INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2019, VOL. 65, NO. 1, PP. 19-24
Manuscript received July 12, 2018; revised January, 2019. DOI: 10.24425/123560

Security of Electronic Patient Record using Imperceptible DCT-SVD based Audio Watermarking Technique

Aniruddha Kanhe, Aghila Gnanasekaran

***Abstract*—A robust and highly imperceptible audio watermarking technique is presented to secure the electronic patient record of Parkinson's Disease (PD) affected patient. The proposed DCT-SVD based watermarking technique introduces minimal changes in speech such that the accuracy in classification of PD affected person's speech and healthy person's speech is retained. To achieve high imperceptibility the voiced part of the speech is considered for embedding the watermark. It is shown that the proposed watermarking technique is robust to common signal processing attacks. The practicability of the proposed technique is tested: by creating an android application to record & watermark the speech signal. The classification of PD affected speech is done using Support Vector Machine (SVM) classifier in cloud server.**

***Keywords*—Watermarking, DCT-SVD, Parkinson's Disease, EPR**

I. Introduction

INCREASE in wearable devices and growth in cloudlet based technology has increased the popularity of telediagnosis systems. This requires the transmission of data from patient to doctor over an unsecured network. Thus, the security and confidentiality of Electronic Patient Record (EPR) is the key issue in telediagnosis systems [1]. The sharing of such critical data over the network always suffers with a potential risk of privacy, security and authenticity. The attacker having unauthorized access to one's medical data can instigate various attacks such as: false data injection, selective reporting, alteration in health data. Security of EPR for medical images using watermarking techniques have already been proposed for e-healthcare system. Application of watermarking for security of EPR using speech signal is an emerging field and less work has been reported so far.

Speech feature based PD detection techniques are proposed in recent years utilizing SVM and k-Neural Network (k-NN) based classifiers. In [3], proposed PD detection technique using genetic algorithm (GA) based SVM classifier. The optimized speech features are selected using GA for classifying the healthy and PD affected speech using SVM classifier. Little et al. In [4], proposed SVM based classification technique by finding the optimized set of features. Zhang [5], proposed a smart phone based approach for PD detection using deep neural network method and achieved a better accuracy of detection. These studies were focused on accuracy in identification of PD affected speech without considering the security aspect of the EPR.

However, security of EPR for medical images using watermarking techniques are presented in [6], [7], [8], [9] for e-healthcare system. Application of watermarking for security of EPR using speech signal is an emerging field and less work has been reported so far. Alhusssein and Muhammad [10] proposed cloud based framework for EPR security and PD detection. In [10], the DWT-SVD based watermarking technique is used for securing the EPR but the classification of PD based speech is not performed using the watermarked audio. Z. Ali et al. [11], proposed a cryptographic approach for securing the PD audio.

The watermarking techniques in audio signals are broadly classified as temporal domain techniques and frequency/wavelet domain techniques. The temporal domain techniques utilizes least significant bit (LSB) substitution and echo hiding techniques [12], [13], [14], [15], [16], [17], in which the watermark bits are embedded in LSB positions or by adding small echoes in the original audio. In frequency/wavelet domain watermarking techniques the watermark is embedded in the coefficients of transforms such as Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) and DCT [18], [19], [20], [21]

In this paper, we present a framework to ensure the security of EPR of a PD suspected person and its smart telediagnosis using cloud computing environment. The EPR is embedded in the patient's speech signal, using proposed discrete cosine transform - singular value decomposition (DCT-SVD) based watermarking technique through an android application. The watermarked speech is transmitted to the cloud for automatic classification of PD affected person's speech from healthy person's speech using SVM classifier as shown in Fig. 1. Jitter, shimmer, pulse parameters, pitch parameters and harmonicity parameters are extracted from the watermarked PD suspected speech for automatic classification.

II. Framework

The proposed cloud-based telediagnosis framework consist of cloud management module (CMM), resource management module (RMM) and service manager. The watermark speech along with the side data enters the cloud through CMM where the patient is registered with an unique ID and the authenticity is verified. The data required to extract the watermark is

Aniruddha Kanhe is with Department of Electronics and communication Engineering, National Institute of Technology Puducherry, Karaikal, India (e-mail: aniruddhakanhe@nitpy.ac.in).

