INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2019, VOL. 65, NO. 3, PP. 527-532
Manuscript received October 17, 2018; revised August, 2019. DOI: 10.24425/ijet.2019.129809

Localization of Copy-Move Forgery in Speech Signals Through Watermarking Using DCT-QIM

N.V. Lalitha, Ch. Srinivasa Rao, and P.V.Y. JayaSree

***Abstract*—Digital speech copyright protection and forgery identification are the prevalent issues in our advancing digital world. In speech forgery, voiced part of the speech signal is copied and pasted to a specific location which alters the meaning of the speech signal. Watermarking can be used to safe guard the copyrights of the owner. To detect copy-move forgeries a transform domain watermarking method is proposed. In the proposed method, watermarking is achieved through Discrete Cosine Transform (DCT) and Quantization Index Modulation (QIM) rule. Hash bits are also inserted in watermarked voice segments to detect Copy-Move Forgery (CMF) in speech signals. Proposed method is evaluated on two databases and achieved good imperceptibility. It exhibits robustness in detecting the watermark and forgeries against signal processing attacks such as resample, low-pass filtering, jittering, compression and cropping. The proposed work contributes for forensics analysis in speech signals. This proposed work also compared with the some of the state-of-art methods.**

***Keywords*—watermarking, copy-move forgery, Discrete Cosine Transform, Quantization Index Modulation, Hash bits**

I. INTRODUCTION

WITH rapid strides in multimedia technology leads to easy access and transmission of multimedia content. This creates a demand for ownership authentication and forgery detection of the same. A well-known and traditional way of solution for this problem is watermarking [1], [2]. It is the process of embedding ownership information as a watermark into a speech signal. This information is further extracted to claim the authentication. If watermark information directly embedded into signal samples then that type of techniques are called time domain techniques [3], [4], [5]. If watermark is inserted into transform coefficients then those are called transform domain techniques [6], [7], [8]. Transform domain techniques are more robust than the other one.

M. A. Nematollahi et.al., [9] proposed a semi-fragile and blind speech watermarking technique based on the Discrete Wavelet Packet Transform (DWPT) and QIM. In this paper, watermark is embedded within an angle of the wavelet's sub-bands. Same authors in [10], proposed a robust speech watermarking technique based on DWPT. Here watermark embeds in the amplitudes of the wavelet's sub-band. S. Wang et.al., [11] presented a speech watermarking based on Double DCT (DDCT) and QIM. Authors claim that this technique provides satisfactory robustness and tamper resistance. M. A. Nematollahi et.al., [12] presented a blind speech watermarking based on Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD). Here, speech signal is divided into frames and applied DWT then SVD to compute eigen values of approximation coefficients. Authors reported that this work is robust against attacks. In [13], authors proposed a synchronized blind speech watermarking based on two-level DWT and Adaptive Mean Modulation (AMM). Here, watermark bits and synchronized bits are embedded in selected second level detail and approximation sub-bands using AMM method.

Most of these works concentrate on watermarking process and are robust to signal processing attacks but cannot detect Copy-Move Forgery (CMF). In speech CMF, voiced part of the speech signal is copied and pasted to a specific location which alters the meaning of the speech signal. To address this problem, in the proposed method, speech signal is divided into voice segments to embed the watermark. Hash bits are generated and inserted in that watermarked voice segments which made the method novel in detecting the CMF. Watermark is useful for copyright protection purpose and hash bits are useful for forgery detection.

The paper organization is as follows. Materials and Methods are discussed in section II. Methodology of the proposed work is discussed in section III. Experiment results and discussion about the work is given in section IV. Conclusions of this work are reported in section V.

II. MATERIALS AND METHODS

In this proposed method, speech signal is transformed by using DCT and QIM technique is used for watermark insertion. These are discussed in this section.

