IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 26, 2017, accepted January 26, 2018, date of publication January 30, 2018, date of current version March 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2799604

New Zero-Watermarking Algorithm Using Hurst Exponent for Protection of Privacy in Telemedicine

ZULFIQAR ALI[1], M. SHAMIM HOSSAIN[2], (Senior Member, IEEE), GHULAM MUHAMMAD[1], AND MUHAMMAD ASLAM[3]

[1]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[3]Department of Computer Science and Engineering, University of Engineering and Technology at Lahore, Lahore 54890 , Pakistan

Corresponding authors: M. Shamim Hossain (mshossain@ksu.edu.sa) and Ghulam Muhammad (ghulam@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, through the Research Group under Project RGP-1436-023.

ABSTRACT Telemedicine has numerous potential applications in the medical field due to the significant progress of telecommunication and information technology in recent years. In any category of telemedicine, such as offline, remote monitoring, and interactive, medical data and personal information of an individual must be transmitted to the healthcare center. An unauthorized access to the data and information is unacceptable in a telemedicine application, because it may create unavoidable circumstances for a person's private and professional life. To avoid any potential threat of identity exposure in telemedicine, a zero-watermarking algorithm to protect the privacy of an individual is proposed in this paper. The proposed algorithm embeds the identity of a person without introducing any distortion in medical speech signals. Two measures, namely, Hurst exponent and zero-crossing, are computed to determine the suitable locations in the signal for insertion of identity. An analysis of the signals indicates that unvoiced speech frames are reliable in insertion and extraction of identity, as well as robust against a noise attack. In the proposed zero-watermarking algorithm, identity is inserted in a secret key instead of a signal by using a 1-D local binary operator. Therefore, imperceptibility is naturally achieved. Experiments are performed by using a publicly available voice disorder database, and experimental results are satisfactory and show that the proposed algorithm can be reliably used in telemedicine applications.

INDEX TERMS Vocal fold disorders, speech signals, Hurst exponent, zero-crossing, local binary operator.

I. INTRODUCTION

With the recent development in telecommunication and information technology, the application of telemedicine has been proposed in various areas in the medical field [1]–[3]. Telemedicine applications provide clinical healthcare from a distance, which benefits patients in remote regions and isolated communities. Some of the substantial benefits of telemedicine applications are frequent follow-ups and fewer unwanted visits to consult doctors or specialists [4].

Generally, telemedicine is classified into three main categories [5]. The first one is store and forward. This category involves obtaining medical data, such as medical speech signals and images and their transmission over wireless communication channels to the specialists for diagnosis at a convenient time. For this offline diagnosis, the presence of both parties concurrently is unnecessary. The second category is remote monitoring, which is also called self-monitoring. This type of telemedicine allows health professionals to monitor the patients remotely through sensors and monitoring devices. The third category is interactive telemedicine where patients and healthcare staff have real-time conversation for the diagnosis, counseling, and monitoring (e.g., through videoconferencing).

In all forms of telemedicine, the potential threat is the disclosure of patients' information and unauthorized access to medical data. Unauthorized access to patients' information may affect a person's private and professional life. Thus, telemedicine applications must be able to protect medical data and should not reveal patients' information, including name, age, telephone number, address, and so on.

The privacy of patients may be vulnerable when proper security is not guaranteed [6]. The privacy of a patients' information can be protected by using various approaches. A comprehensive survey on security and privacy issues in radio frequency identification systems and their solutions was presented in [7]. A comparison of three different watermarking algorithms based on discrete wavelet transformation (DWT), singular-value decomposition (SVD), and DWT-SVD was presented in [8] to protect the privacy of patients in terms of medical images. Several studies have presented watermarking techniques for medical images to protect patients' privacy [9]–[11]. Moreover, Vellaisamy and Ramesh [12] described the inversion attack resilient zero-watermarking scheme for medical image authentication.

The current study aims to protect patients' information by developing a new zero-watermarking algorithm for medical speech signals. No literature is available on watermarking for privacy protection in medical speech signals. Most of the works have been conducted for the privacy protection in medical images. However, a cloud-based healthcare framework [44] was proposed by Alhussein and Muhammad [13]. The designed framework depends on the watermarking in speech signals of patients with Parkinson's disease [14]. An image containing a patient's identity is inserted in the speech signal to protect the identity. An algorithm that combines DWT and SVD is utilized to embed and recover the identity. The insertion of the patient's identity in speech signal may change its characteristics. Therefore, the evaluation of the watermarking algorithm for imperceptibility is important. The authors mentioned that the archived imperceptibly is approximately 4.68 on a scale of one to five. Despite good imperceptibility, insertion of identity distorts the signals, which may ultimately lead to false diagnosis due to change in the speech signal.

The distortion in the signal can be avoided by using zero-watermarking instead of traditional watermarking. The zero-watermarking algorithm embeds the patient's information in the generated secret key instead of the signal. In such algorithms, the imperceptibility is ideal and the signal before and after the insertion of the watermark remains unaltered. Therefore, a zero-watermarking algorithm for medical speech signals was proposed by Ali *et al.* [15] to protect the identity of patients. Privacy protection is two-fold because identity is not embedded directly in the secret keys. Two secret shares of the identity are created by applying visual cryptography [16], which are inserted into the secret keys. Recovering both secret shares is needed to disclose the identity of a patient. The only drawback of the study is the size of the watermark because it increased by many folds due to visual cryptography.

The medical speech signals used in the existing study were recorded from patients with vocal fold disorders. During voice production, the vocal folds come closer to each other and the air pressure generated by the lungs makes the vocal folds vibrate. The produced voice then travels through a person's mouth to produce a sound [17]. The vocal folds reside on top of the trachea (windpipe) and are made of small folds of muscles and tissues. Vocal folds open and close at regular intervals during phonation to produce a healthy voice. Vibration of the vocal folds becomes irregular during voice production due to vocal misuse. Consequently, the pattern in the waveform of the signal becomes transient and complex and makes the voice breathier, weaker, strained, and harsh [18]. These changes in the voice can be regarded as vocal fold disorder or dysphonia. Dysphonia can be defined as an alteration in performance and production of the voice that may interfere in communication [19]. According to the medical dictionary [20], dysphonia is difficulty in speaking, usually evidenced by hoarseness.

Many vocal fold disorder assessment systems for detection, as well as classification of various types of vocal fold disorders, exist in the literature [21]–[27], [42], [43]. Most of the systems were developed by using speech samples containing sustained vowels. In comparison with sustained-vowel-based assessment systems, only few systems have been developed for the detection and classification of vocal fold disorders using running speech [28], [29]. The development of running-speech-based assessment methods is difficult due to the presence of unvoiced speech segment, pauses, variations in fundamental frequencies, and rapid voice onset and offset terminations [30]. However, running-speech-based vocal fold disorder detection and classification systems are more natural than the systems based on sustained vowels. The characteristics of running speech, such as voice breaks and onset–offset information, can play important roles in accurate diagnosis. Such characteristics are missing in sustained vowels [31].

