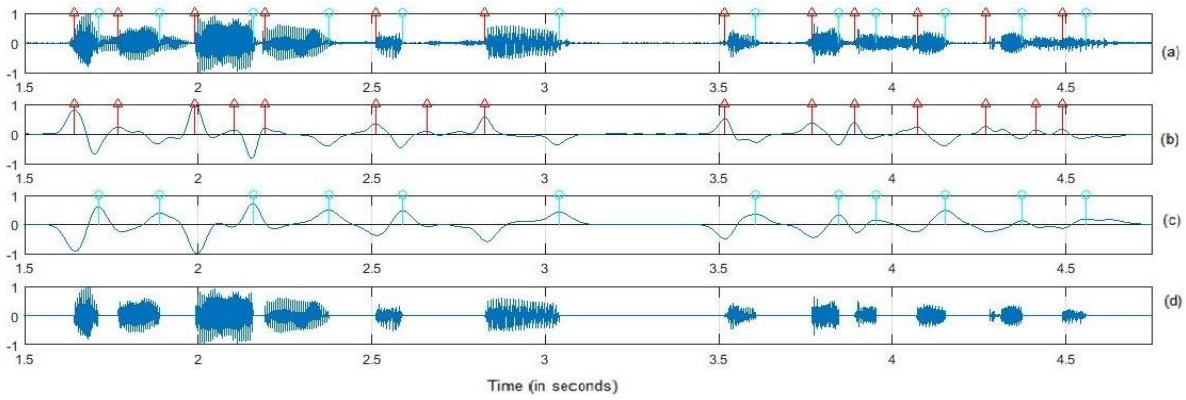

Fig.1. The detection of Vowel regions from speech utterance:(a) manual positions of the vowel starting and ending transition points, (b) and (c) show the detected positions of vowel starting and ending points in the vowel demonstrated signal, respectively, and (d) Extracted vowel regions

3.1. Watermarking Scheme using the FFT Method

The DFT method is useful to transform the finite length signal form spatial to frequency domain using Fourier matrix (FM). This method is periodic and symmetric in nature, so that the FM and its transpose are equal to each other. DFT converts real valued into complex samples of same length. The N- point 1-D DFT and IDFT equations can be written using Eqs. (1) and (2), respectively.

$$Y(k) = \sum_{n=0}^{N-1} y(n) e^{\frac{-2\pi kn}{N}} \quad (1)$$

$$y(n) = \frac{1}{N} \sum_{n=0}^{N-1} Y(k) e^{\frac{2\pi kn}{N}} \quad (2)$$

Where N is the length of the input or N-point DFT, y(n) is the input sequence in discrete time, Y(k) is the output frequency samples, and $0 \leq k, n \leq N-1$.

$W_N = e^{\frac{-j2\pi}{N}}$ is called twiddle factors. Eqs. (1) and (2) can be rewritten as:

$$Y(k) = \sum_{n=0}^{N-1} y(n) W_N^{nk} \quad (3)$$

$$y(n) = \frac{1}{N} \sum_{n=0}^{N-1} Y(k) W_N^{-nk} \quad (4)$$

Where, Wk is known as N^{th} root of the unity. The performance of DFT is dependent on several complex multiplications and additions. The overall time complexity of DFT algorithm is $O(N^2)$. In this work, we have considered FFT algorithm in place of DFT to reduce the time complexity and increase the efficiency of the proposed speech watermarking scheme. The FFT algorithm works based on the divide and conquers approach. It divides the input sample sequence into even and odd indexed subsequence of samples. The overall time complexity of FFT algorithm is $O(N\log_2 N)$.

A. Proposed Embedding Process

The proposed method takes input as a speech signal in a '.wav' format. The speech utterance is converted to frequency domain by computing the FFT. After, watermark data is selected in the form of a grey scale image of same dimension 64×64. This image has been encrypted using Arnold transform method to make it more robust. In the next step, the embedding process has been used with the help of the Min algorithm, which is discussed in subsection 3.2. The speech signal's frequency domain is then changed to accommodate the watermark data, yielding a watermarked signal. Fig. 2 depicts the block diagram of the proposed FFT and vowel-based speech watermarking method.