A. Discrete Cosine Transform (DCT)

DCT transformation is used in this work to convert the signal from spatial to frequency domain. DCT [14] transformation decomposes the signal into series of cosine harmonics and it is computationally simple than FFT. One dimensional DCT transformation of signal $f(x)$ and inverse DCT transformation of $D(n)$ are shown in Equation (1) and Equation (2).

$$D(n) = \beta(n) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x+1)n}{2N}\right], \quad n = 0,1,2,\dots N-1 \qquad (1)$$

$$f(x) = \sum_{n=0}^{N-1} \beta(n) D(n) cos\left[\frac{\pi(2x+1)n}{2N}\right], \qquad x = 0,1,2,\dots N-1 \quad (2)$$

$$\text{Where, } \beta(n) = \begin{cases} \sqrt{\frac{2}{N}}, & for\ n \neq 0 \\ \sqrt{\frac{1}{N}}, & for\ n = 0 \end{cases} \qquad (3)$$

N.V.Lalitha is with the Dept. of Electronics and Communication Engineering, GMR Institute of Technology, Rajam, A.P, India (e-mail: lalitha.nv@gmrit.edu.in).

Ch.Srinivasa Rao is with the Dept. of Electronics and Communication Engineering, JNTU-K University, Vizianagaram, A.P, India (e-mail:chsrao.ece@jntukucev.ac.in).

P.V.Y.Jayasree is with the Dept. of Electronics and Communication Engineering, GIT, GITAM University, Visakhapatnam, A.P, India (e-mail: pvyjayasree@gitam.edu).

B. Quantization Index Modulation (QIM)

In this QIM [15] method, watermark is inserted into the signal by changing the quantizers shown in Equation (4). This method produces high imperceptibility and robustness to amplitude-related attacks.

$$em_i = f_Q(s_i, w_i, \emptyset) \tag{4}$$

Where, f_Q is the quantization modulation function, s_i is the host signal, w_i is the watermark, $\emptyset$ is the quantization strength and em_i is the embedded signal.
In extraction process, watermark is extracted by using below Equation (5)

$$ew = \begin{cases} 1, & if\ g_Q(em_i', \emptyset) > Threshold \\ 0, & otherwise \end{cases} \tag{5}$$

Where, g_Q is the quantization demodulation function, em_i' is the watermarked signal and ew is extracted watermark.

III. METHODOLOGY

In this proposed work, speech signal is divided to voiced segments. For each segment, N X N size pre-processed watermark is embedded to protect the speech signal for unauthorized claiming. Hash bits also generated to each voiced segment and inserted into the binary form of speech segment for detection and identification of CMF

A. Watermark embedding process

N X N size binary watermark image is used for embedding process. This watermark is preprocessed using Gaussian map [16] chaotic encryption for increasing the security to the watermark. Gaussian chaotic map is defined below Equation (6) and Equation (7).

$$P_{i+1} = \exp(-\varphi P_i^2) + \sigma \tag{6}$$

Where P_i is the initial value in the range 0 and 1, and φ, σ are real parameters.

$$g_i = \begin{cases} 1, & if\ P_i > 1/4 \\ 0, & otherwise \end{cases} \tag{7}$$

This g_i binary sequence is EX-OR with watermark binary sequence to get encrypted sequence e_i.

N X N size encrypted watermark is embedded into DCT coefficients of each voiced segment based on QIM rule shown in Equation (8) below.

$$em_i = \begin{cases} round\left[\frac{as_i}{\emptyset}\right]\emptyset, & if\ e_i = 0 \\ \left(floor\left[\frac{as_i}{\emptyset}\right]\emptyset\right) + \frac{\emptyset}{2}, & if\ e_i = 1 \end{cases} \quad i = 1,2,\ldots\ldots N \times N \tag{8}$$

Where, as_i is the sample of the transformed voice segment, $\emptyset$ is embedding strength and em_i is embedded coefficient.

This multiple embedding of watermark is useful in extraction of watermark from attacked speech signal. Due to attack, any segment is corrupted then watermark cannot be recovered properly, hence watermark can be extracted from any other segment. Generate 128-bit hash [17] for each watermarked segment and insert those bits into binary form of same watermarked segment for forgery detection. This process is repeated for all voice segments and it is shown in Fig. 1.