Therefore, in this study, signals containing running speech are used in the proposed zero-watermarking algorithm to protect the privacy of patients. The identity of an individual is embedded in the generated secret key instead of the speech signal. Consequently, the speech signal will not be degraded, and the imperceptibility will be achieved naturally. However, the generation of the secret key depends on the characteristics of the speech signal. Determining the appropriate regions in the speech signals where the characteristics remain stable and do not change in case of malicious attack is also important. In this study, Hurst exponent and zero-crossing are used to find the suitable regions in a speech signal. The characteristics of the speech signals are computed by applying one-dimensional local binary operator based on signal amplitude. The proposed zero-watermarking algorithm is evaluated by performing various experiments. The experimental results show that the proposed zero-watermarking algorithm is reliable in inserting and extracting identity. In addition, the proposed algorithm is robust against noise attack.

The rest of the paper is organized as follows. Section II describes the various components of the proposed zero-watermarking algorithm. This section also explains the insertion and extraction processes of the proposed algorithm. Section III provides the evaluation of the proposed algorithm

through various experiments. Finally, Section IV draws some conclusions.

II. PROPOSED ZERO-WATERMARKING ALGORITHM AND ITS COMPONENTS

Two measures, namely, Hurst exponent and zero-crossing, are calculated for the divided frames of the signal to determine the appropriate locations for the insertion of identity. Moreover, the characteristics of the suitable frames are computed through the one-dimensional local binary operator. The insertion and extraction processes of the proposed zero-watermarking algorithm are also discussed in this section.

A. HURST EXPONENT FOR SPEECH SIGNAL

The Hurst exponent is one of the reliable tests for detecting long memory [32] and fractal analysis of time series [33]. This test was proposed in 1951 by English hydrologist [34]. The Hurst exponent can be calculated by using the rescaled range (R/S) analysis, and it has been found better than autocorrelation, variance analysis, and spectral analysis [35].

In this study, Hurst exponent is calculated by using R/S analysis to determine the appropriate locations for watermark insertion. To computer Hurst exponent for a speech signal U containing the running speech, the signal is divided into shorter frames of length L. Analyzing the speech signal is important because speech, especially running speech, varies quickly over time. The segments of speech in the obtained frames remain stationary, and they become easy to analyze. The length L of each frame is determined by Eq. (1), such that the duration of frames is less than 30 ms, and it contains number of samples equal to a power of 2.

$$L = 2^{\lfloor \log_2(0.03*fs) \rfloor}, \tag{1}$$

where fs stands for the sampling frequency of the speech signal.

The first step in computation of the Hurst exponent for the ith frame F^i of length L is the calculation of mean μ, which is obtained by using Eq. (2).

$$\mu = \frac{1}{L}\sum_{j=1}^{L} f_j^i \quad \text{where } F^i = [f_1^i, f_2^i, f_3^i, \ldots, f_L^i] \tag{2}$$

In Eq. (2), f_j^i represents the samples of F^i. The obtained mean μ is subtracted from F^i for normalizing it, as shown as follows:

$$E_k^i = F_k^i - \mu \quad \text{for } k = 1, 2, 3, \ldots, K, \tag{3}$$

where F_k^i is represents the subframes of F^i, and K can be obtained as

$$K = \frac{L}{2^0}, \frac{L}{2^1}, \frac{L}{2^2}, \ldots, 2^4. \tag{4}$$

In this way, a normalized subframe for all values of K will be computed. Cumulative sum C_k^i for each subframe E_k^i is then computed as follows:

$$C_K^i = \sum_{k=1}^{K} E_k^i. \tag{5}$$

A sequence R containing range of cumulative sum of series for all subframes is obtained by using Eq. (6).

$$R_K^i = \max\left(C_1^i, C_2^i, C_3^i, \ldots, C_K^i\right) - \min\left(C_1^i, C_2^i, C_3^i, \ldots, C_K^i\right) \tag{6}$$

The sequence R_K^i is divided by the standard deviation S_K^i to rescale the range, and it can be calculated as:

$$S_K^i = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(f_k^i - m)} \quad \text{where } m = \frac{1}{K}\sum_{k=1}^{K} f_k^i \tag{7}$$

Then, the obtained R/S values for the subframe K of F^i is shown as follows:

$$\left(\frac{R^i}{S^i}\right)_K = \frac{R_K^i}{S_K^i}. \tag{8}$$

The R/S values are averaged over the subframe. Similarly, R/S values for all subframes for all values of K given in Eq. (4) are computed, and a graph is plotted for $log_2(R/S)$ versus $log_2(K)$. The slope of the plotted line provides the estimation for Hurst exponent.

The value of Hurst exponent close to 0.5 describes a total random walk. The value of H between 0.5 and 1.0 indicates a persistent behavior; hence, if the waveform is in an increasing trend, then it will continue in the same way; otherwise, if the waveform is in a decreasing trend, it will continue in the same trend. Moreover, the value of Hurst component between 0 and 0.5 indicates that the behavior of the waveform is anti-persistent, that is, the increasing trend of the waveform converts into a decreasing trend or vice-versa.

In Fig. 1, a speech signal (AOS21R.wav) affected by the vocal fold disorder is divided into shorter frames. The various frames of the signal denote different behaviors (Fig. 1). Hurst exponents of these frames are also computed and shown in Fig. 1. The appropriate frame for the insertion of identity of an individual can be determined by applying some threshold on the computed Hurst exponents.

B. ZERO-CROSSING IN A SPEECH SIGNAL

The second measure to determine the appropriate frames for the insertion of watermark is a zero-crossing. The zero-crossing is a point where the speech signal crosses the x-axis. In other words, zero-crossing represents the location where the speech signal changes the sign from negative to positive or vice versa. A signal experiences high zero-crossing in case of unvoiced and silent frames. However, unvoiced frames contain higher amplitude as compared to silent frames. Moreover, the voiced frames have high amplitudes and low zero-crossing rate. A threshold can be adjusted on zero-crossing to determine the appropriate frame

