Robust and secure watermarking scheme for breath sound

Baiying Lei\textsuperscript{a,b,*}, Insu Song\textsuperscript{a}, Shah Atiquur Rahman\textsuperscript{a}

\textsuperscript{a} James Cook University Australia, School of Business/IT, Singapore
\textsuperscript{b} Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, School of Medicine, Shenzhen University, Shenzhen 518060, China

\textbf{A R T I C L E   I N F O}

Article history:
Received 8 March 2012
Received in revised form 4 February 2013
Accepted 8 February 2013
Available online 17 February 2013

Keywords:
Digital watermarking
Robust
LWT–DCT
SVD
PSO
Breath sound

\textbf{A B S T R A C T}

Due to the development of the Internet, security and intellectual property protection have attracted significant interest in the copyright protection field recently. A novel watermarking scheme for breath sounds, combining lifting wavelet transform (LWT), discrete cosine transform (DCT), singular value decomposition (SVD) and dither modulation (DM) quantization is proposed in this paper as a way to insert encrypted source and identity information in breath sounds while maintaining significant biological signals. In the proposed scheme, LWT is first performed to decompose the signal, and then DCT is applied on the approximate coefficients. SVD is carried out on the LWT–DCT coefficients to derive singular values. DM is adopted to quantize the singular values of each of the LWT–DCT blocks; thus, the watermark extraction is blind by using the DM algorithm. The novelty of our proposed method also includes the introduction of the particle swarm optimization (PSO) technique to optimize the quantization steps for the DM approach. The experimental results demonstrate that the proposed watermarking scheme obtains good robustness against common manipulation attacks and preserves imperceptivity. The performance comparison results verify that our scheme outperforms existing approaches in terms of robustness and imperceptibility.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Advances in modern communication technologies, in particular the Internet and health social networks, allow anyone to have access to, make copies of, and distribute digital contents. However, this also causes a problem, as anybody can modify or tamper with important data, such as breath sounds for detecting the spread of infectious diseases. Tampered data can lead to false alarms or incorrect diagnoses of patients. As a result, copyright protection, data authentication and security have become challenging issues due to the illegal modification and distribution. As a data hiding and extraction method, digital watermarking is one of the solutions to address this problem and is used for digital rights protection, ownership verification and security purposes (Bender et al., 1996). For instance, some systems, such as Hospital Information System (HIS), address security problems and provide confidentiality, integrity and authentication via watermarking. Watermarking applications for medical purposes have been extensively investigated with reference to their security (Arsalan et al., 2011; Fakhari et al., 2011; Ko et al., 2012). For the breath sounds containing medical information, conventional watermarking is not appropriate because of the signal characteristics and distortion problems. Therefore, the evidence strongly suggests that a new watermarking scheme should be designed.

It is known that breath sounds (Pasterkamp et al., 1997) are a kind of sound signals which differ from musical or speech signals because they have different acoustic properties and watermarking requirements. Normal breath sounds have a narrow frequency range between 100 and 1000 Hz (Pasterkamp et al., 1997), unlike these sounds, musical sounds contain continuous polyphonic harmonic sounds characterized by series of chords. Breath sound (Gaviely, 1995; Gaviely and Cugell, 1996) is often characterized as the band-limited noise in that the amplitude and spectrum of the breath sound change in a characteristic manner and as a function of the airflow rate over a breath cycle. Breath sound may adopt only a small portion of the human perceptual range, and thus it is very acoustically rich. Compared to the 44.1 kHz sampling rate for CD-audio, breath sound signal is usually sampled at 4–32 kHz. Therefore, less information can be embedded in the breath sound signal. The expected channel noise is another difference. The channel noise for CD-audio is usually negligible, while the channel noise has a more effect on the low band–width breath sound, and thus it will be not suitable for real-time communication. Consequently, watermarking of the breath sound signals is different from the better known audio watermarking for intellectual property protection in several ways. However, the small band width breath sound signals can still use the same analysis techniques because of their similar sound signal characteristics. A new watermarking
scheme to protect the sound content and prevent information tampering should be designed to embed sensitive medical information which can be accessed only by authorized individuals for the purpose of copyright protection, integrity check and/or access control. Moreover, for the watermarking of breath sounds, the watermark should be encrypted in such a way that this watermark can only be checked by the owner and sound segments cannot be replaced with another watermark. Recently, the chaotic cat map (also called Arnold transform) (Wang and Zhao, 2006) has been identified as an efficient tool to encrypt the watermark data. As a kind of security mechanism in steganalysis, the scrambling of the hidden message should be adopted to shuffle the watermark before the public network transfer. Watermark security depends on the secret seed but not on the image. If there is no protection for the message hidden in the breath sounds, the identification is easily detected by attackers. To make the system more secure, one way is to adopt a chaotic system. Therefore, this study will use a chaos-based watermarking to improve security. In our work, cat map is utilized to achieve this encryption purpose. Generally, watermarking techniques that are based on transform domain are more popular than those based on time domain, since they provide higher quality and much more robustness. The most common transform domain techniques are Fourier transform, discrete cosine transform (DCT) (Huang et al., 2002) and discrete wavelet transform (DWT) (Wang and Zhao, 2006; Wu et al., 2005). Recently, other transforms, such as lifting wavelet transform (LWT) (Tao et al., 2010), and singular value decomposition (SVD) (Bhat et al., 2011, Bhat et al., 2010) are becoming a hot research topics, too. LWT (Sweldens, 1996) is often referred to as fast LWT in the sense that integer wavelet and scaling coefficients, rather than the floating point coefficients, can be obtained with lifting. SVD has also been widely applied in the watermarking field (Al-Nuaimy et al., 2011; Bhat et al., 2010, 2011; Lei et al., 2011, 2012a,b,c) as an effective and desirable transform technique. The basic principle of SVD is that most general signal processing attacks, such as rotation, scaling and flipping, will not affect the singular values of the cover object. As a result, SVD-based watermarking techniques have attracted a lot of interest (Al-Nuaimy et al., 2011; Bhat et al., 2010, 2011; Lei et al., 2011, 2012a,b,c).

