Journal of King Saud University – Computer and Information Sciences (2015) 27, 58–67

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Blind digital speech watermarking based on Eigen-value quantization in DWT

Mohammad Ali Nematollahi *, **S.A.R. Al-Haddad, Faraneh Zarafshan**

Department of Computer & Communication Systems Engineering, Faculty of Engineering, University Putra Malaysia, UPM Serdang, 43400 Selangor Darul Ehsan, Malaysia

Received 1 November 2013; revised 17 February 2014; accepted 13 March 2014
Available online 29 December 2014

KEYWORDS

Blind digital speech watermarking;
Eigen-value;
Discrete wavelet transform;
Singular value decomposition

Abstract This paper presents a new blind digital speech watermarking technique based on Eigen-value quantization in Discrete Wavelet Transform. Initially, each frame of the digital speech was transformed into the wavelet domain by applying Discrete Wavelet Transform. Then, the Eigen-value of Approximation Coefficients was computed by using Singular Value Decomposition. Finally, the watermark bits were embedded by quantization of the Eigen-value. The experimental results show that this watermarking technique is robust against different attacks such as filtering, additive noise, resampling, and cropping. Applying new robust transforms, adaptive quantization steps and synchronization techniques can be the future trends in this field.

1. Introduction

In recent years, the digital media is distributed through the Internet by many companies, organizations, and users in the society. However, the security and copyright protection of the media are always the main concerns for them. Digital watermarking is the proper technique to protect and monitor the digital media (Phadikar, 2013). Although many audio watermarking techniques have been proposed (Bhat et al., 2010), they are not suitable to be used for digital speech watermarking. Digital speech is different from audio signal in respect to factors like production model, perception, bandwidth, loudness, and intensity (Kent and Read, 2002).

Watermarking is the process of embedding extra information in the host media in such a way that its presence is not distinguishable. There is a tradeoff between capacity, imperceptibility, and robustness in watermarking. Increasing one of the factors can decrease other factors. Although all of these factors must be considered when designing the digital speech watermarking technique, depending on the application, one of them can be added. For example, in air traffic control, the capacity and payload are more important due to the amount of information which must be sent (Hofbauer, 2009). However, for security improvement in speaker recognition (Faundez-Zanuy et al., 2006; Faundez-Zanuy et al., 2007), more concern is on the robustness of watermarking.

* Corresponding author.
E-mail address: Greencomputinguae@gmail.com (M.A. Nematollahi).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2014.03.012
1319-1578 © 2014 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

The digital audio watermarking techniques proposed in various studies can be classified into five main categories: auditory masking, phase modulation, quantization, transformation, and parametric modeling (Nematollahi and Al-Haddad, 2013). However, in this paper, the watermark bits are embedded into the Eigen-value of Approximation Coefficient (AC) which is computed by Discrete Wavelet Transform (DWT). This proposed technique has good imperceptibility, capacity, and robustness. Furthermore, this technique is more adaptive and flexible for different purposes. For example, the frame size can directly determine the watermark's capacity. If the frame size is small, the capacity is increased. Furthermore, selecting between AC and Detail Coefficient (DC) can directly affect the robustness and imperceptibility of the watermark. Although embedding the watermark in AC increases its robustness, it may affect its imperceptibility.

Although applying DWT and SVD techniques for watermarking is not a new idea (Al-Haj and Ahmad, 2010), the main contributions of this paper are listed as follows for better clarification of its differences from Al-Haj (Al-Haj and Ahmad, 2010) work:

1. In this study, the watermark is embedded in AC of DWT but not DC of DWT. This study shows that embedding the watermark in AC is more robust due to more energy and low frequency obtained.
2. This study mainly concentrates on speech and offers more insights into speech which has its own characteristics.
3. The proposed algorithm is blind, that is, it does not need the original signals in the process of watermark extraction or detection. Furthermore, the proposed algorithm provides more robustness, capacity, and imperceptibility than the algorithms in other DWT–SVD schemes (Al-Haj and Ahmad, 2010).
4. This study shows that selecting Quantization Index Modulation (QIM) for watermarking is simple and needs less time. Furthermore, applying different quantization steps in the QIM technique offers more tradeoff for watermarking requirements in terms of robustness and imperceptibility.