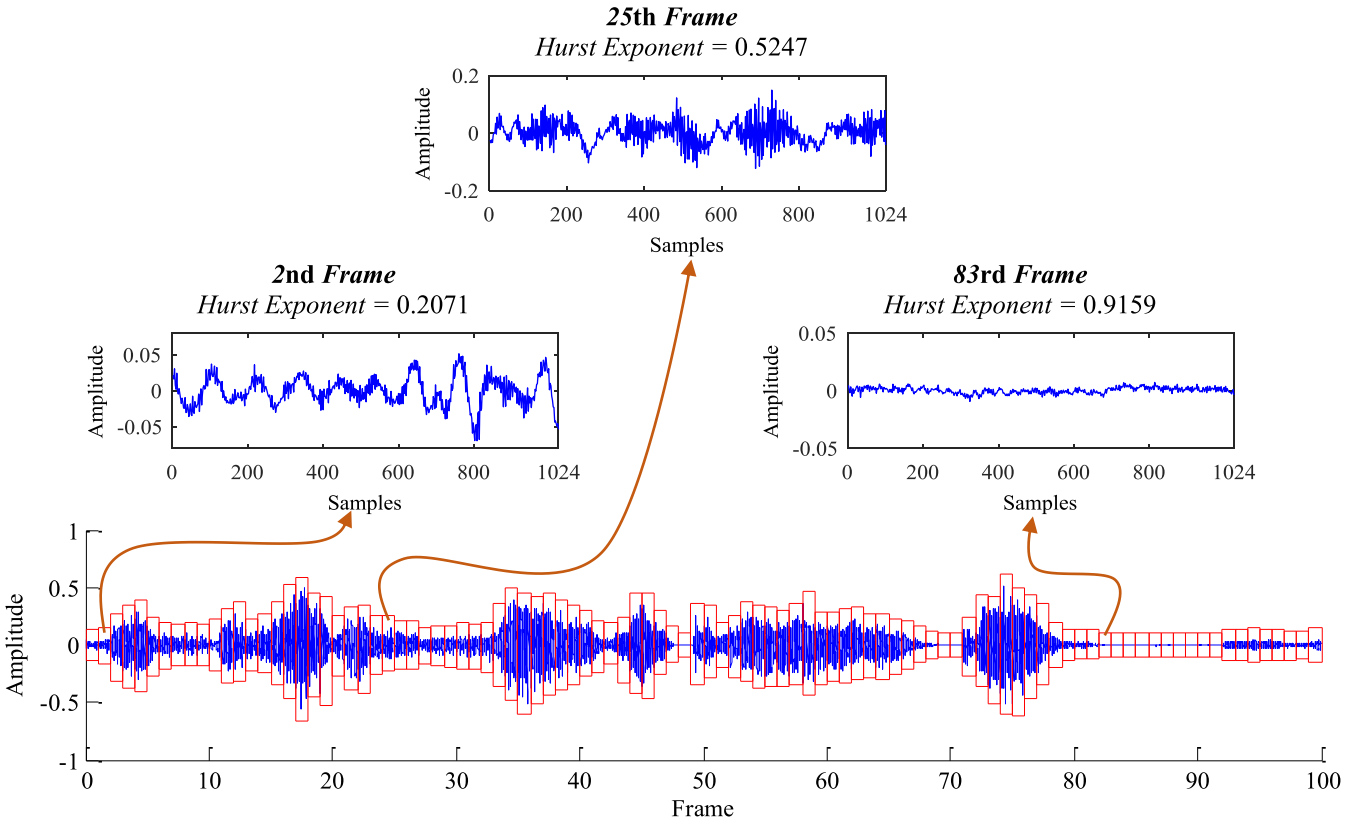

FIGURE 1. Different frames of speech signal (AOS21R) affected by vocal fold disorder and their computed Hurst exponents.

for watermarking [36]. The selected frames will be further analyzed to find some patterns in them.

C. ONE-DIMENSIONAL LOCAL BINARY PATTERN BASED ON THE AMPLITUDE OF THE SIGNAL

The next important step in the proposed zero-watermarking algorithm is the calculation of the characteristics in the suitable speech frames determined by applying a threshold on Hurst exponent. The insertion of the watermark strongly depends on the computed characteristics of the signal.

In this study, the characteristics in each frame are calculated by using a one-dimensional local binary operator [37], [38]. The operator computes a code for every sample of the selected frame assuming amplitudes of samples. The obtained one-dimensional code is represented by LBP^{M}. The LBP^{M} for a segmented window w of the frame F^i with a center element f_c and neighbors $f_\phi, \phi = 1, 2, 3, 4$ are computed by using the relation given by Eq. (9). Elements f_1 and f_2 are on the right side of the center element f_c, whereas, f_3 and f_4 are on the left side of f_c, such as $w = [f_4, f_3, f_c, f_2, f_1]$. Moreover, M represents the average value of the neighboring elements.

$$d_{\phi-1} = \begin{cases} 1 & \text{if } f_c \geq M \\ 0 & \text{if } f_c < M \end{cases} \quad \text{where } M = \frac{1}{4}\sum_{\phi=1}^{4} f_\phi \quad \text{and } \phi = 1, 2, 3, 4 \tag{9}$$

From Eq. (9), a pattern of zeroes and ones $d_3 d_2 d_1 d_0$, is obtained, where d_0 and d_3 are the least and most significant bits, respectively. This pattern is a required LBP^{M} code for the window $w = [f_4, f_3, f_c, f_2, f_1]$, and the frequency of this code in the entire frame is represented by the corresponding bin in a histogram. For example, the frequency of the LBP^{M} code 1010 is represented by 10th bin of the histogram where 10 is the corresponding decimal number of the 4-bit binary number 1010.

In the case of four neighbors, the total number of LBP^{M} codes is 2^n in the range from 0 to $2^n - 1$, and each bin of the histogram shows the frequency of an LBP^{M} code.

D. DATASET

The speech signals containing running speech are taken from the voice disorder database recorded at the Massachusetts Eye and Ear Infirmary (MEEI) voice and speech laboratory [39]. The database is available through PENAX Medical, New Jersey, USA. The database contains speech signals of normal persons, as well as patients with disorders. A subset of the MEEI database containing 173 disordered and 53 normal speech signals are considered in this study. This subset has been used for many studies [24], [28], [40]. The distribution of the normal and disordered speech signals in the selected subset was provided in [41]. The subset contains various types of voice disorders. In addition, age and gender of the subjects are evenly distributed. All samples of the MEEI subset are

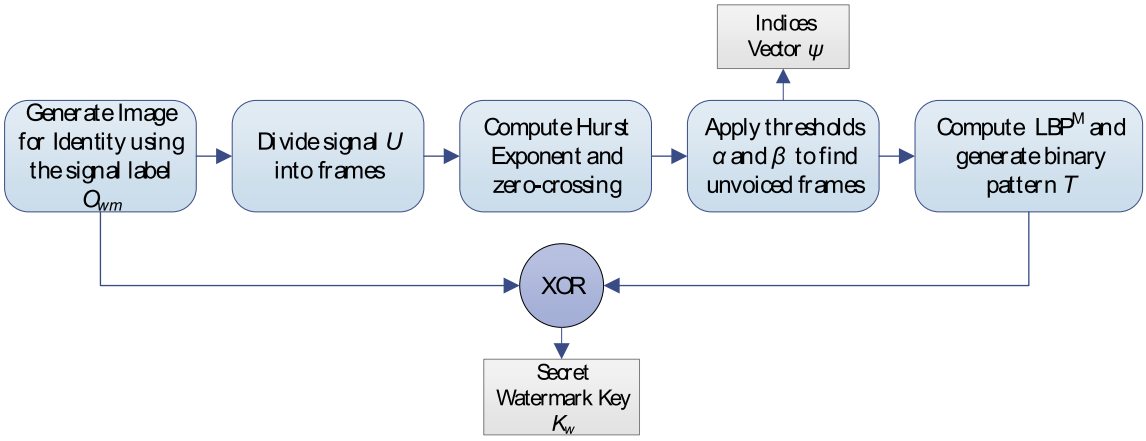

FIGURE 2. Process for insertion of the identity.

downsampled to 25 KHz before the insertion and extraction of the identity.

The labels of the normal and disordered speech signals in the MEEI database are used to generate the black-and-white image for identity of an individual. The dimension of each image is 20 x 126, and it is used as a watermark in the proposed the proposed zero-watermarking algorithm.

E. EMBEDDING PROCESS OF THE PROPOSED ALGORITHM

The block diagram for the watermark insertion is shown in Fig. 2. The watermark, which is the identity of a patient with dysphonia, is embedded as follows.

1. A black-and-white image O_{wm} with dimension $r_o \times c_o$ is created; this image contains the identity of a person or patient. Image O_{wm} is the original watermark that must be inserted to protect the privacy.
2. The host speech signal U is divided into non-overlapping frames of length L, as shown as follows:

$$U = [F^1, F^2, F^3, \ldots, F^i, \ldots, F^N], \tag{10}$$

where N represents the total number of frames in signal U.