Most of the current digital audio and speech watermarking schemes mainly focus on one or two transform combinations. For example, in Wang and Zhao (2006), a novel audio watermarking scheme based on DWT–DCT with synchronization technique was proposed by Wang and Zhao. The watermark bits are inserted into the host audio with the adaptive quantization method. It was claimed by Wang and Zhao that the proposed audio watermarking scheme is robust and imperceptible. However, one drawback of this scheme is that it is susceptible to time scale modifications and resampling attacks. Bhat et al. (2010) proposed a DWT–SVD method for copyright protection. The proposed method achieves good robustness results and outperforms the selected watermarking schemes. However, there is no security measure and the signal to noise ratio (SNR) results reported in this scheme are not very high. Obviously, LWT, DCT and SVD are the existing and known techniques, but the combination of them and the adoption of the adaptive DM quantization method for the watermarking of breath sounds are significant and novel. The justification for the combination of these transforms is that the combination of these transforms for watermarking can achieve the optimal watermarking performance in terms of robustness and imperceptibility. To the best of our knowledge, there is no LWT–DCT–SVD watermarking method for breath sounds in the literature. The reason to apply LWT, DCT and SVD sequentially is that the system can achieve optimal performance in terms of robustness and transparency in this order. Furthermore, our LWT–DCT–SVD watermarking method for breath sounds also differs from the existing watermarking methods.

The performance of watermarking can be improved in several ways. One recently developed and widely used way is to utilize evolutionary and artificial intelligence methods to view the watermarking problem as an optimization problem (Arsalan et al., 2011). In the recent literature (Ishtiaq et al., 2010; Wang et al., 2011), particle swarm optimization (PSO) has been applied in the watermarking field to improve the performance as an evolutionary optimization technique (Kennedy and Eberhart, 2001) inspired by social behavior of bird flocking or fish schooling. In a similar pattern of other evolutionary and stochastic computation techniques, such as genetic algorithm (GA), PSO shares similar characteristics, but it was proven that PSO could obtain better results in a faster and cheaper way compared with GA method. Besides, another fascinating feature of PSO is that there are fewer parameters to adjust compared with GA method. One PSO version can work very well in a wide variety of applications with slight modifications. PSO differs from GA because it has no evolution operators such as crossover and mutation. Initializing a population of random solutions first, PSO algorithm searches for optimal values by updating iterations. Therefore, PSO has been used for approaches in a wide range of fields. For more information about PSO, readers can refer to (Engelbrecht, 2005).

This paper proposes a new and robust watermarking technique based on a unique system in which breath sound data are recorded by digital devices. The proposed watermarking method based on DWT–DCT–SVD for breath sounds in the medical application is integrated with the PSO to improve the watermarking performance. A cat map is utilized to scramble the watermark data and increase confidentiality. The quantization steps are optimized by the PSO heuristic algorithm. The main significance of our proposed method is for the data tampering protection of our self-collected breath sounds for providing health information. Moreover, the meaningful image logo can also provide identification. The organization of this paper is as follows. Section 2 presents our proposed method in detail. Performance analysis is discussed in Section 3. Experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

2. Proposed method

2.1. Particle swarm optimization (PSO)

In PSO, each particle searches for optimal coordinates in the problem space. The coordinates are related with the best solution (fitness or objective value) (Kennedy and Eberhart, 2001). The objective value (called pbest) is also stored. Another “best” value (called gbest) is found by the particle swarm optimizer, which is the best value obtained by any particle in the neighbors of the particle. When a particle takes all the population as its topological neighbors, the best value is a global best. At each time step, the PSO process comprises changing and accelerating the velocity of each particle toward its pbest and gbest locations (PSO local version). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward pbest and gbest locations.

Let D denote the swarm size. Each individual particle i (1 ≤ i ≤ D) has its own position pi and velocity vi. These particles search for the optimal value of a given objective function iteratively, then locate their individual best positions pbest(pi) and keep track of the global position gbest(gbest) from all best positions through a search space. With respect to the two best values, the velocity and position of particle i (1 ≤ i ≤ D) in iteration t + 1 are updated by:

\[ p_i(t + 1) = p_i(t) + v_i(t + 1), \]

\[ v_i(t + 1) = c_1 r_1 (p_{i}(t) - x_i(t)) + c_2 r_2 (g_{best} - x_i(t)), \]
\[ v_i(t + 1) = c_0 \times v_i(t) + c_1 r_1(p^\text{best}_i(t) - x_i(t)) + c_2 r_2(p^\text{best}_g(t) - x_i(t)), \]

\[ x_i(t + 1) = x_i(t) + v_i(t + 1), \]

where \( c_1 \) and \( c_2 \) are referred as cognitive and social parameters respectively, which are positive constants, \( r_1 \) and \( r_2 \) are random numbers uniformly distributed in the range of \([0,1]\), \( c_0 \) is the inertia weight in the range of \([0,1]\), which controls the momentum of the particle and tunes changes in these values. These coefficients control how far a particle moves in a single iteration. \( v_i \) is the moving distance in one-step for a particle \( i \) and is limited to the range of \([v_{\text{min}}, v_{\text{max}}] \), where \( v_{\text{min}} \) and \( v_{\text{max}} \) are the minimum and maximum moving distance in one-step, respectively.