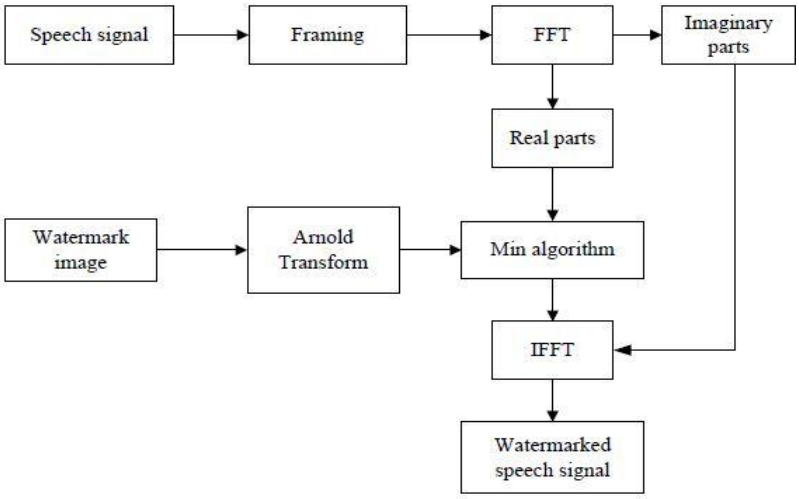

Fig.2. Embedding procedure for the proposed work in watermarking scheme

The embedding process of speech signal into vowel parts is performed by the Algorithm 1.

Algorithm 1: Embedding process

Input S(n): The original speech signal.

Output: Vowel based FFT watermarked speech signals.

Step 1: Take input as the speech signal.

Step 2: Divide the speech signal into distinct frames.

Step 3: Apply the FFT method on each distinct frame and the results are separated in two parts: Real and imaginary.

Step 4: Read a watermark image of same dimension.

Step 5: Apply Arnold transform method for encrypted the watermark image.

Step 6: Embedding process is performed using the Min algorithm (mentioned in section 3.2) on real parts of frequency domain values.

Step 7: Apply the inverse FFT method for reconstructing the speech signal.

Step 8: Finally, watermarked speech signals obtained.

B. Proposed Extraction Process

The watermarked speech signal is used to extract the watermark data. First, watermarked speech signal is converted into the frequency domain and the extract-Min algorithm is applied with the help of reserved watermark information. The pseudo-code for the extract-Min algorithm is discussed in subsection 3.3. In the next step, the decryption process is performed using the inverse Arnold method to extract the watermark. The proposed extraction speech watermarking scheme has been depicted in Fig. 3.

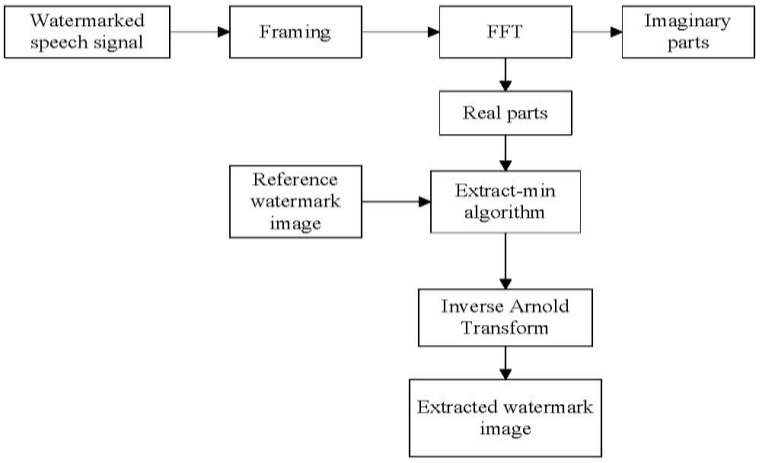

Fig.3. Extracting procedure for the proposed work in watermarking scheme

The extraction process of watermarked speech signal is performedfor the watermark image by the Algorithm 2.

Algorithm 2: Extraction process

Input S(n): The original speech signal.

Output: Vowel based FFT watermarked speech signals.