The rest of this paper is organized as follows: first, the main characteristics of the speech signal are discussed; second, DWT and SVD techniques are discussed, third, related and pervious works in watermarking are explained; fourth, the digital speech watermarking technique is proposed; fifth, the proposed digital speech watermarking is evaluated in the section on experimental results; sixth, the proposed digital speech watermarking is discussed and compared to that of Al-Haj (Al-Haj and Ahmad, 2010); and finally, the conclusions and future trends in this field are discussed.

2. Speech characteristics

Speech has slower time varying signals which are considered as almost stationary with short durations (of between 10 ms and 30 ms). This characteristic causes the speech signal to have a well-established and more predictable spectrum. In (Al-Shoshan and Abdullah, 2006), the speech signal is compared to audio from three viewpoints such as spectral structure, temporal structure, and syntactic or semantic structure. The brief differences are presented as follows:

Tonality: in contrast to music signal with multiple tones, each with a unique distribution of harmonics, speech has concentrated on voice tonality.

Alternative sequence: The speech signal has an alternative sequence of sound segments which are distributed through its spectrum more randomly than in music as speech sound is noise-like whereas music is tonal in shape.

Bandwidth: The power of the speech frequency is mostly located in the low frequency (less than 4 kHz).

Fundamental and formant frequency: Every person has specific or unique frequencies which are used for speaker recognition (Aldhaheri and Al-Saadi, 2004). Although the unique parameters for a specific speaker are not strictly demonstrated in all cases, they can still affect the performance of speaker recognition.

Zero Crossing Count (ZCC): Zero Crossing Count is used in speech processing. Voiced speech tends to have low ZCC.

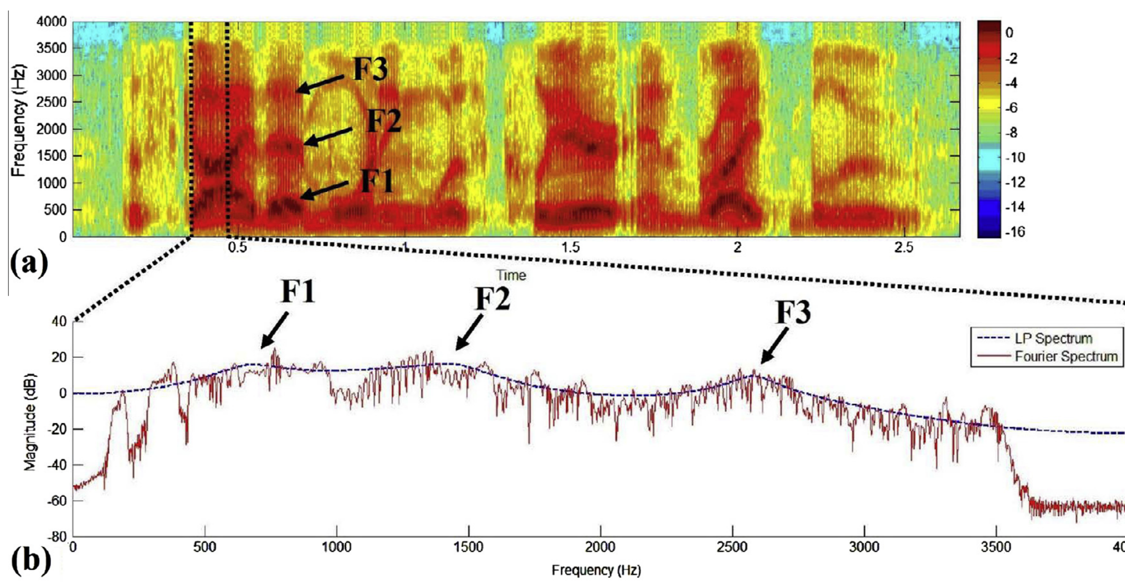

Figure 1 Spectrogram of the speech signal (a). Magnification of specific portions of the spectrogram (b) with main formant frequencies.

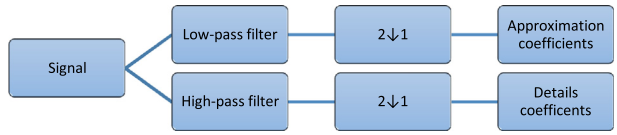

Figure 2 Single level DWT.