The qualitative measure of the selection of PSO algorithm parameter can be found in (Trelea, 2003). From the PSO equations, we can know that the trade-off is related to the PSO parameters such as \( c_0, \) \( \text{Iter}_{\text{max}} \) \( c_1 \) and \( c_2 \). Based on the analysis in (Trelea, 2003), the parameter couples that are close to the center of the stability triangle lead to quick convergence, while parameter couples that are close to its borders need many iterations to converge. After going through the whole process iteratively, objective evaluation function reaches the desired termination criterion.

2.2. Lifting wavelet transform (LWT)

LWT, introduced by Sweldens (1996) initially, adopts the in-place implementation of wavelet transform which leads to the reduction of the execution time. LWT utilizes lifting wavelet in the integer domain to simplify the problem, and thus has a strength over the traditional wavelet transform. There are three main steps of LWT, that is, split/merge, prediction and update. The three steps are described below:

1. Split step: The split step just splits the input signal \( X(n) \) into even and odd parts: \( Xe(n) \) and \( Xo(n) \).

\[ Xe(n) = X(2n), \]
\[ Xo(n) = X(2n + 1). \]

2. Prediction step: The prediction step predicts odd samples from even samples, which is a dual lifting step. The two signal subsets from the split process should be closely correlated. The difference between the predicted value of \( p[Xe(n)] \) and the real value of \( Xo(n) \) is called the detail signal \( d(n) \):

\[ d(n) = Xo(n) - p[Xe(n)], \]

where \( p[\cdot] \) is the predict operator. With the low-pass subband, the prediction step is lifting of the high-pass subband. Therefore, it is usually regarded as a high-pass filter.

3. Update step: The update step introduces the update operator \( U[\cdot] \), which adopts the detail signal \( d(n) \) to update even samples. The update step is the lifting of the low-pass subband with the high-pass subband, and thus it is often viewed as a low-pass filter.

\[ c(n) = Xe(n) + U[d(n)]. \]

In fact, the reconstruction of LWT is an inverse process of decomposition. The lifting scheme of decomposition and reconstruction is illustrated in Fig. 1.

2.3. Singular value decomposition (SVD) principle

SVD (Andrews and Patterson, 1976) is a matrix factorization which decomposes a matrix into three matrices of the same size.

For example, a real matrix \( I \) of size \( N \times M \) can be decomposed into a product of 3 matrices: \( I = USV^T \), where the superscript \( T \) denotes matrix transposition, \( U \) and \( V \) are real orthogonal \( N \times N \) matrices with small singular values, and \( S \) is \( N \times M \) size diagonal matrix with large singular values. Let \( r \) be matrix rank of \( I \), SVD equation can be defined by:

\[ I = USV^T = \sum_{i=1}^{r} \delta(i)u(i)v(i)^T = \sum_{i=1}^{r} u(i)\delta(i)v(i)^T, \]

where \( I \in \mathbb{R}^{N \times M}, U \in \mathbb{R}^{N \times N}, S \in \mathbb{R}^{N \times M}, \) and \( V \in \mathbb{R}^{N \times N}, \) \( \delta(i) \) are the singular values of the matrix \( I \) (diagonal elements of \( S \)) and assumed to be arranged in a decreasing order, that is, \( \delta(1) \geq \delta(2) \geq \cdots \geq \delta(r) > \delta(r+1) = \cdots = \delta(N) = 0 \).

2.4. Cat map encryption

The chaotic sequence often has low correlation sidelobes which means good invariance to disturbance. In order to disturb the pixel space relationship of the binary watermark image, the watermark is first permuted by the chaotic map. To enhance the security of the proposed watermarking scheme, a watermark scrambling algorithm is used at first. By using cat map, the binary watermark image \( B = (b(i, j), \ 1 \leq i \leq W_1, \ 1 \leq j \leq W_2) \) is scrambled to: \( BB = (bb(i, j)) \). The cat map is defined as:

\[ \begin{bmatrix} x(i + 1) \\ y(i + 1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x(i) \\ y(i) \end{bmatrix} \mod N. \]

We should convert the two-dimensional binary image into the one-dimensional sequence in order to embed it into the breath sound signal. The corresponding one-dimensional sequence is given by:

\[ BW = (w(n) = bb(i, j), \ 1 \leq i \leq W_1, \ 1 \leq j \leq W_2, \ n = i \times W_2 + j, \ n = i \times W_2 + j, \ w(n) \in \{0, 1\}). \]

It is further encrypted using random permutation to ensure security. Finally, the cat map encrypted data, \( w(n) \), is used for the watermark embedding.

2.5. Watermark embedding

Fig. 2 describes the block diagram of the watermark embedding algorithm. In our scheme, the combined features of LWT, DCT, SVD and DM are fully utilized.