Step 1: Take input as watermarked speech signals.

Step 2: Input signals is divided into distinct frames.

Step 3: The FFT method is applied and separate into two parts.

Step 4: Read the watermark image and consider it as reference data for the extraction process.

Step 5: Apply the extract-Min algorithm (mentioned in section 3.3) on real parts of the FFT signals for the extraction process.

Step 6: Apply the inverse Arnold transform method for reconstructing the watermark image and finally, we get extracted watermark image.

3.2. Min Algorithm for Watermark Embedding Process

Minimum values have been selected from each frame of the speech signal. Similarly, minimum values from the encrypted watermark image are selected for performing embedding process using Min algorithm. Pseudo code of Min algorithm is as follows.

Step 1: Consider the real part values during FFT computation of the original speech signal as given below:

Z = FFT(frames of original speech);

mag = real_part (Z);

Step 2: Compute minimum values from each frame having real part of FFT, and store them in a variable 'x' as location indices.

(x, indices) = min(real_part, [], 2);

Step 3: Read the watermark image. After that, select minimum values from the watermark image and store them in new variable with new indices.

Wimage = imread(image);

(variable_1, indices_1) = min(W_image, [], 2);

Step 4: variable_2 = variable_1/mean(mean(variable_1));

Step 5: Replace the minimum value of original matrix from minimum values of watermark using following sub-steps:

Step 5.1: for i = 1 to length of variable

Step 5.2: Z(i, indices(i)) = variable_2(i);

Step 5.3: for loop end

Step 6: Finally, the minimum values of a watermark image is inserted into each frame of the speech signal.

3.3. Min Algorithm for Watermark Extraction Process

After insertion of the watermark, the minimum value location may be changed in the watermarked speech signal. To handle this problem, we must save the new indices with values of watermarked speech for the extraction process. We have performed extract-Min algorithm to extract the watermark image. Pseudo code of extract-Min algorithm for extraction process as follows:

Step 1: Compute the actual minimum value indices from the watermark data, and then consider the real part values from the FFT of watermarked speech.

wd = FFT(frames of watermarked speech);

wd_mag = wd_real (wd);

(wd_variable, indices_3) = min(wd_mag, [], 2);

Step 2: Store the watermark indices along with values in the memory to use these indices.

(variable_1, indices_1) = min(W_image, [], 2);

Step 3: variable_2 = variable_1/mean(mean(variable_1));

Step 4: Replace the minimum value of original matrix from minimum values of watermark image.

Step 4.1: for i = 1 to length of watermark image
Step 4.2: W_image (i, indices_1(i)) = wd (i, indices_3 (i));
Step 4.3: W_image (i, indices_1(i)) = W_image (i, indices_1(i))*(variable_2);
Step 4.4: for loop end.

Step 5: Finally, the minimum values of a watermark image is extracted from each frame of the watermarked speech signal.

3.4. Arnold Transform(AT)

On image scrambling, the Arnold transform technique is built. It can be used for equal size resolution, which is only defined for squares. It has also been used to create rectangles, but in order to produce a square; we would need to add extra rows or columns. Because it is straightforward and periodic, it can be used in digital watermarking procedures. The watermark image will eventually revert to its original appearance because of its periodic nature [23]. The AT function is expressed using following equation:

$$\begin{Bmatrix} a' \\ b' \end{Bmatrix} = \begin{Bmatrix} 1 & 1 \\ 1 & 2 \end{Bmatrix} \cdot \begin{Bmatrix} a \\ b \end{Bmatrix} . (\text{mod } n)$$

Where n is the image's width. a and b are individual pixel's coordinates.

The AT function can be rebuilt and written as follows:

$$\begin{Bmatrix} a'_1 \\ b'_1 \end{Bmatrix} = \begin{Bmatrix} 2 & -1 \\ -1 & 2 \end{Bmatrix} \cdot \begin{Bmatrix} a_1 \\ b_1 \end{Bmatrix} . (\text{mod } n)$$

4. Quality Evaluation Parameters for Speech Watermarking Techniques

The proposed schemes are evaluated using various parameters such as PSNR, NC, structure similarity, and BER as discussed in the following subsections:

4.1. Peak Signal to Noise Ratio (PSNR)

The PSNR values evaluate using following equation [16]:

$$psnr = 10 * \log_{10} \frac{m^2}{M} \quad (5)$$

Where m represents maximum value and M is the MSE value.