3. Hurst exponent and zero-crossing are computed for each frame of the host signal U to determine the suitable locations for the insertion of the watermark O_{wm}.
4. Two thresholds, namely, α and β, are adjusted for the computed Hurst exponent and zero-crossing, respectively, to identify the appropriate frames. In the threshold $\alpha = [\alpha_1, \alpha_2]$, α_1 and α_2 are two different non-negative real numbers between 0 and 1, inclusively, whereas the threshold β is a positive integer.
5. The indices of the suitable frame of signal U are determined by using the criteria given in Eq. (11).

$$\psi = \left\{\gamma \,\middle|\, H(F^\gamma) \in [\alpha_1, \alpha_2] \quad \text{and } Z(F^\gamma) \geq \beta \right\} \quad \text{where } \gamma = 1, 2, 3, \ldots, N \tag{11}$$

$H(F^\gamma)$ *and* $Z(F^\gamma)$ represent the Hurst exponent and zero-crossing of a frame, respectively; and ψ contains indices of all suitable frames.

6. Binary pattern T is generated by using LBP^{M} by considering the criteria given in Eq. (12). Pattern T contains a zero if LBP^{M} of a segmented window w of a frame has a zero or one transition of 0-to-1 or 1-to-0. For example, LBP code of a window 0011 has one transition of 0-to-1 or 1-to-0. Likewise, the code 1111 has zero transitions. Pattern T contains a one if an LBP^{M} of a window w has two or more transitions of 0-to-1 or 1-to-0. For example, codes 1010 and 1001 have three and two transitions, respectively.

$$T(\xi) = \begin{cases} 0 & \textit{if trans}\left(LBP^M(w_\theta^\psi)\right) \in \{0, 1\} \\ 1 & \textit{if trans}\left(LBP^M(w_\theta^\psi)\right) \in \{2, 3\} \end{cases}$$

where

$$\theta = 3, 4, 5, \ldots, L-2$$
$$\xi = 1, 2, 3, \ldots, (r_o \times c_o) \tag{12}$$

In Eq. (12), *trans* denotes the number of transition 0-to-1 and 1-to-0. In addition, the number of windows for a frame, represented by θ, are L-4 because zero padding is not performed.

7. Finally, the watermark key K_w is generated by performing exclusive-OR (XOR) operation between the watermark O_{wm} and the pattern T, as shown by Eq. (13).

$$K_w = T \oplus O_{wm} \tag{13}$$

The generated key K_w with signal U band the vector of indices ψ of the appropriate frames must be transmitted to the authorized staff through secure channel for use in disclosing the identity of the individual.

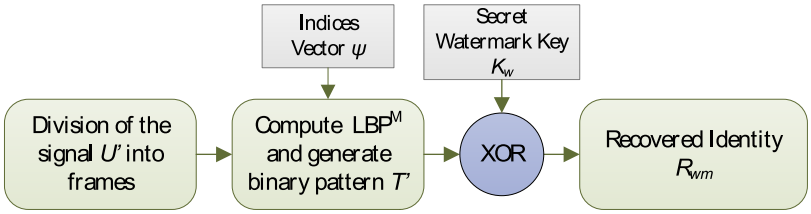

FIGURE 3. Process for extraction of the identity.

F. EXTRACTION PROCESS OF THE PROPOSED ALGORITHM

The authorized healthcare staff can recover the identity of an individual as follows:

1. Divide the watermarked audio U' into N non-overlapping frames of length L.
2. Identify the suitable frames by using the index vector ψ and computing the binary pattern T' by using the LBP^{M} codes of the segmented windows w' through Eq. (12).
3. Perform exclusive-OR (XOR) operation between the received watermark key K_w and the computed pattern T' to recover the identity of an individual, by using Eq. (14).

$$R_{wm} = T' \oplus K_w. \tag{14}$$

The block diagram explaining the extraction process is depicted in Fig. 3.

III. EVALUATION OF THE PROPOSED ALGORITHM

Various experiments are conducted to observe the insertion and extraction reliability of the proposed zero-watermarking algorithm.

FIGURE 4. Identity of an individual (image with dimension of 20 × 126).

A. QUALITY OF WATERMARKING

In privacy-protected healthcare systems, the identity of an individual should be inserted reliably, such that no one can reveal it without having the access to the relevant secret key. The privacy of the identity primarily depends on how the identity is inserted. Therefore, evaluating the proposed algorithm is important to determine the quality of the inserted identity, which can be observed by finding the difference between the original identity O_{wm} and the secret key K_w (which contains the identity). The identity of an individual is shown in Fig. 4.

In the proposed algorithm, the watermark is inserted by determining the Hurst exponent and zero-crossing to ensure that a frame is suitable for watermarking. Voiced (periodic) frames of a speech signal are unsuitable for the insertion of identity, which is performed by the inspection of different frames. Fig. 5(a) shows that the identity is not properly hidden. The identity, which is an image with dimension of 20 × 126 (2520 pixels), took three (two complete and the third partial) frames of length to hide 1024 samples. The last two frames shown in Figs. 5(c) and 5(d), which contain the second half of the identity (approximately 1496 pixels), are voiced frames, and they fail to hide the identity. Moreover, the first half of the identity is inserted in the unvoiced frame shown in Fig. 5(b), and that part of the identity is completely disguised. This frame does not exhibit any periodicity at all, and it provides the best suitable locations for the watermarking.

The unvoiced frames are determined automatically by applying the threshold on the computed Hurst exponent and zero-crossing of the frames. For unvoiced frames, the Hurst exponent should lie within the interval [0.3, 0.4], and the zero-crossing must be greater than 300. These two conditions should be satisfied simultaneously to find the unvoiced frames. It can be seen in the Fig. 1 that the speech signal contains many unvoiced frames such as frame 28 to frame 33. These frames provide sufficient space for the insertion of identity.

The identity of an individual inserted in three unvoiced frames is shown in Fig. 6(a). The identity was hidden completely, and no guess can be made about an individual's information. The three unvoiced frames are shown in Figs. 6(b)–6(c), respectively; the Hurst exponents for these frames are 0.3277, 0.3135, and 0.4000, whereas the zero-crossing for these unvoiced frames are 473, 393, and 388, respectively.

In comparison with the unvoiced frames, the voiced frame in Figs. 4(c)–4(d) have Hurst exponents in the interval [0.3, 0.4]; however, the zero crossing, which is 52 and 30, respectively, is low. Therefore, a frame should comply with both conditions to be an unvoiced frame.

