Authentications and Temper Recovery using Speech Watermarking Techniques: A Review

(IJCSNT) International Journal of Communication Systems and Network Technologies
ISSN Number: 2053-6283
Volume 9, Issue No. 1, 32–39
©The Author(s) 2020
https://journals.mirlabs.in/index.php/ijcsnt
DOI-10.18486/ijcsnt/9.1.114
IJCSNT

Rajeev Kumar[1], Jainath Yadav[2]

Abstract

Due to the gradual increase of Internet users, the main problem arises as duplicity and copying that led to content integrity and protection. We need some techniques to handle/secure the speech signals from an illegitimate user of digital content. These issues can be solved using speech watermarking schemes that provide to ensure content authentication and temper recovery. Content authentication is the validation of content integrity. Digital speech watermarking techniques is now in limelight to protecting our speech content from unauthorized copying. In this paper, we have considered the speech watermarking methods along with their important properties. In this paper, we have focused on the theoretical analysis of the important aspects of the forgery. This paper has also included a summary of work on authentication and tamper detection.

Keywords

Speech Watermarking, Authentication, Temper Recovery, Integrity, Attack

Received: 03 January 2020; **Revised:** 31 March 2020; **Accepted:** 24 April 2020;
Published: 30 April 2020;

[1]Department of Computer Science, Central University of South Bihar, Gaya, India.
[2]Department of Computer Science, Central University of South Bihar, Gaya, India.

Corresponding author:
Email: rajeev6136@gmail.com

Introduction

Digital media is communicated throughout the world by the use of the Internet. Hence, the content of authentication, and copyright comfort, temper detection, integrity, etc., of the media are the major concerns for them. Digital speech content authentication having some drawbacks, i.e., substitution attack[1], they cannot locate the frames which are attacked[2], and they cannot verify the authenticity of the watermarked signal detected[3,4]. The best possible solution for these concerns can be ensured and protected by the use of digital watermarking techniques. Digital speech is unique concerning the audio signal regarding factors like creation model, recognition, data transmission, loudness, and intensity. Digital watermarking strategies have truly been utilized to guarantee security as far as possession assurance and tamper-proofing for a wide assortment of information positions[5].

The current watermarking strategies are founded on either the time or frequency domain. However, in both cases, the time-frequency characteristics of the watermark do not relate to the time-frequency attributes of speech signals. It might cause watermark discernibility because the watermark will be available in the time-frequency areas where speech segments do not exist. There are numerous methodologies for speech watermarking, including Least significant Bit (LSB)[6], spread spectrum (SS), auditory masking, patchwork, transformation[7–9], and parametric modeling. In the SS approach, a pseudo arbitrary grouping is utilized to spread the range of the watermark information and add it to the frequency spectrum of the host signal. The patchwork methodology embeds the watermark information by controlling two arrangements of the signal to decide the contrast between them. The transformation approach embeds the watermark information into the transformation spaces. The parametric modeling embeds the watermark by adjusting the coefficients of the autoregressive (AR) model.

In this paper, we have discussed the theoretical analysis of content authentication and temper recovery using speech watermarking techniques. Also, we reviewed quality and robustness factors and their values from various work. This paper has been organized as follows: the first section includes the introduction about speech content authentication and tamper detection. The second section consists of literature reviews based on speech content authentication. Section 3 shows the result discussion and the last section represents the summary of this paper along with future work.

Review based on Authentication

In this section, we have discussed the digital watermarking technique based on the authentication process, presented below:

Shi et al.[10] have proposed the integrity authentication algorithm based on the perceptual hash function and learned dictionaries method. This method is performed based on gammatone filter model. The proposed scheme is performed in terms of the SNR, ODG, and SDG. This method is verifying its robustness against common signal processing. Revathi et al.[11] have described enhancing the security of speaker authentication using a biometric system. For this, they used the DWT method for the watermark embedding area and the feature selection process is used for authentication. Simulation parameters are based on PSNR, BER, and Perceptual evaluation speech quality (PESQ).

Sun et al.[12] have proposed the speech authentication method based on high-capacity watermark embedding techniques. The embedded process is done in the low-frequency area, which is selected using segmentation and the DCT method. Simulation parameters are BER, objective difference grades (ODG), and subjective difference grades (SDG). The result shows effective robustness, verifying the content and

security. Liu et al.[13] have proposed authentication and tamper recovery for speech watermarking using the DCT method. The watermark information is generated by frame number and compressed signal. Digital requirements are achieved using BER, SDG, and ODG.