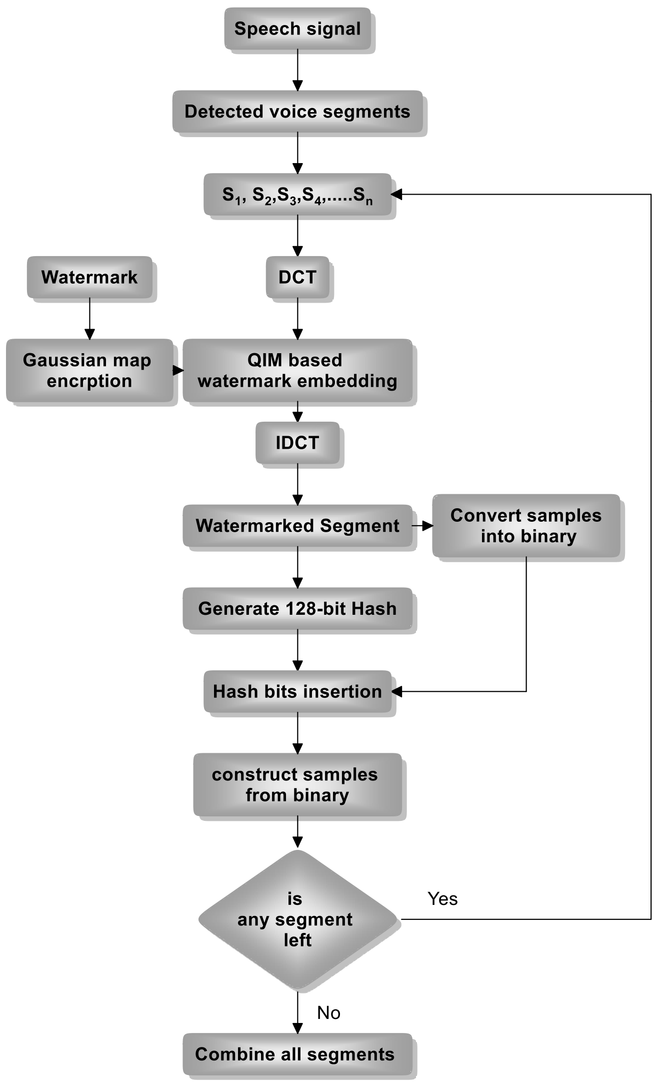

Fig. 1. Watermark and Hash embedding flow

B. Watermark extraction process

In extraction process, divide the incoming forged speech signal into voice segments. Extract the encrypted watermark from the DCT coefficients of speech segment using QIM extraction rule as shown in Equation (9) below.

$$e_i' = \begin{cases} 1, & if\ \frac{\emptyset}{4} \le mod(as_i', \emptyset) < \frac{3\emptyset}{4} \\ 0, & otherwise \end{cases} \tag{9}$$

Where, as_i' is the i^{th} transformed coefficient of the incoming forged speech signal and e_i' are the extracted bits. These bits are further decrypted by inverse Gaussian map to get extracted watermark. BER is calculated to this extracted watermark and original watermark. This watermark extraction process is applied to each segment and which segment got minimum BER that segment is considered as a valid extracted watermark. Hence, watermark is extracted from all segments and low BER for extracted watermark validates the watermark. This watermark extraction flow is shown in Fig.2.

C. CMF detection and identification process

In this process, forged speech signal is divided to voice segments and converted into binary samples. Recover the 128 bit hash from those binary samples and this process is repeated for all segments. Recovered hash bits are compared with other segment hash bits using BER. If calculated BER is less than the predefined threshold then those two corresponding segments are identified as copy-move forged segments, otherwise incoming speech signal is considered as no forgery signal. This process flow is shown in below Fig. 2.