4.2. Normalised Correlation (NC)

The NC values have been assessed in order to determine how resistant the watermark is to various attacks. When comparing the original and extracted watermark images, the NC parameter is employed. SSIM is also used to compare the quality between two signals [14].

5. Simulation Results and Discussion

We have performed experiments on 300 speech signals from TIMIT database and a watermark image to perform watermarking procedure. The sample of speech is divided into frames with 25 ms frame duration, and the sampling rate is 8 kHz. The FFT method has been applied on each frame of original signals to transform the spatial domain into the frequency domain for further processing. The FFT methods have been performed on TIMIT database.

Table 1 shows the PSNR and NC values obtained between the original and extracted watermark signals. The average PSNR value is around 51 dB, and the average NC value is 1, indicating that original image is statistically similar to extracted watermark image. We have determined that the proposed method gives better results.

Table 1. Quality and robustness measures using PSNR and NC parameters based on original and extracted watermark signal

Types of signals	**Average estimated values using vowel and FFT method**	
	PSNR (in dB)	**NC**
Watermarked speech	43.512	1.000
Extracted watermark	51.352	1.000

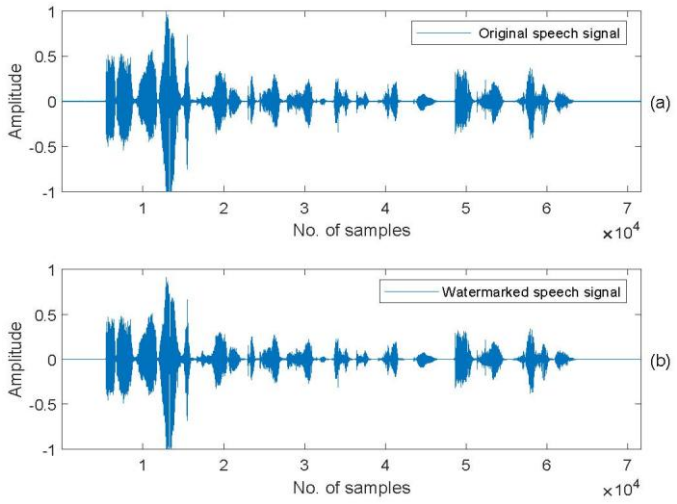

Fig.4. Speech utterance: (a) original speech utterance, and (b) watermarked speech utterance

From fig. 4, we displayed the original speech and watermarked speech signals waveforms. In waveform, both the signals have same signature. Thus, we have concluded that the signals are same. From fig. 5, it shows watermark and extracted watermark images of same dimension 64 × 64.

Fig.5. (a) Watermark lena and (b) extracted lena watermark images

Table 2 shows the results comparison on various attacks using parameters PSNR, NC, and SSIM between FFT methods. The first column represents the name of attacks. Columns 2, 3, and 4 show the computed PSNR, NC and SSIM values, respectively for FFT method.

Table 2. Performance of attacks based on parameters PSNR, NC and SSIM for the proposed watermarked speech signal

Name of attacks	Proposed method		
	PSNR (dB)	NC	SSIM
No attack	43.512	1.0000	1.0000
salt and pepper noise (0.02)	19.1719	0.2548	0.5864
speckle noise (0.04)	18.1623	0.6052	0.5897
Gaussian noise	17.1519	0.2039	0.0234
Poission noise	23.1892	0.6169	0.6086
Re-quantization	21.1252	0.4882	0.6082
Re sampling	21.2528	0.4312	0.5881
Low pass filtering	23.1052	0.5882	0.5974

From Table 2, we have concluded that the FFT method gives better results. The imperceptibility and robustness performance of the proposed FFT method is high on each evaluation parameters such as PSNR, NC and SSIM. The main reason for the higher performance is that the FFT method uses simple transformation of complex values which contains high and low energies. Due to this, the quality of speech is not degraded in the case of the FFT method. In the FFT method, it transforms into the frequency domain then split into complex values. That is why the FFT method may change some values, and consequently, the quality is degraded.