To determine the quantitative difference between the original watermark and the secret key containing the watermark K_w, the objective analysis is conducted by using two performance measures, which are defined in Eqs. (15) and (16). The first relation describes the peak signal-to-noise ratio (PSNR). In Eq. (15), I_1 and I_2 are two images representing O_{wm} and K_w. B denotes the number of bits per pixel in O_{wm} and K_w. MSE stands for mean square error between O_{wm} and K_{wr} and MSE is the squared norm of the difference divided by the number of elements in the images.

$$\text{PSNR}\,(I_1, I_2) = 20\log_{10}\left(\frac{2^B - 1}{MSE}\right) \tag{15}$$

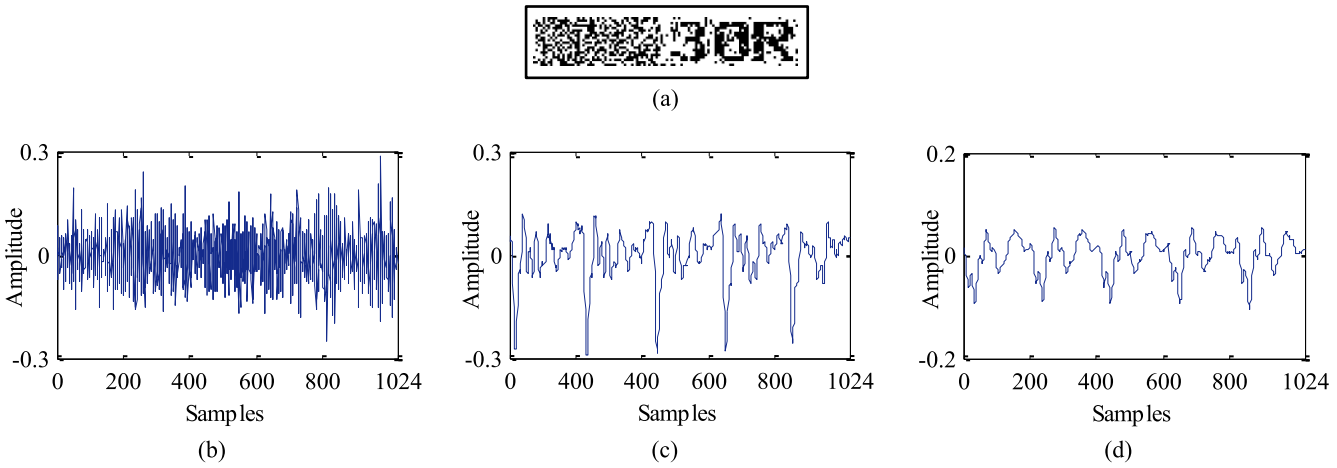

FIGURE 5. (a) Identity inserted into voiced and unvoiced frames. (b) Unvoiced frame with Hurst exponent = 0.3277 and zero-crossing = 473. (c) Voiced frame with Hurst exponent = 0.3436 and zero-crossing = 52 (d). Voiced frame with Hurst exponent = 0.3034 and zero-crossing = 30.

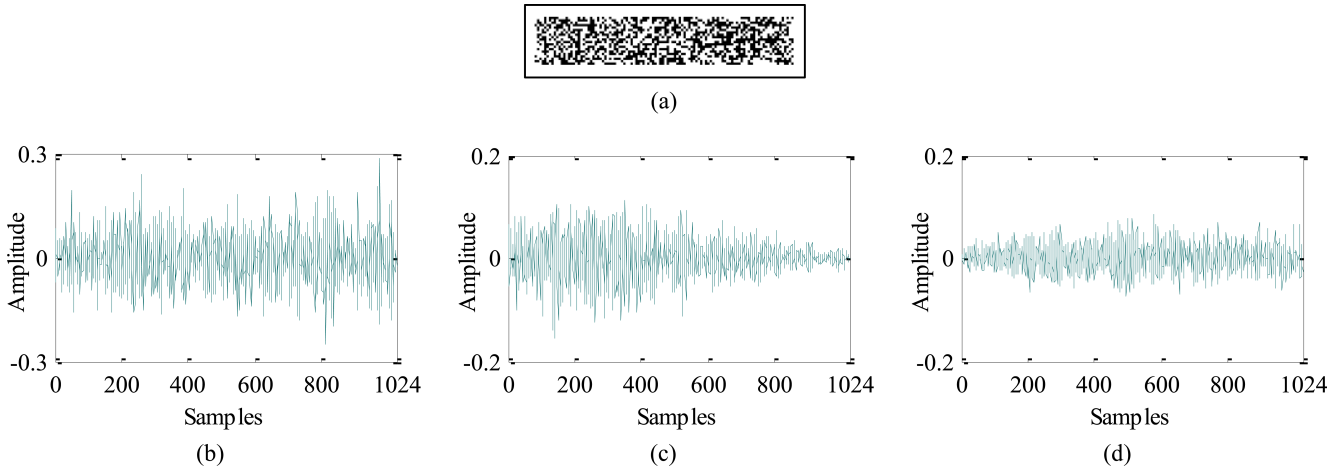

FIGURE 6. (a) Identity inserted into unvoiced frames only. (b) Unvoiced frame with Hurst exponent = 0.3277 and zero-crossing = 473 (c) Unvoiced frame with Hurst exponent = 0.3135 and zero-crossing = 393. (d) Unvoiced frame with Hurst exponent = 0.4000 and zero-crossing = 388.

The second measure is the bit error rate B_{rate}, which determines the difference between the corresponding bits of O_{wm} and K_w. It describes how different the two images are. B_{rate} is defined as

$$B_{rate} = \frac{NDB}{r_0 \times c_0}, \tag{16}$$

where NDB represents the number of erroneous bits and $r_0 \times c_0$ denotes the total number of bits.

Both measures, PSNR and B_{rate}, are computed for O_{wm} and K_w and provided in Table 1. Table 1 shows that when the watermark is inserted in the voiced and unvoiced frames, the PSNR and B_{rate} are 48.09 and 35.65, respectively. The PSNR of 48.09 is high and indicates no significant difference is observed between O_{wm} and K_w, which is unacceptable. Similarly, low B_{rate} means that the O_{wm} and K_w are similar to each other. Meanwhile, when the identity is inserted in unvoiced frames only, the PSNR and B_{rate} for O_{wm} and K_w are 10.17 and 60.21, respectively. The low PSNR and high B_{rate} show that O_{wm} and K_w are significantly different from each other.

TABLE 1. Quantitative analysis of O_{wm} and K_w.

Identities		Performance Measures	
		PSNR	B_{rate}
O_{wm}	K_w (voiced and unvoiced frames)	48.09	35.65
O_{wm}	K_w (voiced frames only)	10.17	60.21

The quantitative analysis strengthens the fact that the identity does not reveal any information about an individual when it is inserted in the unvoiced frame. Therefore, the identity of an individual is inserted in the unvoiced frames in the

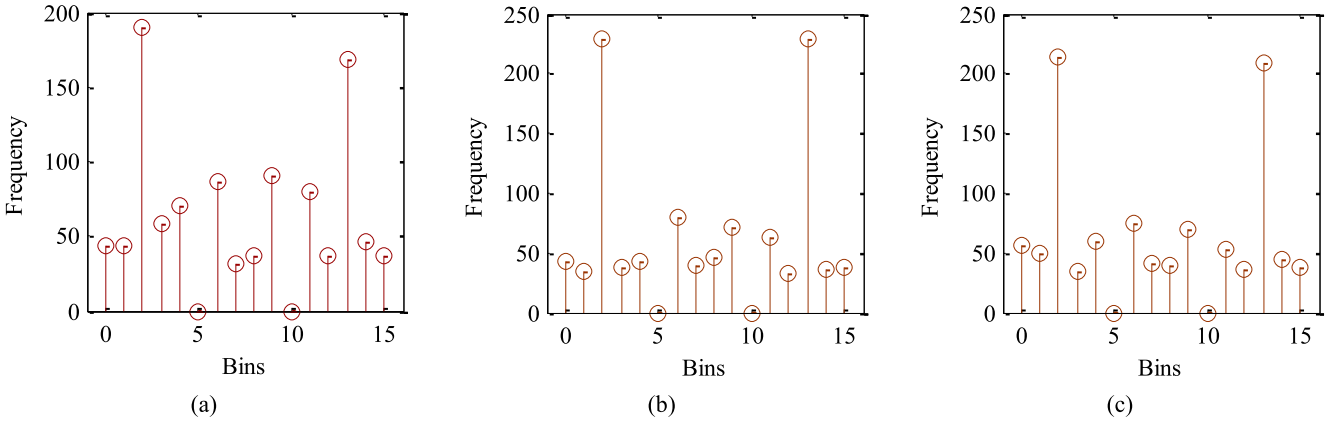

FIGURE 7. Corresponding histograms for the unvoiced frames shown in Figs. 6(b)–6(d).

proposed zero-watermarking algorithm. The PSNR and B_{rate} for O_{wm} and K_w (for unvoiced frame) are computed for all speech samples in the subset of the MEEI database. The average PSNR and B_{rate} for these samples, which are 13.51 and 55.95, respectively, are good and indicate that the proposed algorithm is reliable in hiding the identity.