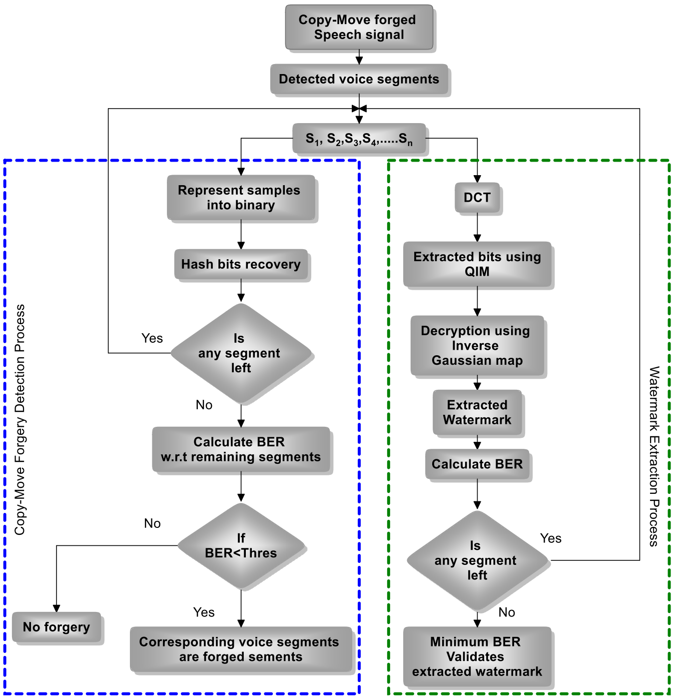

Fig. 2. Watermark Extraction and CMF identification process

IV. Experiment Results

The proposed method is tested on two datasets i. VoxForge database [18] of 4 seconds and ii. Speech signals of 10 seconds. Second dataset is developed by recording voice articulated by male, female and child in three different languages. All are .wav, mono, 44.1 KHz sampling frequency and 16-bit quantization speech signals. 64 X 64 binary image is considered as watermark and it is encrypted using Gaussian map. Fig.3 illustrates original watermark image and its encrypted version.

To measure the imperceptibility of the proposed work metric SNR is calculated based on the Equation (10) shown below.

$$SNR = 10\,log\left[\frac{\sum_{n=1}^{L} s^2(n)}{\sum_{n=1}^{L}(s(n)-s'(n))^2}\right] \quad (10)$$

Where, L is the speech length, $s(n)$ is the original speech and $s'(n)$ is the speech after embedding watermark and Hash bits. SNR values of 21 speech signals (SS_1 to SS_21) after embedding watermark and hash bits is shown in Table I.

Fig. 3. Watermark image and its encrypted image

Average SNR is 46.0633 dB and it is >20 dB (as per IFPI requirement). This SNR is comparatively high when compared with [19], [6], [20], [21] and [22]. SNR of a speech signal is varied with respective to embedding quantization strength and that plot is shown in Fig. 4.

TABLE I
SNR VALUES FOR DIFFERENT SPEECH SIGNALS

Speech signal	SNR	Speech signal	SNR	Speech signal	SNR
SS_1	45.724	SS_8	46.0839	SS_15	43.3783
SS_2	43.3886	SS_9	40.737	SS_16	55.3804
SS_3	39.1009	SS_10	35.5507	SS_17	53.8065
SS_4	46.324	SS_11	47.7097	SS_18	51.7504
SS_5	43.0961	SS_12	37.6553	SS_19	59.8948
SS_6	42.3989	SS_13	45.4921	SS_20	55.4633
SS_7	35.7496	SS_14	41.88	SS_21	46.3054

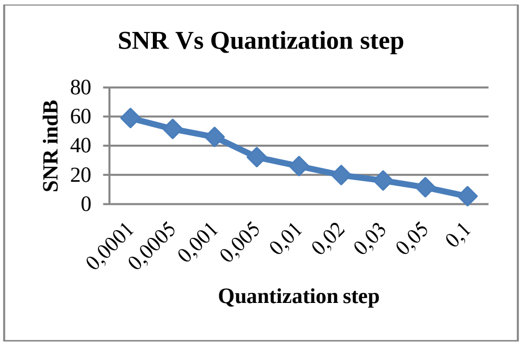

Fig. 4. Plot for SNR Vs embedding quantization strength for a signal

SNR and capacity of this proposed method for two different length signals given in Table II. If length of the signal is increased, capacity decreases because same size of watermark (64 X 64) is embedded in both the signals. In this proposed method, watermark is embedded in the voiced segments only because of this reason; SNR of 4 sec speech signal got more compared to 10 sec speech signal.

Original speech signal and its separated voice segments are shown in Fig. 5a and Fig. 5b respectively. In this example voice segment S3 is copied and pasted in place of segment S7 shown in Fig. 5c. This copied and pasted voice segments are detected by using proposed method and shown in Fig. 5d.