Liu et al.[14] have proposed a dual image watermarking system for authentication and copyright protection. They achieved this technique using a robust and fragile watermarking method. The DWT and quantization methods are used for watermark insertion in YCbCr color space. Extraction is done using blind fragile based on the LSB method for image authentication. The simulation results are evaluated based on PSNR and SSIM methods and it shows effective results on various attacks. Sarreshtedari et al.[15] proposed a watermarking scheme for digital speech self-recovery. They introduced digital self-embedding speech signals and the self-recovery feature. Simulation results based on the Tolerable Tampering Rate (TTR) and PSNR show that the method is robust and secure.

Nematollahi et al.[16] have described an effective, robust, and inaudible audio and speech watermarking algorithm based on the discrete wavelet transform (DWT) and the singular value decomposition (SVD). The simulation results obtained the effectiveness of audio watermarking as a reliable solution to the copyright protection problem which is facing the music industry. The simulation results show that the proposed scheme is more effective compared to estimated audio quality in terms of some quality parameters with all types of music files, but on various attacks, it is not much more effective.

Saraswathi[17] proposed a speech authentication method detecting error based on Mel Frequency Cepstral Coefficients (MFCC) features section process. The beginning, middle, and ending parts are selected to embed the watermark into a low-intensity value position. Any modification is done the forgery is detected in the extracted features. Jiao et al.[18] have proposed speech signals for authentication based on feature extraction such as Linear prediction coding (LPC) and the DCT method. This model is based on two phases; feature extraction and hash modelling. The linear spectrum frequencies (LSF) are used for hash generation. The low frequency of DCT coefficients is taken to embed the watermark. In this way, results show that robustness against content preserving operations.

Comparative Analysis of Results on Authentication

In this section, we have analyzed the quality of watermarked speech signals in terms of objective difference grade (ODG) and subjective difference grade (SDG). The major concern is to maintain the high imperceptibility. The quality of the original speech signal should not degrade after inserting the watermark. Now, the problem is how to evaluate accurately, and it has very difficult to acoustically differentiate the original speech from the watermarked speech. In this paper, we have considered two parameter values of different methods: the first one is ODG and the second one is SDG. The ODG method has measures the objective quality of the watermarked speech signals, and the SDG has used to measure the subjective quality of the watermarked speech signals.

Table 1 represents the general grading system for ODG and SDG methods, and it consists of three columns. The first column holds the ODG/SDG grade values, i.e., “0”, “-1”, “-2”, “-3”, and “-4”; the second column contains speech quality, i.e., “Excellent”, “Favourable”, “Fair”, “Poor”, and “Bad”. The last column shows the watermark inaudibility, i.e., “Imperceptible”, “Audible but not annoying”, “Audible but slightly annoying”, “Audible but annoying”, and “Audible but very annoying”. In this table, ODG/SDG range has varied from 0 to -4. Each value has a different meaning in terms of speech quality

and the human auditory system, i.e., 0 means speech quality is excellent, -4 means very bad quality which is not audible form.

Table 2 consists of five columns: the first column represents the range in terms of minimum, mean, and maximum values, the second column represents the result of DCT based method, the third column is based on the high capacity embedding method. The fourth column has based on a novel method, and the last column consists of perceptual speech hash and learned dictionaries methods. This table consists of the values based on the ODG and SDG on different existing works. In this table, we compare the state of art-works concerning ODG and SDG. We consider three ranges value-form each existing method and each paper result has shown the quality of watermarked speech signal is lies between [0,-1]. It means that the objective and subjective method represents approximately excellent quality. The watermarking concepts of[9,19] have also been considered for the getting better results of authentication, which lies in the high range of performance.

Table 1. The general grading system for ODG and SDG

ODG/SDG	Speech Quality	Watermark Inaudibility
0	Excellent	Imperceptible
-1	Favorable	Audible but not annoying
-2	Fair	Audible but slightly annoying
-3	Poor	Audible but annoying
-4	Bad	Audible but very annoying

Table 2. Result comparison among existing work of watermarked signal on ODG and SDG

Range	In[13]	In[12]	In[20]	In[10]
Based on ODG				
Minimum	-0.892	-0.342	-0.709	-0.411
Mean	-0.764	-0.274	-0.673	-0.372
Maximum	-0.637	-0.206	-0.638	-0.266
Based on SDG				
Minimum	-0.7182	0	-0.8542	-0.38
Mean	-0.6611	0	-0.7918	-0.33
Maximum	-0.6041	0	-0.7295	-0.23

Table 3 represents the performance of robustness on various attacks which has evaluated using the Bit Error Rate (BER) method. In this table, we compare four different existing work based on various types of attacks. In the case of "0" BER, it represents no error occur at all, and if BER is more than this is called some errors have occurred that depend on the value. This table holds three different attacks which have evaluated at different scale and we have concluded that values from each paper. In this way, we can say that the robustness has also maintained and also content authentication.