Aghila Gnanasekaran is with Department of Computer Science and Engineering, National Institute of Technology Puducherry, Karaikal, India (e-mail: aghila@nitpy.ac.in).

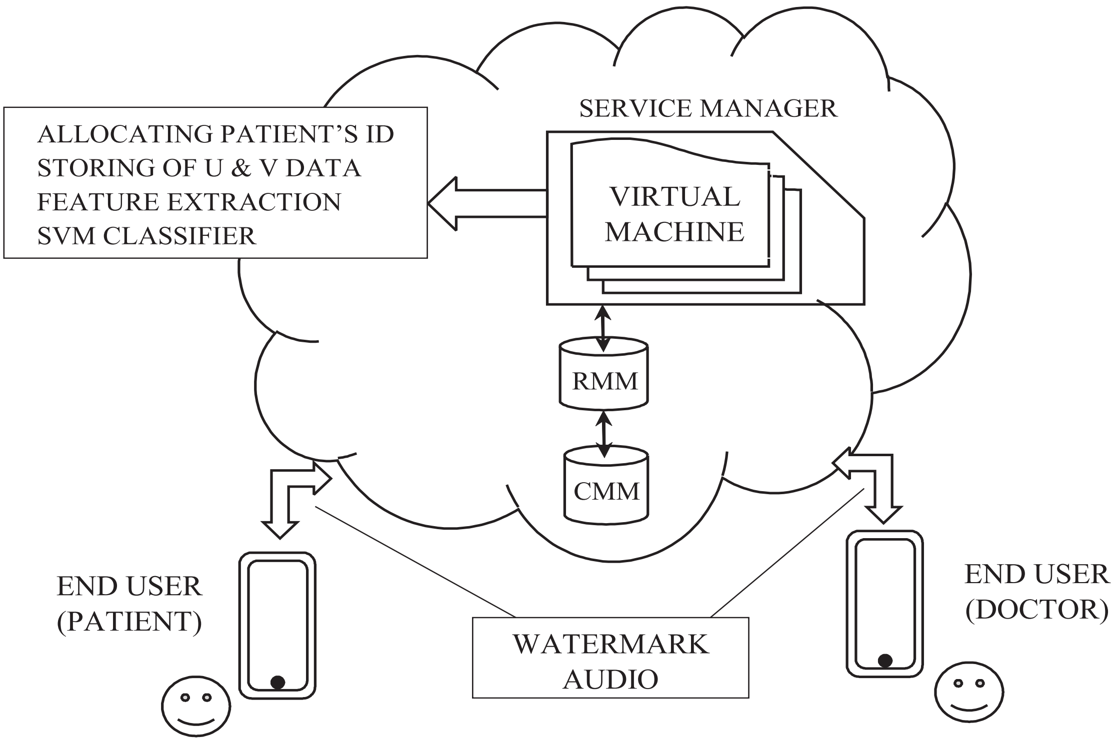

Fig. 1. Cloud-based framework for watermarking and classification using SVM

defined as side data. The RMM allocates different virtual machines (VM) to store the ID and side data to extract the watermark. These VM servers does the processing of watermarked speech to extract spectral features and classify them using SVM. Once the classification of speech is done, the generated initial report is securely transmitted to the patient, using unique patient's ID and the watermarked speech is transmitted to doctor's mobile phone module. The system consist of three parts: watermark embedding, classification using SVM and watermark extraction.