TABLE II
SNR AND CAPACITY OF THE TWO DIFFERENT LENGTH SPEECH SIGNALS

Signal duration	4 sec	10 sec
SNR (dB)	53.7668	42.2846
Capacity (bps)	1024	409.6

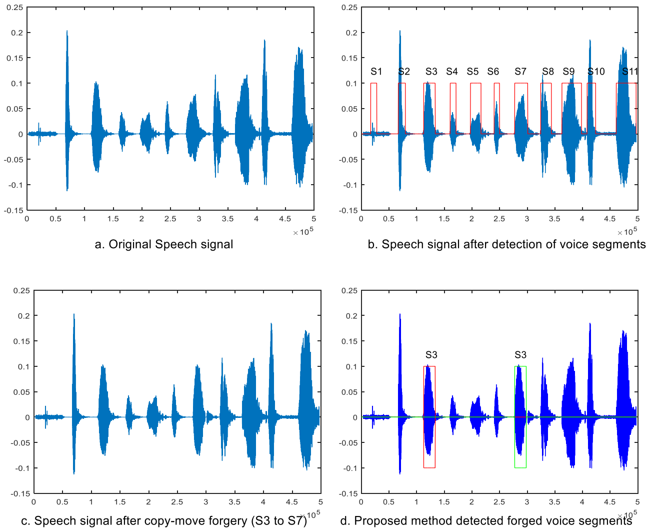

Fig. 5. Original speech signal, forged speech and detected voice segments

The performance of the proposed work is tested on some signal processing attacks and descriptions of these attacks are given below.

a. Without attack (W_A): without any signal processing attack, watermark is extracted.
b. Resampling (R): original speech signal sampling frequency is 44.1 KHz. In this attack, signal is sampled to half of the original frequency and resampled back.
c. Low-pass filtering (LPF_1): signal is passed through 10 KHz cut-off frequency low-pass filter.
d. Low-pass filtering (LPF_2): signal is passed through 16 KHz cut-off frequency low-pass filter.
e. Jittering (J): removing of one sample among one lakh samples.
f. MP3 Compression (MP3_1): signal is compressed with 256 Kbps and back to .wav.
g. MP3 Compression (MP3_2): signal is compressed with 160 Kbps and back to .wav.
h. Cropping (C_1): starting 1000 samples of the signal are replaced with zeros.
i. Cropping (C_2): middle 1000 samples of the signal are replaced with zeros.
j. Cropping (C_3): ending 1000 samples of the signal are replaced with zeros.

The robustness of the proposed work on the attacked signals are evaluated with BER metric and is measured based on the below equation. These BER values listed in Table III.

$$BER = \frac{Number\ of\ error\ bits}{Total\ number\ of\ bits} \tag{11}$$

TABLE III
BER VALUES OF EXTRACTED WATERMA1

Speech signal	W_A	R	LPF_1	LPF_2	J	MP3_1	MP3_2	C_1	C_2	C_3
SS_1	0	0.0004	0.3483	0.0744	0	0	0.0075	0	0	0
SS_2	0	0.0012	0.4670	0.1106	0	0.0009	0.0473	0	0	0
SS_3	0	0	0.1601	0.0051	0	0	0.0002	0	0	0
SS_4	0	0	0.3325	0.0771	0	0	0.0046	0	0	0
SS_5	0	0	0.3359	0.0949	0	0	0.0109	0	0	0
SS_6	0	0	0.2644	0.0605	0	0	0.0163	0	0	0
SS_7	0	0	0.1359	0.0417	0.3527	0	0.0004	0	0	0
SS_8	0	0	0.3015	0.0715	0	0	0.0051	0	0	0
SS_9	0	0	0.3457	0.1418	0	0.0002	0.0097	0	0	0
SS_10	0	0	0.2570	0.0073	0	0	0.0024	0	0	0
SS_11	0	0	0.3681	0.0949	0	0	0.0046	0	0	0
SS_12	0	0	0.4511	0.2722	0	0.0021	0.1909	0	0	0
SS_13	0	0	0.3732	0.1347	0	0.0009	0.0866	0	0	0
SS_14	0	0	0.4047	0.2426	0.0004	0.0017	0.1010	0	0	0
SS_15	0	0	0.4152	0.2897	0	0.0034	0.2346	0	0	0
SS_16	0	0.0126	0.4863	0.4362	0	0.0341	0.3549	0	0	0
SS_17	0	0.1674	0.4565	0.4494	0	0.0546	0.4638	0.0097	0	0
SS_18	0	0.0378	0.4973	0.3859	0	0.0285	0.2695	0	0	0
SS_19	0	0.1901	0.4282	0.4211	0.4465	0.1301	0.4956	0	0	0
SS_20	0	0.0087	0.4755	0.4597	0.4643	0.0090	0.2690	0	0	0
SS_21	0	0.0498	0.4875	0.4775	0.4542	0.0292	0.3852	0	0	0