Fig. 1 shows the spectrogram of the speech signal. As seen, the speech signal has a narrow bandwidth (0 Hz–4 kHz). Formants (F1, F2, and F3), pitch (Muhammad, 2010), and fundamental frequencies of the speech signal can be observed. The majority of the speech energy is concentrated on the frequencies of less than 4 kHz. Furthermore, Linear Prediction (LP) spectrum and Fourier spectrum have been applied to show the frequencies of the formants in small segments of the speech spectrogram.

3. Decomposition techniques

All traditional frequency transforms, i.e., Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) and DWT, are utilized to decompose a signal into a standard or basic set. DWT is the most common or traditional technique for audio watermarking (Bhat et al., 2010). DWT represents an analog signal in the time–frequency domain with Sines and Cosines functions and the coefficients are calculated by using recursive algorithms, e.g., Mallat's pyramid algorithm (Mallat, 1989). The general procedure for single level DWT is illustrated in Fig. 2. DWT decomposes a signal to approximation coefficients (low frequency part) and detail coefficients (high frequency part) by applying low-pass and high-pass filters respectively. The results can be sent to another set of low pass and high pass filters for further decomposition. The subsequent filters are applied using low-pass or high-pass synthesis filters according to the mother wavelet. The results based on DWT become a powerful multi-resolution tool for the analysis of non-stationary signals with good time localization information (Mallat, 1989).

As the traditional frequency transforms are not always optimal, other numerical techniques are utilized for the decomposition of the signals into basic sets. Singular Value Decomposition (SVD), as a numerical technique, acts on diagonal matrices and breaks the signal into basic states optimally (Andrews and Patterson, 1976; Aldhaheri and Al-Saadi, 2004). For N*N matrix A, SVD on matrix A is presented in Eqs. (1) and (2). The diagonal entries of matrix S are singular values of matrix A, arranged in a decreasing order of $\sigma(i) > \sigma(i+1)$. The columns of matrices U and V are left singular vectors and right singular vectors of A respectively.

$$A = USV^T \tag{1}$$

$$\begin{bmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mn} \end{bmatrix} [S11 \quad \cdots \quad Snn] \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{r1} & \cdots & v_{rn} \end{bmatrix}^T \tag{2}$$

Normally a slight variation in the elements of S does not change the perception of an audience regarding the quality of the signal. This feature is basically utilized in audio watermarking. Therefore, the watermark information can be simply added to the singular values of the diagonal matrix S with no serious effects on the perceptibility or audibility of the signal. This audio watermarking algorithm is also robust (Al-Haj and Ahmad, 2010) and tolerates transpose attacks, filliping, scaling, and rotation (Biglieri and Yao, 1989; Al-Haj and Ahmad, 2010).

4. Watermarking

The audio and the speech signals have different features of production and perception. Watermarking on speech signal is discussed in this section (further discussions can be found in our previous (Nematollahi and Al-Haddad, 2013)).

Transform domain (Tempest, 1985; Schroeder et al., 1979; Djebbar et al., 2010) masks the watermarking information and embeds them into unimportant perceptual components of speech signal. In order to ensure that final signal is not affected by watermarking and is not distinguished by human, a Human Auditory System (HAS) Zwicker and Fastl, 1990 should be utilized for analysis of the result. The idea behind this technique is to hide the lower sound (maskee) from human audience by using the louder sound (masker). Amplitude coding performs frequency masking transformation for better intelligibility and inaudibility and capacity, while embedding watermark information. It calculates the wideband magnitude speech spectrum and finds the secure embedding area (between 7 kHz and 8 kHz) for embedding the watermark data.

The other technique called as Spread spectrum (SS) Khan, 1984; Cox et al., 1996; Cheng and Sorensen, 2005 inserts the watermarking information in form of hidden pseudorandom data throughout the frequency spectrum. The watermark information is then extracted at destination by calculating correlation between pseudo-random noise data and watermarked speech signal (Cox et al., 1996; Cheng and Sorensen, 2005).