III. Watermark embedding

The name and age of patient is considered as the watermark for the present work and is converted into binary bit stream before embedding. The masking property of human auditory system (HAS) is utilized to embed the watermark in high energy low frequency voiced frames (HELF) of the patient's speech signal. The speech signal are first divided into frames of $10ms$ duration and HELF frames are identified by short time energy (STE) using:

$$E_m = \sum [s(n)w(m-n)]^2 \quad (1)$$

$$w(n) = \begin{cases} 0.54 - 0.46 * cos\frac{2\pi n}{L-1} & for\ 0 \leq n \leq L-1 \\ 0 & Otherwise \end{cases} \quad (2)$$

where $s(n)$ represents a signal, and E_m is the short time energy and zero crossing count (ZCC):

$$ZCC = \sum_{n=0}^{N-1} 0.5|sign(s[n]) - sign(s[n-1])| \quad (3)$$

where

$$sign(s[n]) = \begin{cases} +1 & if\ s[n] \geq 0 \\ -1 & Otherwise \end{cases} \quad (4)$$

The DCT operation is performed on these HELF frames using:

$$X(k) = w(k) \sum_{n=0}^{N-1} x(n) cos \frac{(2n+1)k\pi}{2N}, k = 0, 1, ..., N-1 \quad (5)$$

where

$$w(k) = \begin{cases} \sqrt{\frac{1}{N}} & if\ k=0 \\ \sqrt{\frac{2}{N}} & Otherwise \end{cases} \quad (6)$$

where, $x(n)$ is the input voiced frame and N is length of DCT.

The DCT coefficients are arranged in 4×4 matrix and the SVD operation is performed to get U, S and V matrices. In the singular matrix S, 12 secret scaled EPR bits are embedded as watermark and the SVD operation is performed again to generate side data. The watermark speech is generated by performing the inverse SVD operation on modified singular matrix S followed by inverse DCT transform. The energy compaction property of DCT and robustness of SVD against signal processing operation [2] is exploited in the proposed watermarking technique. The patient's android phone module accomplishes the task of watermarking the speech and transmitting the watermarked speech to cloud. Fig. 2 shows the PD affected speech sample before and after embedding watermark using proposed watermarking technique.

IV. Watermark extraction

The extraction of EPR is done using the proposed android app. Whenever the EPR is requested by authentic enduser the service manager of cloud provides the pre-stored U and V matrices and the watermarked speech frames are identified using HELF method used in embedding technique. The DCT followed by the SVD operation is performed on the watermarked HELF frames to obtain the scaled singular matrix using pre-stored side data. The EPR bits are extracted by upscaling the non-diagonal elements of the matrix obtained from previous operation.

The watermark extraction requires $[U_1]$ and $[V_1]^T$ matrices; the steps involved in extraction of EPR are as follows: label=Step 0:,ref=Step 0, wide=0pt

1) The steps 1 and 2 of embedding process is repeated with the watermarked speech to obtain the coefficient matrix $[B]$.
2) The SVD operation is performed on this $[B]$ matrix to obtain the singular matrix $[S_w]$.
3) Using $[U_1]$ and $[V_1]^T$ matrices, inverse SVD operation is performed on $[S_w]$ to generate the $[D_w]$ matrix containing scaled watermark bits.

$$[U_1] \times [S_w] \times [V_1]^T = [D_w]$$

4) The embedded EPR bits are extracted from $[D_w]$ by:

$$b_i = \begin{cases} 1 & for \quad D_{w(ij)} \geq \epsilon \\ 0 & for \quad D_{w(ij)} < \epsilon \end{cases} \tag{7}$$

where,

$$\epsilon = avg[D_{w(ij)}] \quad \forall \quad i \neq j$$

Whenever the EPR is requested by authentic enduser to the service manager, these steps are repeated for all the watermarked frames to obtain the EPR.

As per the procedure mentioned in section 3.1.1, after the embedding of EPR, the speech is transmitted to cloud where the feature extraction is performed to classify the PD affected speech. The implementation of classifier is presented in next Subsection.

V. Classification of speech SVM

The classification of watermarked PD affected speech from healthy person's speech is done using SVM classifier in the cloud server. The training and testing dataset are created by extracting frequency parameters, pulse parameters, amplitude parameters, voicing parameters, pitch parameters and harmonicity parameters. The performance of the classifier is evaluated by computing classification accuracy.