We chose the popular DM quantization method in the embedding process because of its good robustness and blind nature. The breath sound is first pre-processed with the normalization method for providing data consistency and improving the watermarking performance. Specifically, the following steps describe the embedding process:
**Step 1:** Perform LWT on the host breath sound signal, SB, to generate the coefficients of approximations and details: $A(i)^K$, $d(i)^K$, $d(i)^{K-1}$, ..., $d(i)^1$. $A(i)^K$ is the approximate signal, and $d(i)^K$, $d(i)^{K-1}$, ..., $d(i)^1$ are the detail signals. $K$ is the decomposition level.

$$\begin{align*}
[A(i)^K, d(i)^K, d(i)^{K-1}, ..., d(i)^1] &= \text{LWT}(SB).
\end{align*}$$

**Step 2:** In most watermarking schemes, there is a need for a trade-off between robustness and imperceptibility, and thus coefficients in the certain frequencies are usually selected for watermark embedding to achieve balance. Since we want to develop a watermarking scheme for source identification and data tampering protection of low bit rate breath sound signals, the watermark should survive potential attacks over the entire channel. Besides, we need to embed as many watermark bits as possible in the LWT coefficients. It was found that the high frequency band of the signal has low energy values and robustness, while the low frequency LWT coefficients of the signal usually have higher energy values; hence, they are more robust against common signal processing manipulations. The high frequency of the signal can be easily removed by distortions such as low-pass filtering, too. To achieve optimal performance, DCT is only applied to the low frequency approximate coefficients of the signal, that is, $A(i)^K$.

$$A(i)^{\text{DCT}} = \text{DCT}(A(i)^K).$$

**Step 3:** Cut the generated LWT–DCT coefficients, $A(i)^{\text{DCT}}$, into different blocks. Specifically, the block length is equal to dividing the length of the approximate coefficients by the watermark length.

**Step 4:** Conduct SVD transformation on each block to produce singular values:

$$A(i)^{\text{SVD}} = USV^T.$$  

**Step 5:** In order to guarantee robustness and transparency of watermarking scheme, the watermarking system inserts the watermark bits into the LWT–DCT coefficients by adaptive DM quantization method. The first singular values of each block have the highest energy values, and thus they are utilized to embed the watermark. The embedding rule is:

$$S' = \begin{cases} 
\text{round} \left( \frac{S}{\mu} \times \mu + \frac{3}{4} \mu \right) & w(n) = 1 \\
\text{round} \left( \frac{S}{\mu} \times \mu + \frac{1}{4} \mu \right) & w(n) = 0
\end{cases}$$

where $\mu$ is the quantization step, round $[\cdot]$ is rounding to the nearest integer value, $w(n)$ is cat map encrypted signal. It is obvious that the quantization step is significant in terms of both robustness and inaudibility. The larger the quantization step is, the more robust, but less transparent, the watermarking scheme is. Therefore, the quantization step should be specially developed to achieve optimal performance.

**Step 6:** Inverse SVD transformation is conducted to obtain the watermarked block:

$$A(i)^{\text{SVD}} = USV^T.$$
optimized. Therefore, different host signals or watermarks would require different quantization steps depending on their signal characteristics. To address this issue, the evolutionary algorithm, PSO, is used to search for the best quantization steps for different host signals and watermarks. In our scheme, PSO should be applied under this constraint because the quantization step for our watermarking is in the range of [0, 1].

A lot of common signal processing attacks can be applied to measure the watermarking performance. Among these attacks, resampling, MP3 compression, noise addition and cropping are commonly used, and hence they are selected to be embedded in our PSO training procedure. The weights should also be taken into a special consideration in the evaluation of objective values. In the watermarking field, the properties of watermark robustness, imperceptibility and capacity are mutually contradictory to each other. Since we only want to achieve a tradeoff between watermark imperceptibility and robustness, the watermark capacity is fixed in the PSO process.

The objective value function should first be designed as a function of both imperceptibility and robustness to obtain the optimal performance by PSO. In our method, the robustness between the original watermark \( w \) and extracted watermark \( w' \) is measured by normalized correlation, \( \text{Corr}(w, w') \), which is defined as:

\[
\text{Corr}(w, w') = \frac{\sum_{n=1}^{N_w} w(n)w'(n)}{\sqrt{\sum_{n=1}^{N_w} w^2(n)} \sqrt{\sum_{n=1}^{N_w} w'^2(n)}}.
\]  

Similarly, the popular objective measure of the watermarking perceptual quality, signal to noise ratio (SNR) and segmental signal to noise ratio (SSNR) are employed in the PSO training process to evaluate the imperceptibility performance. SNR and SSNR are defined as follows:

\[
\text{SNR} = 10 \log_{10} \left( \frac{\sum_{i=1}^{L} S^2(i)}{\sum_{i=1}^{L} (S(i) - S(i))^2} \right),
\]

\[
\text{SSNR} = 10 \log_{10} \left( \frac{\sum_{m=0}^{K-1} \sqrt{\sum_{i=1}^{r} S^2(i)}}{\sum_{m=0}^{K-1} \sqrt{\sum_{i=1}^{r} (S(i) - S(i))^2}} \right),
\]

where \( S(i) \) and \( S'(i) \) correspond to the original and watermarked signal respectively.