Table 3. Result comparison among existing work on various attacks using BER

Types of attack	In [13]	In [12]	In [20]	In [10]
Based on BER				
MP3 Compression (32 kbps)	3.68	1.26	2.3826	0
MP3 Compression (64 kbps)	0	0	0.7235	0
Re-sampling (44.1 → 8 → 44.1 kHz)	–	0	0.3954	0
Re-sampling (44.1 → 11.025 → 44.1 kHz)	0.3583	0	0	0
Re-sampling (44.1 → 16 → 44.1 kHz)	0.2156	0	0	0
Low pass filtering (11,025 Hz)	0	0	0.2651	0

Figure 1 shows the comparison of the existing work based on ODG from table 2. In this figure, we have considered three value ranges, i.e., maximum, mean, and minimum. It also represents the ODG grade values have not exceeded -0.9, which means watermarked speech signals have maintained the quality for each case.

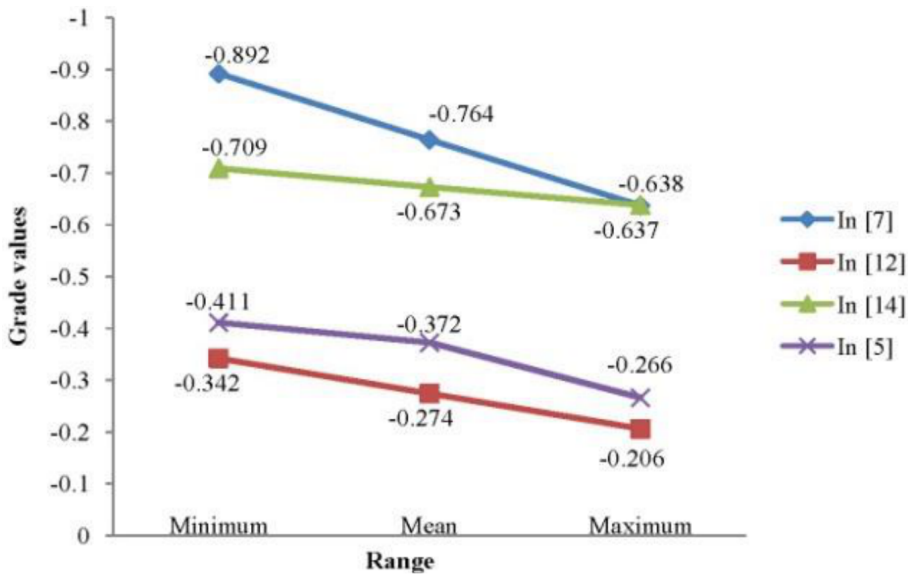

Figure 1. Graphical representation of ODG results based on existing works

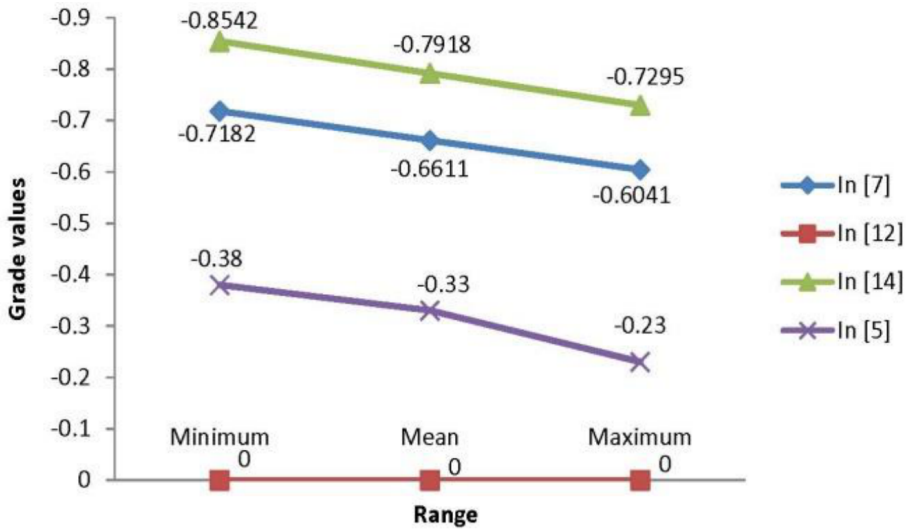

Figure 2. Graphical representation of SDG results based on existing works

Figure 2 shows the comparison of the state-of-art methods based on SDG from table 2. In this figure, we have considered three value ranges, i.e., maximum, mean, and minimum. It also represents the SDG grade values have not exceeded -0.9, which means watermarked speech signals have maintained the quality for each case.