The next important step in the evaluation of the proposed zero-watermarking algorithm is to observe the imperceptibility of the speech signal after watermark insertion.

B. IMPERCEPTIBILITY OF SPEECH SIGNAL

One of the benefits of zero-watermarking over traditional watermarking is the insertion of the identity in the secret key instead of the speech signal. In this way, the speech signal remains the same and exhibits no irregularities compared with the original signal. The privacy protection in the medical signal requires that the characteristic of the signal should be unaltered after the insertion of the watermark, otherwise, the speech signal may produce a false diagnosis, which is unacceptable. A false diagnosis may create discomfort and anxiety for individuals. For example, if a person is affected by vocal fold disorders, and the disorder detection system does not diagnose this correctly because of alteration in the signal after watermarking, then the person will not visit a doctor after this false diagnosis, and the disorder will become more severe. In addition, if a person is healthy and the system detects the vocal fold disorder, then the person will not only be in mental stress but will also waste time and money on hospital visit and consultation with a doctor.

This study seeks to avoid such circumstances by protecting the privacy of an individual through zero-watermarking where the speech signal will not exhibit any change and a detection system can correctly detect the vocal fold disorder. Moreover, the watermark is inserted in the secret key instead of the speech signal, thereby achieving imperceptibility naturally. The relation given in Eqs. (15)–(16) can also be used for the objective analysis of the imperceptibility. The only difference is the one-dimensional signals instead of identity images.

The original signal and the signal after the insertion of the identity are the same. Hence, an infinite PSNR is achieved because in case of similar signals, the MSE is zero and the division with a zero in Eq. (15) results in infinity. Moreover, the B_{rate} for the signals is zero because no difference is observed in the corresponding bits of the signals. In addition, the proposed algorithm is analyzed for the reliable recovery of the identity.

C. DETECTION RELIABILITY OF WATERMARK

The identity of an individual is inserted in the secret key, which is determined by performing the XOR operation between the pattern T and the identity image O_{wm}. The pattern T depends on the one-dimensional LBP^M codes, which are computed by comparing the magnitude of the neighboring elements with the center element in a segmented window w of an unvoiced frame. The window w contains four elements, and 16 LBP^M codes in the range of 0 to 15 are obtained. The obtained codes are shown in the histograms plotted in Fig. 7. These histograms are attained from the three unvoiced frames depicted in Figs. 6(b)–6(d). Each bin of the histograms represents the frequency of the respective LBP^M code.

The detection reliability of the watermark indicates that the watermark should be recovered accurately by using the transmitted secret key K_w. During the watermark extraction process, when healthcare staff computes the LBP^M codes by using the transmitted signal and secret key, the staff should obtain the same histograms as those computed during the embedding process because the characteristics of the signal are unchanged given that nothing is inserted in it. Therefore, the recovered identity R_{wm} of an individual will be exactly similar to the inserted identity O_{wm}. Moreover, the PSNR and B_{rate} between O_{wm} and R_{wm} are infinity and zero, respectively. This result suggests that the proposed zero-watermarking algorithm can recover the identity of an individual reliably, as performance parameters indicate no difference between the original and the recovered identities.

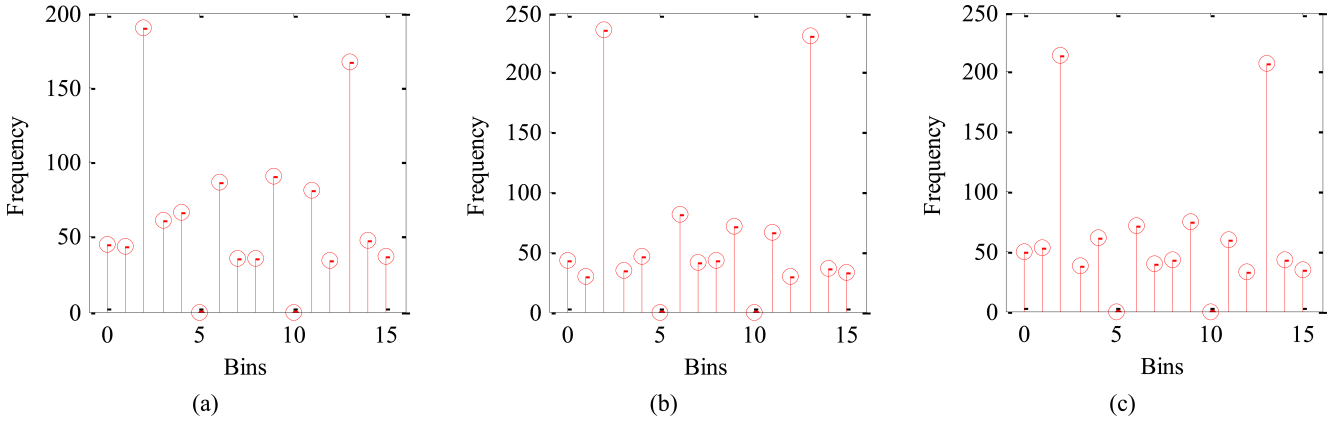

FIGURE 8. Corresponding histograms for the unvoiced frames shown in Figs. 6(b)–6(d) after the attack of 20 dB white Gaussian noise.

D. SURVIVABILITY AGAINST NOISE ATTACK

A watermarking algorithm must be able to recover the identity of an individual in case of a malicious attack. White Gaussian noise of different signal-to-noise (SNR) is added in the watermarked signals to observe the robustness of the proposed zero-watermarking algorithm against an attack.

The embedding and recovery of the individual's identity strongly rely on the computed LBP^{M} codes. Therefore, observing the robustness of the proposed algorithm is important to determine how much variation appears in the frequency of the LBP^{M} codes. For this purpose, the histograms for the unvoiced frame shown in Figs. 6(b)–6(d) are computed after an attack of 20 dB. The histograms for the attacked frames are depicted in Fig. 8. The histograms before and after the attack, shown in Figs. 7 and 8, respectively, did not exhibit any substantial difference.

The binary pattern T' during the recovering process will be similar to the pattern T obtained during the embedding process because of the ample similarity between the histograms. The identity recovered from the attacked signal will be closer to the original identity despite the significant attack of 20 dB. Therefore, an attack of high SNR does not distract the recovery process of the proposed zero-watermarking.