Moreover, another new perceptual measurement, log likelihood ratio (LLR) (Crochiere et al., 1980), is also adopted to further measure the imperceptibility of the watermarking scheme and is included in the PSO training. It is assumed that the audio segment can be represented by a \( p \)th order all-pole linear predictive coding model of the form and denoted as:

\[
S_n = \sum_{m=1}^{p} a_m S_{n-m} + G_n u_n,
\]

where \( S_n \) is the \( n \)th audio sample, \( a_m \) (\( 1 \leq m \leq P \)) are the coefficients of an all-pole filter, \( G_n \) is the gain of the filter and \( u_n \) is an appropriate excitation source for the filter. The LLR metric based on this assumption is defined as:

\[
\text{LLR} = \log \left( \frac{\tilde{a}_T \tilde{a}^T}{\tilde{a}_T \Sigma_T a^T} \right).
\]

Accordingly, the final objective function in the proposed method is designed as:

\[
\text{Objective} = \alpha \times (\text{SNR} - 20) + \beta \times (\text{SSNR} - 20) + \lambda \times \text{LLR} + \frac{1}{R} \sum_{i=1}^{K} \gamma_i \text{Corr}(w, w'),
\]

where \( \alpha, \beta, \lambda \) and \( \gamma_i \) are weights to have a good balance between robustness and imperceptibility, \( R \) is the total number of attacks in the PSO optimization process. As SNR and SSNR are often above 20 dB for a good watermarking scheme, 20 is used to balance the objective value too. Fig. 4 shows the PSO training process in our watermarking algorithm.

3. Performance analysis

3.1. Watermark capacity

The watermark capacity (also known as the payload) of a watermarking scheme is defined as the number of bits that can be embedded and recovered in the audio stream, and often measured in bits-per-seconds (bps) (Bhat et al., 2010; Lei et al., 2012a,b,c). Here, the duration time of the host signal relative to the size of the watermarked signal is used, and denoted as:

\[
\text{Payload} = \frac{N_w \text{Time}}{10^6},
\]

where \( \text{Time} \) is the duration of the host audio and \( N_w \) is the watermarked length. For our scheme, \( N_w = 4096 \) bits is embedded in a host audio of duration of 6 s, thus the payload of our method is 170.67 bps. This is a relatively high payload as a typical payload is 20–50 bps.
3.2. Error analysis

Bit error rate (BER) is adopted to evaluate the proposed watermarking scheme, which is defined as below:

$$\text{BER} = \frac{1}{N_w} \sum_{n=1}^{N_w} w(n) \oplus w'(n),$$  \hspace{1cm} (27)

where $\oplus$ is the XOR (exclusive or) operator. BER is used to evaluate our water marking methods against signal distortions or collusion attacks. A procedure is defined to evaluate whether our watermark scheme has a sufficient confidence to counter on the collusion attack to decide the sign of the hidden watermark or not. This procedure adopts the commonly used watermark detection and is regarded as an efficient approach in the existing feature-based watermarking methods (Lu and Hsu, 2007; Seo and Yoo, 2004; Tsai et al., 2011; Bhat et al., 2010).

To decide whether there is a watermark, the original watermark $w$ is compared with the extracted watermark $w'$. If the BER between $w$ and $w'$ is less than a user-defined threshold $L$, there is a watermark, otherwise, there is no watermark. Actually, the threshold $L$ is determined by the probability of the detection error due to a false alarm or rejection detection (Bhat et al., 2010). To determine the watermark threshold $T$, the false alarm and rejection are usually taken into consideration. The false alarm rate $P_{FA}$ is defined as the probability of marking the unwatermarked signal as watermarked. The false rejection rate $P_{FR}$ is the probability of rejecting a watermarked signal as unwatermarked one. The selection of the threshold is done by a tradeoff between $P_{FA}$ and $P_{FR}$. It is very difficult to analyze $P_{FA}$ due to many processing steps since the exact characteristics are not known.

A block (e.g., a 6 s of duration of audio) is marked as watermarked if at least $L$ number of bits of a watermark is detected correctly in the block. The presence of a watermark in the host breath sound is determined if at least one block is determined to be watermarked. The false alarm rate (i.e., the probability of the false positive detection of the watermark in a signal) can then be defined as follows using the binomial distribution approximation:

$$P_{FA} = \sum_{i=L}^{N_w} \binom{N_w}{i} (P_{FP-B})^i (1 - P_{FP-B})^{N_w - i}.$$  \hspace{1cm} (28)

The value of $L$ could be determined according to the desired probability of false-positive detection of the host signal. $P_{FP-B}$ is the probability that a bit is falsely detected as a watermark bit. From the above derived process, it can be noted that the false alarm probability $P_{FA}$ for a host signal depends on both the correlation detection threshold $L$ and $N_w$. Here $P_{FP-B}$ is assumed to be $1/2$, and $L = (1 - \text{BER}) \times N_w$, where BER is bit error rate between the original watermark bits and the extracted watermark bits. According to the watermarking requirements, the desired false alarm rate should be smaller, the better. The false alarm rate of our scheme is illustrated in Fig. 5, which demonstrates the good error probability of the proposed algorithm. As can be seen in Fig. 5, the curves of $P_{FA}$ are plotted over the watermark length of each BER values for detection threshold. It can be observed that the false positive error is higher if the BER value is higher and watermark length is shorter. The average BER of our approach on the test data sets is 0.008975 which is smaller than BER = 0.1 (see Table 3). In our simulation, $N_w$ is equal to 4096, and thus $P_{FA}$ is near zero (less than $10^{-8}$) when the watermark length is 4096.

Suppose that watermark bit 0 is negative, while bit 1 is positive, we will justify our confidence in collusive estimation of a watermark sign using binomial probability distribution. Suppose each is regarded as a trial $\text{sgn}(w_b^i(1))$, $b \in N_w$, and that the trials are independent. Let $\beta$ be the random variable denoting the number of...
observed in $N_w$. As a consequence, the confidence can be expressed as:

$$P_s \left( \beta > \frac{|N_w|}{2} \right) = \sum_{n=|N_w|/2+1}^{|N_w|} \binom{|N_w|}{n} 0.5^{N_w}.$$

Looking at the table of binomial probabilities, we can find that $P_s$ will increase rapidly (usually it will be larger than 0.8) as long as it is slightly larger than $|N_w|/2$. The larger $P_s$, the more confident the scheme is. Therefore, the watermark scheme has a sufficient confidence to counter on the collusion attack to decide the sign of the hidden watermark.