The high computational power of cloud is used for feature extraction and classification. The person affected with PD typically suffers with dysphonia [4]. It discribes the vocal impairment of PD affected person where the patient cannot produce the normal voice[22]. The dysphonia can be identified by three major speech features which includes: perturbation in fundamental frequency (jitter measures), perturbation in amplitude (shimmer measures) and signal to noise ratios (harmonics to noise measures) [25].

A. Feature extraction and classification

The SVM classifier is used for classifying the PD affected person's speech and healthy person's speech because of its simplicity and high accuracy [22]. The training and test sets are created using leave-one-subject-out (LOSO) validation scheme. LOSO scheme provides higher accuracy than the conventional bootstrapping or leave-one-out validation scheme [22].

The description of data set used for experimentation and the results associated with testing of proposed watermarking technique against various signal processing attacks and the classification accuracy using SVM classifier are presented in the following section.

VI. Experimental Results

The proposed watermarking technique is implemented and tested using MATLAB R2016b software. These codes are converted into C code using MATLAB coder. These C codes are used to develop an android app using android studio software where the speech can be recorded/uploaded by patients. The proposed framework is tested using PD speech signal database created by Department of Neurology in Cerrahpasa, Faculty of Medicine, Istanbul University [22]. The database consists of voice samples of 20 PD affected persons which include 6 females and 14 males and 20 healthy individuals consisting of 10 females and 10 males. Multiple speech samples are recorded from all individuals which includes vowels, words, number and short sentences. The speech samples of healthy persons are obtained from Saarbruken voice disorder database [23].

The robustness of the proposed watermarking technique is evaluated by computing bit error rate (BER) against common signal processing attack such as re-sampling, re-quantization and MP3 compression. The SNR and BER results obtained for the proposed watermarking technique are listed in Table I.

TABLE I
SNR AND BER RESULTS

Parameter	Proposed Technique
SNR	86dB
BER for re-sampling at 6KHZ	0.001%
BER for re-quantization at 24	0%
BER for MP3 compression at 96kbps	0.0011%

A. Evaluation of watermarking technique

The performance of watermarking technique is evaluated on three basic parameters: imperceptibility, robustness and payload [18]. The imperceptibility is quantified using signal to