Phase modulation as the other watermarking technique shifts the phase of speech and preserves the power spectrum without any changes. As the original and watermarked signals have the same power spectrum, the signal is not distorted. Phase modification and phase coding are two famous methods of phase modulation. Phase modification utilizes different bands for watermarking, while phase coding uses one frame for the whole watermark data. Embedding watermark information in the cepstrum coefficients of a log spectral domain is a robust and inaudible method of watermarking into the speech signal (Li and Yu, 2000; Gopalan, 2005; Gopalan, 2009).

Parametric modeling models the speech signal by using an all-pole filter (Autoregressive (AR)) and is utilized for watermarking. Linear predictive coding (LPC) or line spectrum pair (LSP) indirectly modifies or quantizes (AR) parameters for embedding the watermark (Gurijala and Deller, 2002; Hatada et al., 2002; Yan and Yin-Jing, 2011; Hofbauer and Kubin, 2006), then watermark data are embedded in the bit stream of the codec, e.g., ACELP (Geiser and vary, 2008), G.729 (Singh et al., 2009), G.711-PCMU (Aoki, 2008) and G.723.1 (Huang et al., 2011), for bypassing the speech compression attacks during or after speech compression.

The other famous method is called as the Patchwork method. It obtains the distance between two sets of speech signals for embedding the watermark data. This procedure is