### 3.3. Key space analysis

For a secure watermarking scheme for breath sounds, security is a very important issue. To improve confidentiality, the key space should be large enough to boost confidentiality and render attacks, especially brute force attacks, impossible. Therefore, secret keys are adopted for security purposes. Actually, the security of the watermarking scheme mainly depends on the security of the initial seeds rather than the host and watermark signal. In the proposed secure watermarking scheme for breath sounds, the secret keys $K_1$ and $K_2$ are adopted for security purposes. It is common that digital computers store floating-point numbers using 32 bits, which consist of an exponent being represented using eight bits and a significant number being represented using 24 bits. In our scheme, keys or initial conditions are floating-point numbers. Hence, the exponent is fixed while the significant number may be varied. As a result, the total number of possible initial conditions is 2 to the power of 24 which is more than 16 million. This number can also be greatly increased if so desired through the use of double precision floating point numbers. Actually, the security of an information system depends on keys rather than the privacy of the scheme. In our proposed watermarking scheme for breath sounds, we adopt the two keys $K_1$, $K_2$ to generate a cat map, thereby enhancing the security of the proposed scheme. The size of the key value space influences the security of the proposed scheme. As keys $K_1$, $K_2$ are both utilized in our scheme, we take $K_1$ as an example and compute its key value space as follows:

Suppose $K_1 = \{0 < K_1(i) < 1 | i = 1, 2, ..., N_1\}$, $N_1$ is an integer used to generate the sequences and should be large enough to produce sequence, $Z = \{z(i, j) | i = 1, 2, ..., N_1, j = 1, 2, ..., N_2\}$, where $N_1$ denotes the number of chaotic bit sequences and $N_2$ represents the length of each chaotic sequence. In the same way, the disturbance of the key $K_1$ generates $K'_1$, $K'_1 = \{0 < K'_1(i) + d < 1 | i = 1, 2, ..., N_1\}$. With the same method, $K_1$ is generated and used to obtain another similar chaotic sequence: $Z' = \{z'(i, j) | i = 1, 2, ..., N_1, j = 1, 2, ..., N_2\}$. A simple function, $F = f(d)$, is used to test key space of $K_1$ as below:

$$F = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} |z(i, j) - z'(i, j)|}{N_1 \times N_2}.$$

![Fig. 6. Key space under different initial value difference.](Image)

Fig. 6 plots the function $F = f(d)$. It is observed that $f$ is equal to 0 when $d = 10^{-17}$ in our method. Thus the key space of $K_1$ is $1/d = 10^{17}$. Similarly, the key spaces of $K_2$ can be calculated in the same way as $K_1$. As can be seen from Fig. 6, the key space of $K_2$ is $1/d = 10^{17}$. Therefore, the total key space of our watermarking scheme is $10^{34}$, which means that there is enough key space to guarantee high confidentiality of our proposed watermarking system (Fan and Wang, 2009; Lei et al., 2011). As $K_1$ and $K_2$ are two different initial values of $x$ and $y$ in the cat map, $f(d)$ values are quite similar but not the same though the first initial values of $x(1)$ and $y(1)$ are the same, which verify the random nature of the chaotic sequence. Based on this key
space analysis, it can be concluded that the embedded watermarks are secure to attackers who try to exhaustively or statistically detect and read them. All in all, our proposed scheme with such a long key is adequate for reliable and practical use (Fan and Wang, 2009; Lei et al., 2011).

3.4. Security analysis of the estimation attack

The randomization of the watermark by cat map can mitigate the security risk. In order to decrypt the encryption function adopted to construct the content-dependent watermark,

$$ h(y) = \int f(y) \log f(y) dy, $$

where \( Y \) is the support set of the random variable and \( f(y) \) is probability density function. Suppose that a secret key is uniformly distributed over the interval \([\alpha_{\min}, \alpha_{\max}]\). Considering various radii for generating the watermark regions, the probability density function of a secret radius \( r \) (Tsai et al., 2012) is given by:

$$ f(r) = \begin{cases} 0 & \text{if } \alpha_{\min} \leq \alpha \leq \alpha_{\max} \
\frac{1}{\alpha_{\max} - \alpha_{\min}} & \text{otherwise} \end{cases} $$

The differential entropy of the watermark region detection can be written as:

$$ h(r) = \int_{r_{\min}}^{r_{\max}} f(r) \log_2 \frac{1}{f(r)} dr $$

$$ = -\int_{r_{\min}}^{r_{\max}} \frac{1}{\alpha_{\max} - \alpha_{\min}} \log_2 \frac{1}{\alpha_{\max} - \alpha_{\min}} dr $$

The randomness in the watermark region selection is mainly on the key-dependent pseudo-random number, which are Gaussian distributed with mean \( \mu \) and variance \( \sigma^2 \). The combined values can be considered as the weighted Gaussian distributed random variables, \( y \), with mean and variance shown by:

$$ E(y) = \mu \bullet r, $$

$$ \text{Var}(y) = \sigma^2 \bullet r^2. $$

Its probability density function is given by

$$ f(y) = \frac{1}{\sqrt{2\pi\sigma^2 \bullet r^2}} e^{-\frac{(y-\mu r)^2}{2\sigma^2 \bullet r^2}}. $$