This proposed method has able to detect copy-move forged voice segments. This method is also tested against the same signal processing attacks which are mentioned above. Recovered hash bits in each voice segment are compared with remaining voice segments using metric BER. If BER is less than the predefined threshold value then those corresponding segments are considered as CMF segments otherwise considered as forgery has not occurred. Table IV shows the CMF segments are detected or not for signal processing attacks. If forgery is identified then 'Yes' is reported otherwise 'No' is reported in the Table IV.

TABLE IV
CMF DETECTION WITH ATTACKS

Speech signal	W_A	R	LPF_1	LPF_2	J	MP3_1	MP3_2	C_1	C_2	C_3
SS_1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
SS_2	Yes	No	No	No	No	No	No	Yes	Yes	Yes
SS_3	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SS_4	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
SS_5	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Yes
SS_6	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SS_7	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SS_8	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SS_9	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SS_10	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes
SS_11	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
SS_12	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SS_13	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes
SS_14	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes
SS_15	Yes	No	No	Yes	No	Yes	No	Yes	Yes	Yes
SS_16	Yes	No	Yes	Yes	No	No	No	Yes	Yes	Yes
SS_17	Yes	No	No	No	No	No	No	Yes	Yes	Yes
SS_18	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SS_19	Yes	No	No	No	No	No	No	Yes	Yes	Yes
SS_20	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SS_21	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SP22	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

It is evident from the Table IV, that the proposed method is resilient to all post-processing attacks except attacks jittering and MP3 compression.

SNR, Capacity and BER of the proposed method is compared with state-of-art methods and shown in Table V and Table VI respectively.

TABLE V
SNR AND CAPACITY COMPARISON WITH WORK [6]

Speech (4 sec)	A. Merrad et al. [6]	Proposed method
SNR (dB)	33.1824	53.7668
Capacity (bps)	204.8	1024

TABLE VI
BER COMPARISON OF THE PROPOSED WORK WITH STATE-OF-ART WORKS

	A. Al-haj and A.Mohammad [7]	M.A.Nematollahi et al. [12]	Z. Liu and H. Wang [23]	Proposed method
No attack	0.0924	0	-	0
LPF (cut off=22 kHz)	0.4648	0.3130	-	0
Filter (44.1-11.025-44.1)	-	-	0.0839	0.0031

V. CONCLUSIONS

In this paper, a transform based speech watermarking using DCT and QIM is proposed. Proposed method can provide copyright protection and detect CMF. Hash bits are inserted in watermarked voice segments to make the proposed method to detect CMF in speech signals. Watermark is useful for copyright protection purpose and hash bits are useful for forgery detection. Experimental results shows that the proposed method achieved good imperceptibility and robustness against signal processing attacks viz., resample, low-pass filtering, jittering, compression and cropping. As well, this method can detect CMF in the presence of signal processing attacks. The proposed work can be adopted for forensics analysis in speech signals.

REFERENCES

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, 1997.

[2] P. Taylor, V. S. Verma, and R. K. Jha, “An Overview of Robust Digital Image Watermarking,” *IETE Tech. Rev.*, no. June, pp. 1–18, 2015.

[3] A. N. Lemma, J. Aprea, W. Oomen, and L. Van de Kerkhof, “A temporal domain audio watermarking technique,” *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1088–1097, 2003.