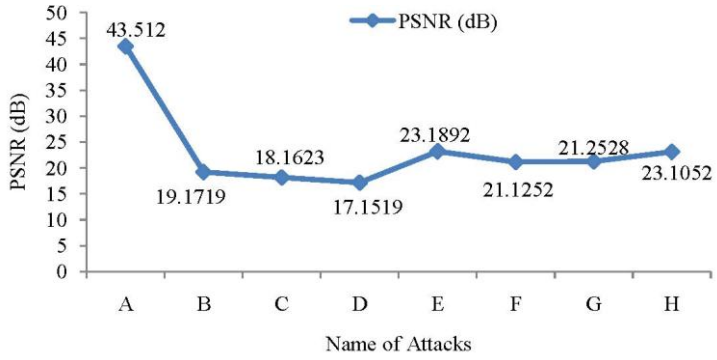

Fig.6. Graphical representation for the different attacks as to PSNR(dB). (Where identification number are as obey: A. No attack, B. salt and pepper noise (0.02), C. speckle noise (0.04), D. Gaussian noise, E. poission noise, F. Re-quantization, G. Re-sampling, H. Low pass filtering, respectively).

The results of several attacks are depicted in Fig. 6 in terms of quality measurements. This graphic's horizontal axis depicts different attack as to identification numbers, as detailed in Table 2. The identification numbers 'A' for no attack, 'B' for salt and pepper noise (0.02), 'C' for speckle noise (0.04), 'D' for Gaussian noise, 'E' for Poisson noise, 'F' for re-quantization, 'G' for re-sampling, and 'H' for low pass filtering are shown in Fig. 6. From Fig. 6, when we analyzed without performing any attack, the quality of watermarked signals have high. It means that there is no degradation occurs in the quality of watermarked signals. After performing some different type of attacks that values are sustained the quality of watermarked signals.

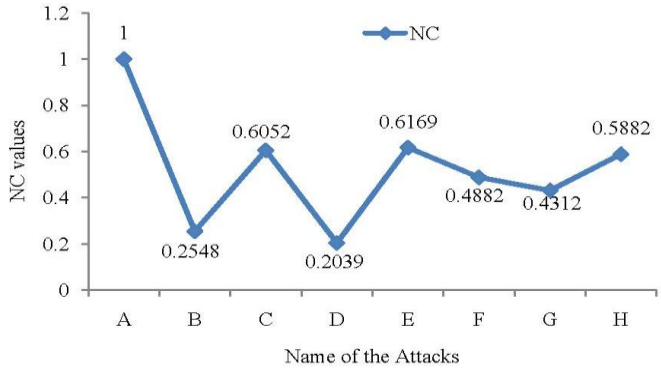

Fig.7. Pictorial demonstration for the various attacks in terms of robustness (NC values)

Fig. 7 demonstrates the robustness as to NC values. The robustness of the watermarked signals is maintained after performing different kinds of attacks. The NC values lies between 0 and 1. The proposed method gives NC value 1, which means robustness is 100%.

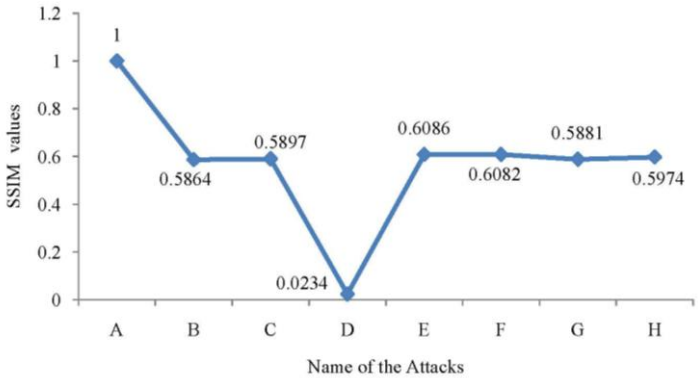

Fig.8. Demonstrate the various attacks in terms of structure similarity (SSIM values)

Fig. 8 represents the structure similarity index measurement in terms of SSIM values. The SSIM values depict how similar the original signals and watermarked signals are to one another (See from fig. 4). The SSIM values of the proposed approach are greater than those of the existing methods, as shown in fig. 8.