Therefore, the differential entropy of the watermark region selection can be written as:

$$ h(y) = \frac{1}{2} \log_2 2\pi e \sigma^2 \bullet r. $$

Changing the base of the logarithm, the differential entropy is:

$$ h(y) = \frac{1}{2} \log_2 2\pi e \sigma^2 \bullet r. $$

Consequently, the degree of randomness to the watermark detection and watermarking region of the proposed scheme can be estimated by Eqs. (33) and (38), respectively. Moreover, the degree of security is enhanced by increasing the interval \( \alpha_{\min}, \alpha_{\max} \) or the variance \( \sigma^2 \) (Tsai et al., 2012). The interval controlling the secret radius is constricted to the watermarking algorithm, the host signal length and the watermark length. As the capacity of watermarking region cannot be smaller than the watermark length, the secret radius should not be too small. Additionally, it should not be too large, as the range of the watermarking region cannot be larger than the length of the host signal (Swaminathan et al., 2006; Lu and Hsu, 2007). The randomness of watermark detection and watermarking region selection prevents an attacker from estimating the exact range and location of the watermarked area.

4. Experimental results

To evaluate the performance of the watermarking scheme, we collected a total of 90 breath sounds from different children under the age of 9. Among the 90 breath sounds, 40 of them were recorded in indoor environments and 50 of them were recorded in outdoor environments. Along with the recording of the breath sounds, general well-being questions were asked about each participant (e.g., “whether the child is energetic or sluggish, having pneumonia, cold or flu”). Information from these questions could be embedded with the breath sounds later and used for research purpose in identifying the state of well-being of the children from these sound data. The recordings were uploaded to a personal computer, converted to the WAV file format with a fixed duration time and then divided into different segments to facilitate the analysis. The sounds were sampled at 32 kHz with 16 bit resolution. Based on an empirical experience and the trial and error, the PSO optimization parameters were set as \( c_0 = 0.4, \ c_1 = c_2 = 1.8 \) to achieve the optimal robustness and transparency. For the objective equation parameters, since all the parameters play the same important role in the optimization process, the same weight parameters were employed in our method, that is, \( \alpha = \beta = \lambda = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.5 \).

A 64 × 64 bit binary JCU image logo was adopted in our watermarking method. The duration time of the breath sound data was 32.768 s. In our observations, a larger decomposition level could not increase the watermarking robustness and caused intensive computation. Thus, we chose 3 as the decomposition level of the Haar wavelet as a tradeoff between robustness and imperceptibility. BER was adopted to evaluate the proposed watermarking scheme.
4.1. Imperceptibility

Two representative breath sound samples were used in our scheme for the performance comparison. Fig. 7 plots the time domain waveforms and frequency domain spectrogram of the original and watermarked breath sounds, respectively. No observed visual differences were found between the time domain waveforms. The low frequency part of the frequency domain spectrum of the original and watermarked signal were almost the same, while the high frequency part of the spectrogram of the original and watermarked signal had only some noise differences, which verified the successful insertion of the watermark and the good imperceptibility of the watermarking scheme. Moreover, the high SNR and LLR values after watermarking also validated good transparency of the breath sound watermarking scheme objectively.

4.2. Subjective listening test

As we know, SNR is a simple way to present the sense of imperceptibility by measuring the signal distortion caused by watermarking. However, human perception may not corroborate well with the SNR measure. Consequently, subjective quality evaluation of the watermarking methods must be conducted to provide a better test of inaudibility based on human perception. In our experiment, we performed an informal subjective listening test to evaluate the perceptual quality of the watermarked audio.

For the subjective test, the popular mean opinion score (MOS) method was utilized. The 5-level MOS scores, “5” (imperceptible), “4” (perceptible, but not annoying), “3” (slightly annoying), “2” (annoying), and “1” (very annoying) were used to report the subjective quality. The average MOS score of the 40 test audio signal was 4.5, which demonstrates good subjective listening results.

Results of PSO technique

The conflicting problem of imperceptibility and robustness of the watermark was addressed by the PSO based optimization method. Figs. 8 and 9 show the results of PSO optimization of our proposed LWT–DCT–SVD watermarking method after 50 generations of the two presentive breath sound signals. The quantization steps, SNR, SSNR, LLR, and the average Corr values of the four attacks gradually converged to the saturation values as the number of generations was increased. These results validate the positive effect of PSO training and the successful optimization technique. These parameters were optimally varied to achieve the most desirable ones for host signals according to their different breath sound characteristics. Moreover, it was observed that 50 generations were enough to achieve the optimal values from these figures. Therefore, 50 generations were used in our simulations.

4.4. Robustness

General signal processing attacks such as resampling (resampling to half of the sampling frequency and then back to the sampling frequency), requantization (16 bits-8bits-16bits), MP3 compression (at bit rate of 112, 96 and 64 kbit/s), cropping (10% of the original samples are cropped) and noise addition (with the SNR 30 dB) were conducted to validate the robustness of our scheme. The robustness results and extracted watermarks under various kinds of attacks are shown in Table 1. It can be noted that the watermarks under attacks are still visually recognizable and can,
therefore, be used for data source identification and data tampering detection purposes for breath sound signals.