[4] M. Unoki and R. Miyauchi, “Robust, blindly-detectable, and semi-reversible technique of audio watermarking based on cochlear delay characteristics,” *IEICE Trans. Inf. Syst.*, vol. E98–D, no. 1, pp. 38–48, 2015.

[5] G. Hua, J. Goh, and V. L. L. Thing, “Time-spread echo-based audio watermarking with optimized imperceptibility and robustness,” *IEEE/ACM Trans. Speech Lang. Process.*, vol. 23, no. 2, pp. 227–239, 2015.

[6] A. Merrad, A. Benziane, S. Saadi, and A. Hafaifa, “Robust Blind Approach for Digital Speech Watermarking,” in *2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, 2018.

[7] A. Al-haj and A. Mohammad, “Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decomposition,” *Eur. J. Sci. Res.*, vol. 39, no. 1, pp. 6–21, 2010.

[8] X.-Y. Wang and H. Zhao, “A novel synchronization invariant audio watermarking scheme based on DWT and DCT,” *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4835–4840, 2006.

[9] M. A. Nematollahi, H. Gamboa-Rosales, M. A. Akhaee, and S. A. R. Al-Haddad, “Semi-fragile digital speech watermarking for online speaker recognition,” *EURASIP J. Audio, Speech, Music Process.*, vol. 2015, no. 31, pp. 1–15, 2015.

[10] M. A. Nematollahi, H. Gamboa-rosales, M. A. Akhaee, and S. A. . Al-Haddad, “Robust Digital SpeechWatermarking For Online Speaker Recognition,” *Math. Probl. Eng.*, vol. 2015, pp. 1–12, 2015.

[11] S. Wang, X. Liu, X. Dang, and J. Wang, “A Robust Speech Watermarking Based On Quantization Index Modulation And Double Discrete Cosine Transform,” in *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2017)*, 2017.

[12] M. A. Nematollahi, S. A. R. Al-Haddad, and F. Zarafshan, "Blind digital speech watermarking based on Eigen-value quantization in DWT," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 27, no. 1, pp. 58–67, 2015.

[13] H. T. Hu, S. J. Lin, and L. Y. Hsu, "Effective blind speech watermarking via adaptive mean modulation and package synchronization in DWT domain," *Eurasip J. Audio, Speech, Music Process.*, vol. 2017, no. 1, pp. 1–14, 2017.

[14] Subir and A. M. Joshi, "DWT-DCT based blind audio watermarking using Arnold scrambling and Cyclic codes," in *3rd International Conference on Signal Processing and Integrated Networks, SPIN 2016*, 2016, pp. 79–84.

[15] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[16] P. K. Dhar and T. Shimamura, "Audio Watermarking in Transform Domain Based on Singular Value Decomposition and Cartesian-polar transformation," *Int. J. Speech Technol.*, vol. 17, pp. 133–144, 2014.

[17] S. Self-recovery, S. Sarreshtedari, S. Member, and M. A. Akhaee, "A Watermarking Method for Digital Speech Self-Recovery," *IEEE/Acm Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 1917–1925, 2015.

[18] VoxForge speech database "http://www.repository.voxforge1.org/ downloads/SpeechCorpus/Trunk/Audio/Original/44.1kHz_16bit/tle." .

[19] D. Avci, T. Tuncer, and E. Avci, "A New Information Hiding Method for Audio Signals," in *6th International Symposium on Digital Forensic and Security (ISDFS)*, 2018, pp. 3–6.

[20] P. K. Dhar and T. Shimamura, "Blind Audio Watermarking in Transform Domain Based on Singular Value Decomposition and Exponential-Log Operations," *Radio Eng.*, vol. 26, no. 2, pp. 552–561, 2017.

[21] H. Hu and J. Chang, "Efficient and robust frame-synchronized blind audio watermarking by featuring multilevel DWT and DCT," *Cluster Comput.*, vol. 20, no. 1, pp. 805–816, 2017.

[22] B. Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT," *Signal Processing*, vol. 91, no. 8, pp. 1973–1984, 2011.

[23] Z. Liu and H. Wang, "A novel speech content authentication algorithm based on Bessel–Fourier moments," *Digit. Signal Process. A Rev. J.*, vol. 24, pp. 197–208, 2014.