Table 3. Comparing the outcomes of proposed and existing approaches for watermarked signals based on PSNR and BER

Reference	PSNR (dB)	BER (in %)
Based on DWT method using speech as a watermark [24]	49.34015	0.5849
Based on DWT method using image as a watermark [24]	22.37976	2.2631
Proposed method	51.3525	0.075

The existing scheme [24] is based on the features of the DWT method. In the DWT method, the speech signal is decomposed into two different sub-bands, i.e., high and low energy sub-bands in the time domain. Due to the division of speech signals, the voice quality is degraded.

Table 3 indicates the comparison between existing and proposed method for an original and extracted watermark. In this table, two parameters PSNR and BER (in percentage) have been considered to compare both existing and proposed methods. From Table 3, it is clear that the proposed vowel and FFT based speech watermarking technique is better than existing methods.

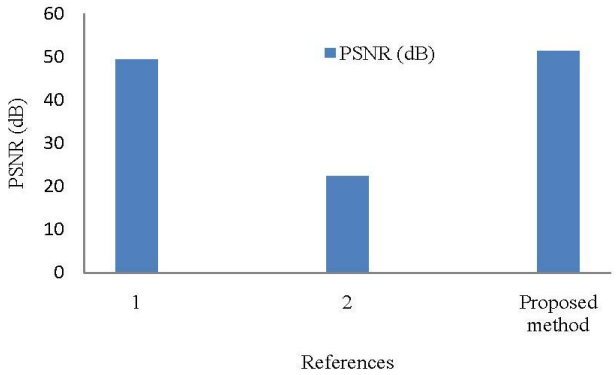

Fig.9. Comparison between existing and proposed methods as to PSNR(dB). (Where numbers are as obey: 1. Based on DWT method using speech as a watermark, 2. Based on DWT method using image as a watermark.)

The comparison of PSNR values between existing approaches and the proposed method is shown in Fig. 9 (See Table 3). Fig. 9 shows that the proposed approach outperforms the existing methods as to PSNR.

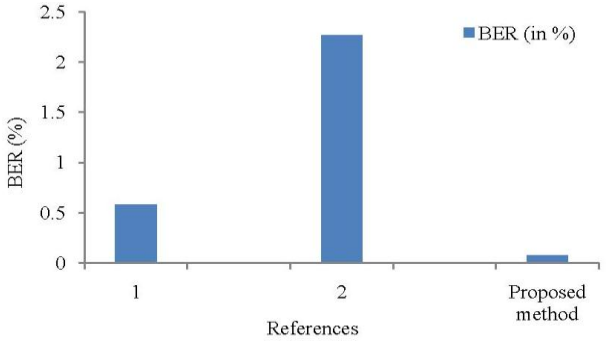

Fig.10. Comparison between existing and proposed methods as to BER(%). (Where identification numbers are as obey: 1. Based on DWT method using speech as a watermark, 2. Based on DWT method using image as a watermark)

Fig. 10 represents the comparison of BER (in %) between existing methods and the proposed method (See Table 3). From fig. 10, it is clear that the proposed method produces less error compared to the existing methods.

6. Conclusions

Using the proposed FFT, the Arnold transform, and Min algorithms, we have presented vowel-based digital speech watermarking approaches in this article. The primary focus of our methods is to provide security and robustness along with copyright protection of speech watermarking. The results of proposed method show that before and after the signals have slight differences, it means that the proposed scheme provide us a best quality signal. We have compared proposed method to other method. The proposed speech production-based FFT scheme produces the BER 0.075% which is approximately 2.2% less than the existing method. The PSNR values for the proposed method are approximately 2 db higher than existing methods. This work may be extended to recover the tamper detection.

References

[1] Y. He and Y. Hu. 'A proposed digital image watermarking based on DWT-DCT-SVD', in 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), May, pp.1214–1218, 2018.