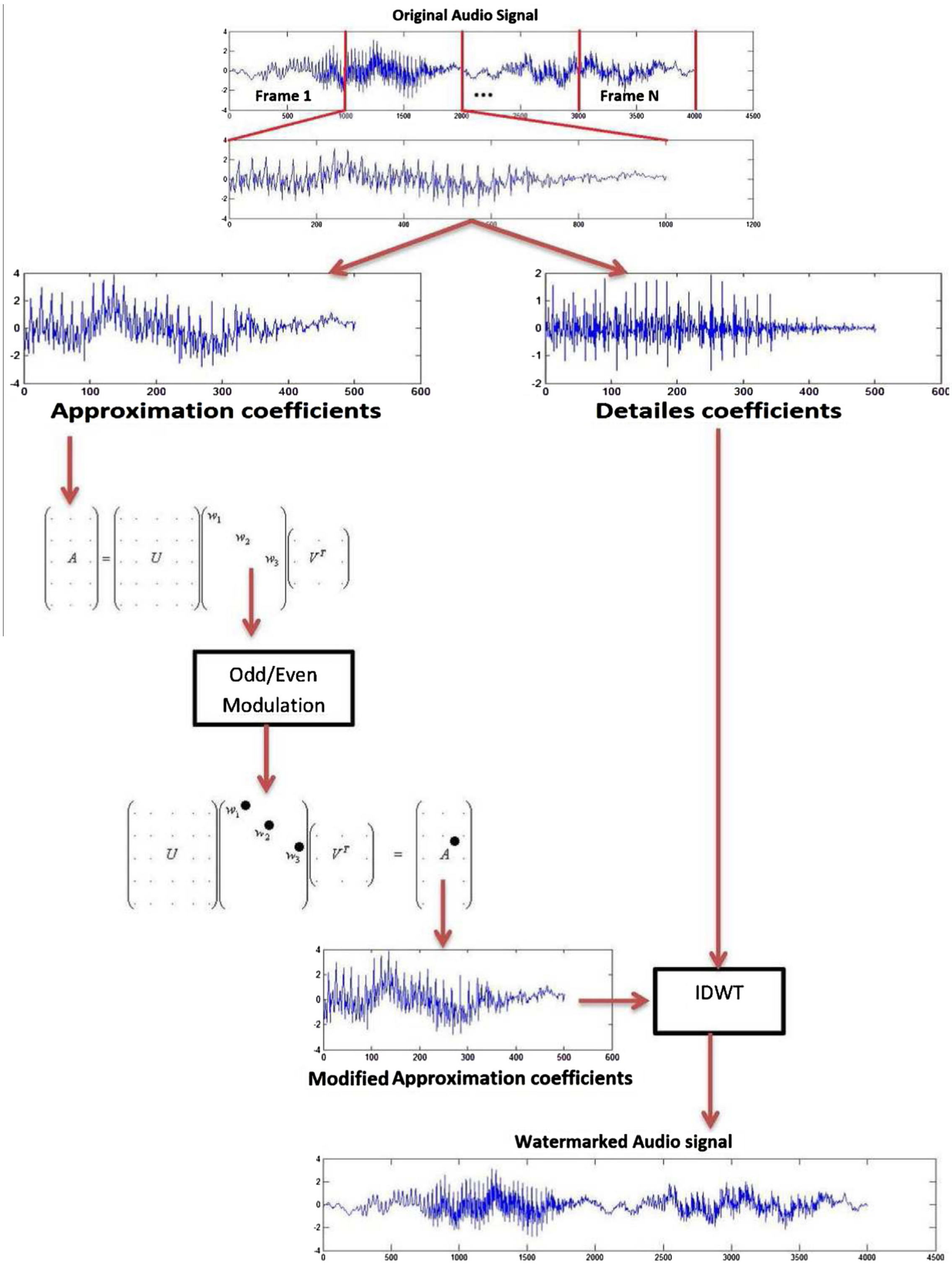

Figure 3 Block diagram of the proposed digital speech watermarking technique.

performed by using some statistical methods to change the variance, the energy, and the mean of the sets (Yan and Yin-Jing, 2011; Hofbauer and Kubin, 2006; Geiser and vary, 2008; Singh et al., 2009). The larger the distance obtained, the easier the watermarking embedding. However, the speech imperceptibility could be degraded.

Quantization Index Modulation (QIM) is a more popular technique of modulation in watermarking scheme that uses Costa scheme (Phadikar, 2013; Costa, May 1983). QIM has two steps. The first is modulating an index or a sequence of indices with the embedded information and the second is quantizing the original signal by the associated quantizer or sequence of quantizers. Applying QIM for watermarking can provide efficient tradeoff among capacity, imperceptibility, and robustness (Chen and Wornell, 2001). Furthermore, the blindness nature in QIM is significant in the process of watermark extraction. Therefore, applying QIM in the proposed technique in this study can offer improvements as compared to the technique used in Al-Haj and Ahmad (2010) in terms of robustness, time, blindness, and imperceptibility.

5. Proposed Digital speech watermarking technique

In this part, a new digital speech watermarking technique is designed based on quantization function which is proposed in Yan et al. (2005).

5.1. Embedding procedure

1. Segment the original speech signal into frames.
2. Apply first level DWT on each frame to calculate approximation and detail coefficients.
3. Apply 2-dimensional (2D) matrix formations by using approximation coefficients.
4. Apply SVD on the matrix to find right Eigen vector V, singular value S, and left Eigen vector U.
5. Use odd or even modulation function (a simple version of QIM) for each Eigen value based on Eq. (3) to embed the watermark bits. Odd or even modulation can preserve the histogram of the original signal. Furthermore, this version ensures that less modification is done with the signal's statistics because the modification occurs in a small neighborhood with less distortion injected into the host signal, as in Eq. (3):

$$\hat{Si} = \left[\frac{S_i + \Delta_i(1 - W_i)}{2\Delta_i}\right] \times 2\Delta_i + W_i\Delta_i \tag{3}$$

where $\hat{Si}$ is modified Eigen value, S_i is ith Eigen value, Δ_i is ith quantization step, and W_i is ith watermark bit.

6. Apply inverse SVD to compute the modified matrix.
7. Convert the modified matrix into modified approximation coefficients.
8. Apply inverse wavelet transform on detail coefficient and modified approximation coefficient to obtain the frame.
9. Reconstruct the signal based on all the modified frames to get the watermarked speech signal.

The overall embedding diagram is shown in Fig. 3.

5.2. Extraction procedure

1. Search for the start position of the watermark by using synchronization bits.
2. Segment the watermarked speech signal into frames.
3. Apply first level DWT on each frame to calculate approximation and detail coefficients.
4. Apply 2 dimensional (2D) matrix formations by using approximation coefficients.
5. Apply SVD on the matrix to find right Eigen vector V, singular value S, and left Eigen vector U.
6. Use inverse odd or even modulation function on each Eigen-value to extract the watermark bits as in Eq. (4):

$$\widehat{W_i} = \left[\frac{\hat{S}_i}{\Delta_i} + \frac{1}{2}\right]\%2 \tag{4}$$

where $\hat{S}_i$ is modified Eigen value, Δ_i is ith quantization step, W_i is ith watermark bit, and % corresponds to mod.