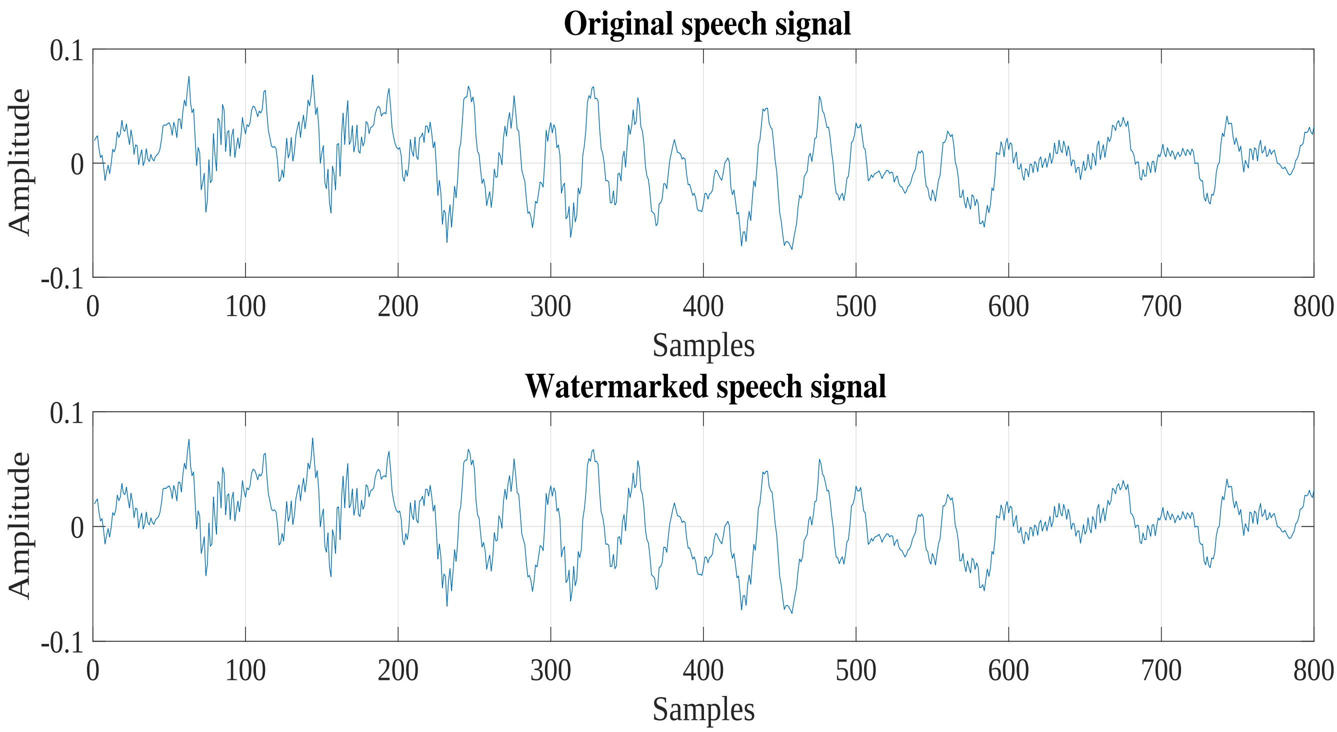

Fig. 2. PD affected audio sample for vowel a before and after watermarking.

noise ratio (SNR) and subjective listening test. Robustness of watermarking technique is analyzed by computing bit error rate (BER) subjected to the signal processing attacks such as: re-sampling, re-quantization, additive white Gaussian noise (AWGN) and MP3 compression. The payload define, number bits embedded in speech signal per second.

a) SNR: To calculate SNR, the embedded bits are considered as noise to the original speech. The mathematical expression for calculating the SNR is given by Eq. 8

$$SNR(dB) = 10\log\left[\frac{\sum |s_o(n)|^2}{\sum |s_o(n) - s_w(n)|^2}\right] \tag{8}$$

$s_o(n)$ and $s_w(n)$ are original and watermarked speech signal respectively and $n = 1, 2, ..., N$, where N represents number of speech frames. The SNR values are computed by embedding the watermark bits in the complete database and the average value along with its comparison is listed in Table II.

The high SNR value resulted in proposed watermarking technique indicates that the embedding of watermark does not affect the quality of speech signal.

b) BER: The bit error rate calculates the bit error from retrieved watermark bits from the watermarked speech signal, after re-sampling, re-quantization, AWGN and MP3 compression attacks. In re-sampling attack the watermarked speech signal is sampled with a sampling frequency different from the original sampling frequency and re-sampled back to the original frequency. Similarly in re-quantization attack the watermarked speech is quantized to different level to corrupt the watermark.

In AWGN attack the white Gaussian noise is added to the watermarked speech signal and the error between retrieved watemark bit and original watermark bit is calculated. Similarly, in MP3 compression attack the watermarked speech is compressed by MP3 standard and de-compressed to corrupt the watermarked bits embedded in the speech. The BER values against the common signal processing attacks are listed in Table III with maximum payload of $6kbps$. The BER values were obtained for MP3 compression rate of $128kbps$, $96kbps$, $64kbps$ and $32kbps$. Comparison of our BER results with another frequency domain based DWT-SVD [2] watermarking technique are shown in the Table III. The proposed technique shows the reduction in BER for re-sampling & MP3 compression attacks by a factor of 10. It shows similar performance for re-quantization & AWGN attack when compared to DWT-SVD based watermarking technique presented in [2]. The BER is zero for 24 & 8 levels of re-qunatization attack, similarly it is zero for $15dB$ & $20dB$ of AWGN attacks.

B. Evaluation of classifier

The classification of watermarked PD affected speech from healthy person's speech is performed using k-NN and SVM classifiers to find the optimized classifier. The training and test dataset is created using PD speech database [22] and Saarbrucken voice disorder database (SbVD) [23]. The PD speech database consists of 168 speech samples recorded from 28 PD affected patients. Each patient was asked to speak sustained vowels "a" and "o" three times. Speech samples of 100 healthy people are taken from SbVD database containing sustained vowels "a" and "o".