[2] S. Bhattacharya, T. Chattopadhyay, and A. Pal, "A survey on different video watermarking techniques and comparative analysis with reference to H. 264/AVC," in 2006 IEEE International Symposium on Consumer Electronics. IEEE, 2006, pp. 1–6.

[3] V. Singh, "Digital watermarking: a tutorial," Cyber Journals, Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), January Edition, 2011.

[4] D. Ambika & V. Radha, "Speech Watermarking Using Discrete Wavelet Transform", Discrete Cosine Transform And Singular Value Decomposition. Int. J. Comput. Sci. Eng. Technol, 5(11), 1089-1093, 2014.

[5] M. A. Nematollahi and S. A. R. Al-Haddad, "An overview of digital speech watermarking," International Journal of Speech Technology, vol. 16, no. 4, pp. 471–488, 2013.

[6] A. Jadhav and M. Kolhekar, "Digital watermarking in video for copyright protection," in 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies. IEEE, 2014, pp. 140–144.

[7] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, "Real-time labeling of MPEG-2 compressed video," Journal of Visual Communication and Image Representation, vol. 9, no. 4, pp. 256–270, 1998.

[8] M. S. Kankanhalli, K. Ramakrishnan et al., "Adaptive visible watermarking of images," in Proceedings IEEE International Conference on Multimedia Computing and Systems, vol. 1. IEEE, 1999, pp. 568–573.

[9] S. P. Mohanty, K. Ramakrishnan, and M. Kankanhalli, "A dual watermarking technique for images," in Proceedings of the seventh ACM international conference on Multimedia (Part 2). Citeseer, 1999, pp. 49–51.

[10] M. M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in Proceedings of international conference on image processing, vol. 2. IEEE, 1997, pp. 680–683.

[11] F. Perez-Gonzalez and J. R. Hernandez, "A tutorial on digital watermarking," in Proceedings IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology (Cat. No. 99CH36303). IEEE, 1999, pp. 286–292.

[12] O. Jane and E. Elbai, 'Hybrid non-blind watermarking based on DWT and SVD', Journal of Applied Research and Technology, Vol. 12, No. 4, pp.750–761, 2014.

[13] I. J. Cox and M. L. Miller, "Electronic watermarking: the first 50 years," in 2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No. 01TH8564). IEEE, 2001, pp. 225–230.

[14] R. Kumar and J. Yadav, "Vowel and non-vowel frame segmentation based digital speech watermarking technique using lpa method," Journal of Information Security and Applications, vol. 68, p. 103218, 2022.

[15] Q. Wu, R. Ding, and J. Wei, "Audio watermarking algorithm with a synchronization mechanism based on spectrum distribution," Security and Communication Networks, vol. 2022, 2022.

[16] A. Kanhe and A. Gnanasekaran, "Robust image-in-audio watermarking technique based on DCT-SVD transform," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2018, no. 1, p. 16, 2018.

[17] A. Singha and M. A. Ullah, "Audio watermarking with multiple images as watermarks," IETE Journal of Education, vol. 61, no. 2, pp. 64–75, 2020.

[18] Z. Liu, Y. Huang, and J. Huang, "Patchwork-based audio watermark- ing robust against de-synchronization and recapturing attacks," IEEE Transactions on Information Forensics and Security, vol. 14, no. 5, pp. 1171–1180, 2019.

[19] M. Charfeddine, E. Mezghani, S. Masmoudi, C. B. Amar, and H. Al- humyani, "Audio watermarking for security and non-security applica- tions," IEEE Access, vol. 10, pp. 12 654–12 677, 2022.

[20] S. Alshathri and E. E.-D. Hemdan, "An efficient audio watermarking scheme with scrambled medical images for secure medical internet of things systems," Multimedia Tools and Applications, pp. 1–19, 2023.

[21] K. O. Boateng, B. W. Asubam, and D. S. Laar, "Improving the effectiveness of the median filter," 2012.