Conclusion

The comparative review study has been done based on content authentication and temper recovery. In this paper, we analyze the objective and subjective result that provides the imperceptibility level for watermarked speech signals. Hence, we include minimum, mean and maximum values of ODG and SDG values concerning grade values. The robustness can be found after attacks by the using of BER that values included from various existing works. In this paper, we consider authentication and tamper detection related work and their results. If the percentage of BER values is zero and the imperceptibility value is high then we can say that watermarking methods are helpful to protect content authentication. In future work, we implement this work and check how much efficiency can achieve.

References

[1] Liu ZH and Wang HX. A novel speech content authentication algorithm based on bessel-fourier moments. *Digital Signal Processing* 2014; 24: 197–208.

[2] Wang HX and Fan MQ. Centroid-based semi-fragile audio watermarking in hybrid domain. *Science in China Series F-Information Sciences* 2010; 53(3): 619–633.

[3] Bai YL, Ing YS and Zhen L. Blind and robust audio watermarking scheme based on svd-dct. *Signal Processing* 2011; 91(8): 1973–1984.

[4] Vivekananda BK, Indranil S and Abhijit D. A new audio watermarking scheme based on singular value decomposition and quantization. *Circuits, Systems, and Signal Processing* 2011; 30(5): 915–927.

[5] Kumar G, Brar ESS, Kumar R et al. A review: Dwt-dct technique and arithmetic-huffman coding based image compression. *International Journal of Engineering and Manufacturing* 2015; 5(3): 20.

[6] Kumar M, Kumar R and Yadav J. A robust digital speech watermarking based on least significant bit, 2020.

[7] Kumar R, Kumar S and Brar SS. Video quality evaluation using dwt-spiht based watermarking technique. In *2016 11th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, pp. 1–6.

[8] Kumar R and Yadav J. Speech watermarking techniques using arnold transform based on multi dimension multilevel dwt method ; 83: 22661–22671.

[9] Bedi SS, Tomar GS and Verma S. Npt based video watermarking with non-overlapping block matching. *Springer-Verlag Berlin Heidelberg Transactions on Computational Science* 2010; 11: 270–292.

[10] Shi C, Li X and Wang H. A novel integrity authentication algorithm based on perceptual speech hash and learned dictionaries. *IEEE Access* 2020; 8: 22249–22265.

[11] Revathi A, Sasikaladevi N and Jeyalakshmi C. Digital speech watermarking to enhance the security using speech as a biometric for person authentication. *International Journal of Speech Technology* 2018; 21(4): 1021–1031.

[12] Sun F, Liu Z and Qi C. Speech content authentication scheme based on high-capacity watermark embedding. *International Journal of Digital Crime and Forensics (IJDCF)* 2017; 9(2): 1–14.

[13] Liu Z, Zhang F, Wang J et al. Authentication and recovery algorithm for speech signal based on digital watermarking. *Signal Processing* 2016; 123: 157–166.

[14] Liu XL, Lin CC and Yuan SM. Blind dual watermarking for color images' authentication and copyright protection. *IEEE Transactions on circuits and systems for video technology* 2016; 28(5): 1047–1055.

[15] Sarreshtedari S, Akhaee MA and Abbasfar A. A watermarking method for digital speech self-recovery. *IEEE/ACM transactions on audio, speech, and language processing* 2015; 23(11): 1917–1925.

[16] Nematollahi MA, Al-Haddad SAR, Doraisamy S et al. Digital audio and speech watermarking based on the multiple discrete wavelets transform and singular value decomposition. In *2012 Sixth Asia Modelling Symposium*. IEEE, pp. 109–114.

[17] Saraswathi S. Speech authentication based on audio watermarking. *International Journal of Information Technology* 2010; 16(1): 34–43.

[18] Jiao Y, Ji L and Niu X. Robust speech hashing for content authentication. *IEEE Signal Processing Letters* 2009; 16(9): 818–821.

[19] Bedi SS, Tomar GS and Verma S. Robust watermarking of image in the transform domain using edge detection. In *IEEE International Conference on simulation UKSIM 2009*. pp. 233–238.

[20] Wang J and He J. A speech content authentication algorithm based on a novel watermarking method. *Multimedia Tools and Applications* 2017; 76(13): 14799–14814.