Figure 4 Binary watermark logo.

6. Experimental setup

In this study, an experiment using simulations was done to study the robustness, imperceptibility, and capacity of the proposed blind digital speech watermarking technique by using MATLAB. Two sets of speech data were used for validation and evaluation. First, a total of 6 signals from SQAM (EBU) were used, with indices of 49–54, sampling frequency of 44.1 kHz, and average time duration of around 22 s and 16 bits resolution. Second, 20 speech data were also used from ATCOSIM speech corpus (Hofbauer et al., 2008), with a sampling frequency of 32 kHz, a resolution of 12 bit, and an average duration of 3.8 s. These signals were selected based on good quality, different lengths, and publically available data. Fig. 4 shows the binary watermark logo, with the size of 22×31, which was used in this experiment.

Table 1 shows the various factors for the proposed blind digital speech watermarking. Bit Error Rate (BER) as in Eq. (5), Bit Per Second (BPS), and Signal to Noise Ratio (SNR) as in Eq. (6) were used to measure robustness, capacity, and imperceptibility respectively for the proposed watermarking technique:

$$\mathrm{BER}(w, \widehat{w}) = \frac{\text{Number of error bits}}{\text{Number of total bits}} = \frac{1}{N}\sum_{i=1}^{N} w(i) \otimes \widehat{w}(i) \tag{5}$$

$$\mathrm{SNR}(w, \widehat{w}) = 10log_{10}\frac{\sum_{i=1}^{N} w(i)^2}{\sum_{i=1}^{N}[w(i) - \widehat{w}(i)]^2}(\mathrm{dB}) \tag{6}$$

where w, $\widehat{w}$ are the original and watermarked signals, $\otimes$ is the exclusive OR (XOR) operator, and N is the length of the signal.

Table 1 shows how increasing the capacity could affect the imperceptibility. Embedding more watermarks into the original signal caused more distortion which affected speech imperceptibility.

Table 2 presents BER and SNR for different quantization steps. It can be seen that less quantization steps caused more imperceptibility in the watermarked signal.

Table 1 Robustness, capacity, and imperceptibility for $\Delta = 0.2$ without any attack.

BER (%)	bps	SNR (dB)
0	441	16.6111
0	220.5	20.2784
0	88.2	23.7455
0	44.1	26.4055

Table 2 Different quantization steps with 100 samples per frame for the proposed watermarking technique for frame length of 200.

Quantization step Δ	BER (%)	SNR (dB)
0.001	0	66.4058
0.01	0	46.4395
0.1	0	26.4830
0.2	0	20.2784
0.5	0	12.3320
1	0	6.2511

Table 3 BER for different attack for the proposed watermarking technique.

Attack	BER (%)
BPF (300–3400 Hz)	46.24
AWGN with 1 SNR per sample	57.09
AWGN with 5 SNR per sample	44.67
AWGN with 15 SNR per sample	16.62
Up-sampling with factor of 25	0

Table 3 shows BER under different attacks. As seen, Band Pass Filter (BPF) destroyed half of BER because many watermark bits were embedded inside the low frequency area. When SNR of Additive White Gaussian Noise (AWGN) was increased, BER was degraded.

7. Discussion

The proposed blind digital speech watermarking technique in this study is compared to Al-Haj's technique (Al-Haj and Ahmad, 2010). Al-Haj's work was selected for the assessment experiments in this paper due to the following: the need to apply similar techniques in signal watermarking with DWT and SVD, the ability to change the watermark intensity factors

Table 4 Ber and snr of proposed and Al-Haj's (Al-Haj and Ahmad, 2010) techniques under different quantization steps for frame length of 200 samples.