[22] R.-S. Run, S.-J. Horng, J.-L. Lai, T.-W. Kao, and R.-J. Chen, "An improved SVD-based watermarking technique for copyright protection," Expert Systems with applications, vol. 39, no. 1, pp. 673–689, 2012.

[23] L. Min, L. Ting, and H. Yu-jie, "Arnold transform based image scrambling method," in 3rd International Conference on Multimedia Technology(ICMT-13). Atlantis Press, 2013/11. [Online]. Available: https://doi.org/10.2991/icmt-13.2013.160

[24] A. Revathi, N. Sasikaladevi, and C. Jeyalakshmi, "Digital speech water- marking to enhance the security using speech as a biometric for person authentication," International Journal of Speech Technology, vol. 21, no. 4, pp. 1021–1031, 2018.

Authors' Profiles

Dr. Rajeev Kumar is an Assistant Professor in the Faculty of Information Technology, Gopal Narayan Singh University, Jamuhar, Sasaram, Bihar. He has passed graduation (B.Tech) from BPUT, Rourkela and post-graduated (M.Tech) from Central University of Punjab, Bathinda in computer science. He has completed Ph.D from central university of south Bihar in computer science stream. His area of research is signal processing, image compression, video, image, and speech watermarking. He has published 14 papers in the reputed conferences, SCOPUS and SCI journals. He has also published three books in different publication.

Kshitiz Singh is a Ph.D. scholar at the Central University of South Bihar in Gaya. He has a Masters of Computer Science from Banaras Hindu University, Varanasi. He has been working as an Information Scientist in the Central Library, Gaya, Bihar, India, since July 12th, 2017. Image processing, networking, audio forgery, and software development are some of his research interests.

Dr. Jainath Yadav is an Associate 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.

Ajay Kumar is working as an Assistant professor in the Department of Computer Science and Informatics at Central University of Himachal Pradesh Dharamshala. He has 1.5 years industry and 9.5 years teaching experience. He is pursuing his Ph.D. from Central University of Jammu and received MCA and B.Sc. (H) degree from University of Delhi and Banaras Hindu University respectively. His research interest includes UAV, IoT, Blockchain Technology and Machine Learning. He was visiting fellow IIT Delhi and IIT Ropar. He has published numerous research papers in international journals and conferences. He has completed one minor project in the area of IoT having title “Make your life easy using smart switches”. He was a jury member of Smart India Hackathon software edition 2020 organized by Ministry of Education govt. of India. He recently got the best paper award at MISS2021. He is reviewer and program committee member of many international journals and conferences. He attended many conferences, FDP’s and workshops. He has organized many conferences and workshops. He is an active professional member of the ACM, IEEE and IAENG.

Dr. Indranath Chatterjee is working as a Professor in the Department of Computer Engineering at Tongmyong University, Busan, South Korea. He received his Ph. D. in Computational Neuroscience from the Department of Computer Science, University of Delhi, Delhi, India. His research areas include Computational Neuroscience, Schizophrenia, Medical Imaging, fMRI, and Machine learning. He has authored and edited 8 books on Computer Science and Neuroscience published by renowned international publishers. To date, he has published numerous research papers in international journals and conferences. He is a recipient of various global awards on neuroscience. He is currently serving as a Chief Section Editor of a few renowned international journals and serving as a member of the Advisory board and Editorial board of various international journals and Open-Science organizations worldwide. He is presently working on several projects of government & non-government organizations such as PI/co-PI, related to medical imaging and machine learning for a broader societal impact, in collaboration with several universities globally. He is an active professional member of the Association of Computing Machinery (ACM, USA), Organization of Human Brain Mapping (OHBM, USA), Federations of European Neuroscience Society (FENS, Belgium), Association for Clinical Neurology and Mental Health (ACNM, India), and International Neuroinformatics Coordinating Facility (INCF, Sweden).

How to cite this paper: Rajeev Kumar, Kshitiz Singh, Jainath Yadav, Ajay Kumar, Indranath Chatterjee, "Vowel based Speech Watermarking Techniques using FFT and Min Algorithm", International Journal of Computer Network and Information Security(IJCNIS), Vol.17, No.1, pp.57-67, 2025. DOI:10.5815/ijcnis.2025.01.05