The objective analysis in the case of noise attack is also performed by using PSNR and B_{rate}. The computed PSNR and B_{rate} for the original identity O_{wm} and the recorded identity R_{wm} from the attacked signal are listed in Table 2. The recovered identity R_{wm} from the speech signal, which is attacked with the noise of 30 dB, is shown in Fig. 9(a). The B_{rate} for this recovered identity is 2.36%, which indicates that the identity is recovered from the attacked signal successfully. In addition, the high PSNR of 83.33 means that the recovered identity R_{wm} is considerably close to the original identity O_{wm}.

Furthermore, the SNR is increased to 20 dB to experience the robustness of the proposed algorithm in case of high SNR. The recovered identity from attacked signal of 20 dB is shown in Fig. 9(b). The recovered identity is clearly readable and has an extremely small B_{rate} of 7%. The PSNR of 76.53 dB also confirms that the proposed zero-watermarking algorithm can recover the identity reliably.

TABLE 2. Quantitative analysis of O_{wm} and R_{wm} after an attack.

Noise	Performance Measures	
	PSNR	B_{rate}
30 dB	83.33	2.36
20 dB	76.53	7.18

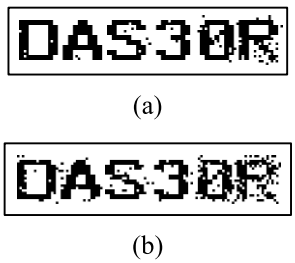

FIGURE 9. Recovered identities from the attacked speech signal with noise of (a) 30 dB and (b) 20 dB.

IV. CONCLUSIONS

A new zero-watermarking algorithm for privacy protection in telemedicine is proposed in this study. The proposed algorithm is based on the Hurst exponent and zero-crossing of the frames. Both measures are computed to determine the unvoiced frames for watermark insertion. The proposed algorithm can be used in telemedicine application successfully. In addition, the robustness of the algorithm against noise ensures that the identity cannot be removed by malicious attacks. The experimental results show that the algorithm provides high bit rate and low PSNR for the original and watermarked identity. Therefore, the algorithm will not reveal the identity of a person in telemedicine applications because of ideal imperceptibility and reliable insertion and extraction

of the identity. Moreover, the experiments for the recovery of the identity obtained extremely low bit rate and high PSNR. The results inferred that the proposed algorithm can reliably extract identities, which is vital for all forms of telemedicine. In future work, multiple secret shares of the identity can be embedded in the speech signal to enhance the security of the identity.

REFERENCES

[1] P. T. Kim and R. A. Falcone, Jr., "The use of telemedicine in the care of the pediatric trauma patient," *Seminars Pediatric Surgery*, vol. 26, pp. 47–53, Feb. 2017.

[2] M. Aguas, J. D. Hoyo, R. Faubel, B. Valdivieso, and P. Nos, "Telemedicine in the treatment of patients with inflammatory bowel disease," *Gastroenterol. Hepatol.*, vol. 40, no. 9, pp. 641–647, 2017.

[3] T. Elliott, J. Shih, C. Dinakar, J. Portnoy, and S. Fineman, "American College of Allergy, Asthma & Immunology position paper on the use of telemedicine for allergists," *Ann. Allergy, Asthma Immunol.*, vol. 199, no. 6, pp. 512–517, 2017.

[4] M. Berman and A. Fenaughty, "Technology and managed care: Patient benefits of telemedicine in a rural health care network," *Health Econ.*, vol. 14, no. 6, pp. 559–573, 2005.

[5] N. Sikka, S. Paradise, and M. Shu. (2014). *Telehealth in Emergency Medicine: A Primer*. Accessed: Nov. 12, 2017. [Online]. Available: https://www.acep.org/uploadedFiles/ACEP/Membership/Sections_of_Membership/telemd/ACEP%20Telemedicine%20Primer.pdf

[6] P. Kumar and H.-J. Lee, "Security issues in healthcare applications using wireless medical sensor networks: A survey," *Sensors*, vol. 12, no. 1, pp. 55–91, 2012.

[7] Y. Xiao, X. Shen, B. Sun, and L. Cai, "Security and privacy in RFID and applications in telemedicine," *IEEE Commun. Mag.*, vol. 44, no. 4, pp. 64–72, Apr. 2006.

[8] S. Mehta, R. Nallusamy, R. V. Marawar, and B. Prabhakaran, "A study of DWT and SVD based watermarking algorithms for patient privacy in medical images," in *Proc. IEEE Int. Conf. Healthcare Informat.*, Sep. 2013, pp. 287–296.

[9] D. S. Chauhan, A. Adarsh, B. Kumar, R. Gupta, and J. P. Saini, "Double secret key based medical image watermarking for secure telemedicine in cloud environment," in *Proc. 40th Int. Conf. Telecommun. Signal Process. (TSP)*, 2017, pp. 626–631.

[10] R. Eswaraiah and E. S. Reddy, "Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest," *IET Image Process.*, vol. 9, no. 8, pp. 615–625, 2015.

[11] E. Walia and A. Suneja, "Fragile and blind watermarking technique based on Weber's law for medical image authentication," *IET Comput. Vis.*, vol. 7, no. 1, pp. 9–19, Feb. 2013.

[12] S. Vellaisamy and V. Ramesh, "Inversion attack resilient zero-watermarking scheme for medical image authentication," *IET Image Process.*, vol. 8, no. 12, pp. 718–727, 2014.

[13] M. Alhussein and G. Muhammad, "Watermarking of Parkinson disease speech in cloud-based healthcare framework," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 10, p. 264575, 2015.

[14] J. W. Langston, "Parkinson's disease: Current and future challenges," *NeuroToxicology*, vol. 23, pp. 443–450, Oct. 2002.

[15] Z. Ali, M. Imran, W. Abdul, and M. Shoaib, "An innovative algorithm for privacy protection in a voice disorder detection system," in *Proc. 1st Int. Early Res. Career Enhancement School BICA Cybersecur. (FIERCES)*, 2018, pp. 228–233.

[16] M. Naor and A. Shamir, "Visual cryptography," in *Proc. Workshop Theory Appl. Cryptograph. Techn.*, Perugia, Italy, May 1994, pp. 1–12.

[17] J. Kreiman and D. Sidtis, "Producing a voice and controlling its sound," in *Foundations of Voice Studies*. Hoboken, NJ, USA: Wiley, 2011, pp. 25–71.

[18] P. L. Dhingra and S. Dhingra, *Diseases of EAR, NOSE & THROAT*, 6th ed. New Delhi, India: Elsevier, 2014.

[19] S. R. Schwartz *et al.*, "Clinical practice guideline: Hoarseness (dysphonia)," *Otolaryngol.-Head Neck Surgery*, vol. 141, pp. S1–S31, Sep. 2009.

[20] (Jan. 24, 2016). *The American Heritage Stedman's Medical Dictionary*. [Online]. Available: http://Dictionary.com and http://dictionary.reference.com/browse/dysphonia

[21] Z. Ali, M. Talha, and M. Alsulaiman, "A practical approach: Design and implementation of a healthcare software for screening of dysphonic patients," *IEEE Access*, vol. 5, pp. 5844–5857, 2017.

[22] Z. Ali *et al.*, "Voice pathology detection based on the modified voice contour and SVM," *Biol. Inspired Cognit. Archit.*, vol. 15, pp. 10–18, Jan. 2016.

[23] Z. Ali *et al.*, "Intra- and inter-database study for arabic, english, and german databases: Do conventional speech features detect voice pathology?" *J. Voice*, vol. 31, pp. 386.e1–386.e8, May 2017.