The training and testing dataset is created by extracting 26 features from each speech sample as listed in Table IV.

TABLE II
COMPARISON SNR VALUES

Watermarking Techniques	DCT-SVD (Proposed technique)	DWT-SVD [10]	DWT-SVD [2]	DWT-FFT [24]	DWT [19]
SNR(dB)	87.41	54.12	37.50	34.45	61

TABLE III
BER RESULTS OF PROPOSED WATERMARKING TECHNIQUE SUBJECTED TO VARIOUS SIGNAL PROCESSING ATTACKS

Attacks	Resampling		AWGN		MP3 compression		
	6 KHz	4 KHz	15 dB	20 dB	96 $kbps$	64 $kbps$	32 $kbps$
Proposed Technique	0.001	0.003	0	0	0.0011	0.0013	0.0014
DWT-SVD [2]	0.09	–	0	0	0.0721	0.0820	0.2901

TABLE IV
VARIOUS FEATURES EXTRACTED FROM THE SPEECH SAMPLES FOR CLASSIFICATION

Parameters	Features Extracted
Frequency Parameters	Jitter (local), Jitter (local, absolute), Jitter (rap), Jitter (ppq5), Jitter (ddp)
Pulse Parameters	Number of pulses, periods, Mean period, standard deviation of period
Amplitude Parameters	Shimmer (local), Shimmer (local, dB), Shimmer (apq3), Shimmer (apq5), Shimmer (apq11), Shimmer (dda)
Voicing Parameters	Fraction of locally unvoiced frames, Number of voice breaks, Degree of voice breaks
Pitch Parameters	Median pitch, Mean pitch, Standard deviation, Minimum pitch, Maximum pitch
Harmonicity Parameters	Autocorrelation, Noise-to-Harmonic ratio, Harmonic-to-noise ratio

The accuracy in classification of watermarked PD affected speech from healthy person's speech using SVM classifier is listed in Table V. The classification accuracy is computed by:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

where, TP is true positive, TN is true negative, FP is false positive and FN is false negative.

TABLE V
ACCURACY(%) OF SVM CLASSIFIER IN CLASSIFICATION OF WATERMARKED PD AFFECTED SPEECH FROM HEALTHY SPEECH

Vowels	Accuracy
"a"	88.2%
"o"	89.5%

VII. CONCLUSION

In this article, we have presented a new watermarking technique to ensure the security of EPR in PD affected speech signal. The watermarking technique is highly imperceptible and hence maintains the spectral characteristics. Experimental results shows that the proposed framework is robust to common signal processing attack and provides better classification accuracy for diagnosing PD affected patients by their speech using smart phone device.

REFERENCES

[1] M. M. Baig, H. Gholamhosseini, Smart health monitoring systems: An overview of design and modeling, Journal of Medical Systems 37 (2) (2013) 9898.

[2] A.-H. Ali, An imperceptible and robust audio watermarking algorithm, EURASIP Journal on Audio, Speech, and Music Processing 2014 (1) (2014) 37.

[3] M. Shahbakhi, D. T. Far, E. Tahami, Speech analysis for diagnosis of parkinson‘s disease using genetic algorithm and support vector machine, Journal of Biomedical Science and Engineering 7 (4) (2014) 147–156.

[4] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, L. O. Ramig, Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease, IEEE Transactions on Biomedical Engineering 59 (5) (2012) 1264–1271.

[5] Y. Zhang, Can a smartphone diagnose parkinson disease? a deep neural network method and telediagnosis system implementation, Parkinson‘s Disease 2017.

[6] S. A. Parah, J. A. Sheikh, F. Ahad, N. A. Loan, G. M. Bhat, Information hiding in medical images: a robust medical image watermarking system for e-healthcare, Multimedia Tools and Applications 76 (8) (2017) 10599–10633.