For the robustness test, we also compared our method with existing methods in Wang and Zhao (2006), Tao et al. (2010) and Bhat et al. (2010). We choose these methods in Wang and Zhao (2006), Tao et al. (2010) and Bhat et al. (2010) because they utilize the common and similar transforms, DWT–DCT, LWT and DWT–SVD. Table 2 shows the Corr results under various attacks as well as the performance comparison results with the methods in Wang and Zhao (2006), Tao et al. (2010) and Bhat et al. (2010). From the comparison results, it can be seen that our algorithm is comparable to the algorithms in Wang and Zhao (2006), Tao et al. (2010) and Bhat et al. (2010), as the average Corr values are very similar. As our watermarking method was applied to breath sounds, the robustness performance of our method is relatively lower than that of methods that utilize wide bandwidth audio signals. However, the robustness of our watermarking for breath sounds is still acceptable.

Table 3 presents the BER (%) results under various common signal processing attacks and the comparison results with methods in Wang and Zhao (2006), Tao et al. (2010), Bhat et al. (2010). In Table 3, the schemes in Wang and Zhao (2006) and Tao et al. (2010) are without SVD, while scheme in Bhat et al. (2010) and our scheme are the schemes with SVD. Since the watermark embedding and extraction are inserted into the singular values, therefore, the comparison results with and without SVD are not provided in the paper. It can be observed that BER (%) results in our method are a bit lower than that in the method in Bhat et al. (2010) as our watermarking

![Quantization step, SNR, SSNR, Corr and LLR values of breath sounds 1 with PSO training using 20 particles after 50 generations](image1)

**Fig. 8.** Quantization step, SNR, SSNR, Corr and LLR values of breath sounds 1 with PSO training using 20 particles after 50 generations; (a) best quantization step results (best quantization step = 0.1289); (b) best SNR and SSNR results (best SNR = 24.7532 and best SSNR = 13.2821); (c) best LLR results (best LLR = 1.8515); and (d) best Corr results (best Corr = 0.9932).

![Quantization step, SNR, SSNR, Corr and LLR values of breath sounds 2 with PSO training using 20 particles after 50 generations](image2)

**Fig. 9.** Quantization step, SNR, SSNR, Corr and LLR values of breath sounds 2 with PSO training using 20 particles after 50 generations; (a) best quantization step results (best quantization step = 0.1459); (b) best SNR and SSNR results (best SNR = 23.4612 and best SSNR = 15.28); (c) best Corr results (best Corr = 0.9936); and (d) best LLR results (best LLR = 1.826).
method is based on the low frequency breath sound. Table 3 shows the superiority of methods with SVD over that without SVD.

The overall robustness performance after attacks of our breath sound watermarking scheme is quite good. This watermarking scheme can achieve source identification and data tampering protection based on the robustness performance and the visually recognizable extracted watermarks after attacks.

5. Conclusions

In this paper, a robust and blind watermarking scheme for medical breath sounds based on LWT–DCT with the exploration of SVD properties and DM quantization was proposed. The attractive properties of SVD, LWT, DCT and DM quantization techniques make our scheme very robust to various common signal processing attacks. This robustness was validated by the simulation results. In our optimization process, PSO was used to make an optimal trade-off between imperceptibility and robustness through effective selection of quantization steps to locate the best parameters to insert the watermarks. The quantization steps were optimally adapted to achieve the most suitable performance for various breath sounds with different characteristics for the medical applications. Additionally, the proposed watermarking scheme also achieved very good transparency results. The experimental results revealed that our proposed method can increase the robustness of the embedded watermark against various attacks in comparison to previous work. In general, the proposed method performed much better than the DCT method for almost all of the images and attacks. As the identification information can be exactly extracted without affecting the imperceptibility of the breath sounds, it is especially suited to medical breath sound applications. Therefore, health information, such as patients’ data, digital signatures and identification codes, can be well embedded in medical breath sounds. It is obvious that the LWT–DCT–SVD medical breath sound watermarking is preferable to facilitate data management in health information management systems.

Future work can focus on enhancing patient privacy by tracing the source of an unauthorized release of breath sounds in networks.

Watermarking medical breath sounds for tracing purposes needs to satisfy fidelity requirements to ensure diagnostic quality and high robustness for tracing purposes. Future work can include the investigation of other evolutionary algorithms for the performance improvement with respect to the existing algorithms.

Acknowledgements

This project is funded by a grant from the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative (Grant Number: OPP1032125).

References


Baiying Lei obtained her M.Eng. degree in Electronics Science and Technology from the Department of Information Science and Electronic Engineering, Zhejiang University, China in 2007, and Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore in 2012. Her research interests include signal processing, audio, image content protection with watermarking and encryption, computer vision, health informatics, pattern recognition and machine learning.

Insu Song received the B.Sc. in Physics from Chung-Ang University, Seoul, Korea in 1991, B.InfoTech (Hons.) from Griffith University, Australia in 2004, and Ph.D. in computer science from the University of Queensland, Australia in 2008. He is currently Lecturer of Information Technology at James Cook University Australia. His research interests include biomedical engineering, health informatics, mental health informatics, knowledge engineering, and text mining. He has more than 10 years’ experience in information systems design, embedded system design, and electronics engineering.

Shah Atiqur Rahman received B.Sc. Eng. Degree in computer science and engineering from Khulna University of engineering and technology, Khulna, Bangladesh in 2003 and Ph.D. degree from Nanyang Technological University, Singapore in 2012. Currently he is working as a Lecturer in the school of Business and IT at James Cook University Australia, Singapore campus. His research interests include image processing, signal processing, health informatics, pattern recognition, machine learning, and computer vision.