[24] G. Muhammad *et al.*, "Voice pathology detection using interlaced derivative pattern on glottal source excitation," *Biomed. Signal Process. Control*, vol. 31, pp. 156–164, Jan. 2017.

[25] A. A. Nasheri *et al.*, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *IEEE Access*, preprint, doi: 10.1109/ACCESS.2017.2696056.

[26] G. Muhammad *et al.*, "Automatic voice pathology detection and classification using vocal tract area irregularity," *Biocybern. Biomed. Eng.*, vol. 36, no. 2, pp. 309–317, 2016.

[27] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, and Z. Ali, "Investigation of voice pathology detection and classification on different frequency regions using correlation functions," *J. Voice*, vol. 31, no. 1, pp. 3–15, 2017.

[28] Z. Ali, I. Elamvazuthi, M. Alsulaiman, and G. Muhammad, "Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model," *J. Voice*, vol. 30, no. 6, pp. 757.e7–757.e19, 2016.

[29] J. I. Godino-Llorente, R. Fraile, N. Sáenz-Lechón, V. Osma-Ruiz, and P. Gómez-Vilda, "Automatic detection of voice impairments from text-dependent running speech," *Biomed. Signal Process. Control*, vol. 4, no. 3, pp. 176–182, 2009.

[30] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological VoiceSustained vowels versus continuous speech," *J. Speech, Lang., Hearing Res.*, vol. 44, pp. 327–339, Apr. 2001.

[31] B. Hammarberg, B. Fritzell, J. Gauffin, J. Sundberg, and L. Wedin, "Perceptual and acoustic correlates of abnormal voice qualities," *Acta Oto-Laryngol.*, vol. 90, nos. 1–6, pp. 441–451, 1980.

[32] M. A. S. Granero, J. E. T. Segovia, and J. G. Pérez, "Some comments on Hurst exponent and the long memory processes on capital markets," *Phys. A, Statist. Mech. Appl.*, vol. 387, pp. 5543–5551, Sep. 2008.

[33] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, 1968.

[34] H. E. Hurst, "Long-term storage of reservoirs: An experimental study," *Trans. Amer. Soc. Civil Engineers*, vol. 116, no. 1, pp. 770–799, 1951.

[35] B. Mandelbrot, "Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis," *Ann. Econ. Social Meas.*, vol. 1, no. 3, pp. 259–290, 1972.

[36] M. R. L. Hodges, "Effect of threshold offsets in zero-crossing speech detector," *Electron. Lett.*, vol. 17, no. 19, pp. 682–684, Sep. 1981.

[37] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, and R. Jennane, "One dimensional local binary pattern for bone texture characterization," *Pattern Anal. Appl.*, vol. 17, no. 1, pp. 179–193, 2014.

[38] N. Chatlani and J. J. Soraghan, "Local binary patterns for 1-D signal processing," in *Proc. 18th Eur. Signal Process. Conf.*, 2010, pp. 95–99.

[39] *Massachusette Eye & Ear Infirmry Voice & Speech LAB, Disordered Voice Database Model 4337 (Ver. 1.03)*, Kay Elemetrics Corp., Lincoln Park, NJ, USA, 1994.

[40] Z. Ali, G. Muhammad, and M. F. Alhamid, "An automatic health monitoring system for patients suffering from voice complications in smart cities," *IEEE Access*, vol. 5, pp. 3900–3908, 2017.

[41] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," *J. Speech, Lang., Hearing Res.*, vol. 43, no. 2, pp. 469–485, 2000.

[42] M. S. Hossain, G. Muhammad, and A. Alamri, "Smart healthcare monitoring: a voice pathology detection paradigm for smart cities," *Multimedia Syst.*, 2017, doi: 10.1007/s00530-017-0561-x.

[43] M. Hossain, "Patient state recognition system for healthcare using speech and facial expressions," *J. Med. Syst.*, vol. 40, no. 12, pp. 1–8, 2016.

[44] M. S. Hossain, "Cloud-supported cyber–physical localization framework for patients monitoring," *IEEE Syst. J.*, vol. 11, no. 1, pp. 118–127, Mar. 2017.

ZULFIQAR ALI received the Ph.D. degree from Universiti Teknologi Petronas, Malaysia, in 2017, the master's degree in computational mathematics from the University of the Punjab, Lahore, and the master's degree in computer science from the University of Engineering and Technology at Lahore, Lahore, with the specialization in system engineering. Since 2010, he has been a full-time Researcher with the Digital Speech Processing Group, College of Computer and Information Sciences, King Saud University, Saudi Arabia. His current research interests include speech and language processing, medical signal processing, privacy and security in healthcare, multimedia forensics, and computer-aided pronunciation training systems. He is a member of the Center for Intelligent Signal and Imaging Research, Universiti Teknologi Petronas.

M. SHAMIM HOSSAIN (SM'09) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada. He is currently a Professor of software engineering with King Saud University, Riyadh, Saudi Arabia. He has authored and co-authored around 165 publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include cloud networking, social media, IoT, cloud and multimedia for healthcare, smart health, and resource provisioning for big data processing on media clouds. He is a member of ACM and ACM SIGMM. He has served as a member of the organizing and technical committees for several international conferences and workshops. He has served as the co-chair, the general chair, the workshop chair, the publication chair, and TPC for over 12 IEEE and ACM conferences and workshops. He currently serves as the Co-Chair of the first IEEE ICME Workshop on Multimedia Services and Tools for Smart-Health in 2018. He was a recipient of a number of awards including, the Best Conference Paper Award, the 2016 *ACM Transactions on Multimedia Computing, Communications and Applications* Nicolas D. Georganas Best Paper Award, and the Research in Excellence Award from King Saud University. He is on the Editorial Board of the IEEE Access, the IEEE Multimedia, *Computers and Electrical Engineering* (Elsevier), *Games for Health Journal*, and the *International Journal of Multimedia Tools and Applications* (Springer). He served as a Guest Editor for the IEEE Transactions on Information Technology in Biomedicine (currently JBHI), the *International Journal of Multimedia Tools and Applications* (Springer), *Cluster Computing* (Springer), the *Future Generation Computer Systems* (Elsevier), *Computers and Electrical Engineering* (Elsevier), and the *International Journal of Distributed Sensor Networks*. He currently serves as a Lead Guest Editor for the *IEEE Communication Magazine*, the IEEE Transactions on Cloud Computing, the IEEE Access, the *Future Generation Computer Systems* (Elsevier), and *Sensors* (MDPI).

GHULAM MUHAMMAD received the Ph.D. degree in electrical and computer engineering from the Toyohashi University of Technology, Japan, in 2006. He is currently a Professor with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He has authored and co-authored many publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. He has supervised a number of Ph.D. and master's theses. He owns a U.S. patent. His research interests include serious games, cloud and multimedia for healthcare, resource provisioning for big data processing on media clouds and biologically inspired approach for multimedia and software system, image, and speech processing.

MUHAMMAD ASLAM received the Ph.D. degree in computer science and engineering from the CINVESTAV-IPN, Mexico. He is currently an Associate Professor with the Department of Computer Science and Engineering, University of Engineering and Technology at Lahore, Lahore, Pakistan. He has authored over 50 publications in different journals and conference of national and international repute. His current research interests include artificial intelligence, human–computer interface, image, speech processing, software engineering, computational intelligence, wireless networks, bioinformatics, and embedded systems.