[7] N. A. Loan, S. A. Parah, J. A. Sheikh, J. A. Akhoon, G. M. Bhat, Hiding electronic patient record (epr) in medical images: A high capacity and computationally efficient technique for e-healthcare applications, Journal of Biomedical Informatics 73 (2017) 125–136.

[8] S. A. Parah, J. A. Sheikh, F. Ahad, G. Bhat, High capacity and secure electronic patient record (epr) embedding in color images for iot driven healthcare systems, in: Internet of Things and Big Data Analytics Toward Next-Generation Intelligence, Springer, 2018, pp. 409–437.

[9] N. Dey, A. S. Ashour, S. Chakraborty, S. Banerjee, E. Gospodinova, M. Gospodinov, A. E. Hassanien, Watermarking in biomedical signal processing, in: Intelligent Techniques in Signal Processing for Multimedia Security, Springer, 2017, pp. 345–369.

[10] M. Alhussein, G. Muhammad, Watermarking of parkinson disease speech in cloud-based healthcare framework, International Journal of Distributed Sensor Networks 11 (10) (2015) 264575.

[11] Z. Ali, M. Imran, W. Abdul, M. Shoaib, An Innovative Algorithm for Privacy Protection in a Voice Disorder Detection System, Springer International Publishing, Cham, 2018, pp. 228–233.

[12] W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding, IBM Systems Journal 35 (3.4) (1996) 313–336.

[13] N. Cvejic, T. Seppanen, Increasing the capacity of lsb-based audio steganography, in: Multimedia Signal Processing, 2002 IEEE Workshop on, 2002, pp. 336–338.

[14] K. Bhowal, D. Bhattacharyya, A. Jyoti Pal, T.-H. Kim, A ga based audio steganography with enhanced security, Telecommun. Syst. 52 (4) (2013) 2197–220.

[15] A. Kanhe, G. Aghila, C. S. Kiran, C. Ramesh, G. Jadav, M. Raj, Robust audio steganography based on advanced encryption standards in temporal domain, in: Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on, 2015, pp. 1449–1453.

[16] Y. Erfani, S. Siahpoush, Robust audio watermarking using improved ts echo hiding, Digital Signal Processing 19 (5) (2009) 809–814.

[17] V. Korzhik, G. Morales-Luna, I. Fedyianin, Audio watermarking based on echo hiding with zero error probability, International Journal of Computer Science and Applications, Technomathematics Research Foundation 10 (1) (2013) 1–10.

[18] M. Fallahpour, D. Megias, Audio watermarking based on fibonacci numbers, IEEE/ACM Transactions on Audio, Speech, and Language Processing 23 (8) (2015) 1273–1282.

[19] M. Fallahpour, D. Megias, Robust audio watermarking based on fibonacci numbers, in: 2014 10th International Conference on Mobile Ad-hoc and Sensor Networks, 2014, pp. 343–349.

[20] S. Ahani, S. Ghaemmaghami, Z. J. Wang, A sparse representation-based wavelet domain speech steganography method, Audio, Speech, and Language Processing, IEEE/ACM Transactions on 23 (1) (2015) 80–91.

[21] P. Shah, P. Choudhari, S. Sivaraman, Adaptive wavelet packet based audio steganography using data history, in: 2008 IEEE Region 10 and the Third international Conference on Industrial and Information Systems, 2008, pp. 1–5.

[22] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, O. Kursun, Collection and analysis of a parkinson speech dataset with multiple types of sound recordings, IEEE Journal of Biomedical and Health Informatics 17 (4) (2013) 828–834.

[23] B. Woldert-Jokisz, Saarbruecken voice database, Institute of Phonetics, Saarland University.

[24] Rekik, S., Guerchi, D., Selouani, S.A., Hamam, H.: Speech steganography using wavelet and fourier transforms. EURASIP Journal on Audio, Speech, and Music Processing **2012**(1), 1–14 (2012). https://doi.org/10.1186/1687-4722-2012-20

[25] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, et al., Suitability of dysphonia measurements for telemonitoring of parkinson's disease, IEEE transactions on biomedical engineering 56 (4) (2009) 1015–1022.