	Δ	BER (%)	SNR (dB)
Proposed	0.01	0	52.8794
	0.2	0	26.5456
	1	0	12.4413
Average		0	30.6221
Al-Haj and Ahmad (2010)	0.01	0.4868	7.4802
	0.2	0.0924	6.7205
	1	0.0936	3.4354
Average		0.2243	5.8787

(Δ and α) in the same manner, as well as the popularity and simplicity of the techniques. Fig. 5 shows the original and watermarked signals for the proposed technique. The extracted watermarked logos under various attacks are also shown. All the factors were assumed to be exactly the same for fair and valid comparison. As seen, the extracted watermark using the proposed technique seems to have better visibility than that in Al-Haj and Ahmad (2010).

Table 4 compares robustness and imperceptibility in terms of BER and SNR for different quantization steps for the proposed technique and Al- Haj's technique (Al-Haj and Ahmad, 2010). As seen, the average BER of the proposed technique is more than that in Al-Haj and Ahmad (2010) and shows robustness. SNR of the proposed technique is also more than that in Al-Haj and Ahmad (2010), showing more imperceptibility for the proposed technique.

Table 5 shows robustness in terms of BER for the proposed technique and that in Al-Haj and Ahmad (2010) when the watermarking technique was under different attacks. It is clear that the proposed technique has better robustness because the watermark bits are embedded into AC part of the speech which has more energy than DC part of the speech.

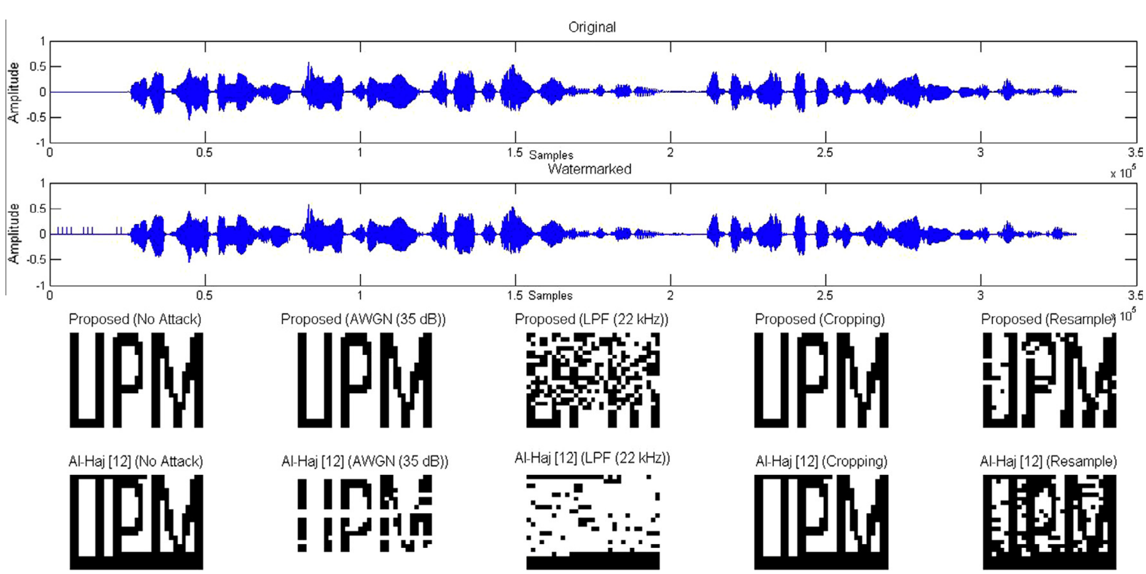

Figure 5 Original, watermarked, and extracted logos for proposed (Δ = 0.2) and Al-Haj's techniques (Al-Haj and Ahmad, 2010) with (α = 0.2).

Table 5 BER comparison for proposed and Al-Haj (Al-Haj and Ahmad, 2010) watermarking techniques.

Attack	Proposed	Al Haj (Al-Haj and Ahmad, 2010)
No Attack	0	0.0924
AWGN (35 dB)	0	0.1645
LPF (Cutoff 22 kHz)	0.3130	0.4648
300 samples of the watermarked signal set to zero randomly	0	0.0926
Resample to 8 kHz, then resample again to original sampling rate.	0.0804	0.1867

Function Name	Calls	Total Time	Self Time*	Total Time Plot (dark band = self time)
extract_AlHaj_DWT_SVD	30	236.443 s	130.227 s	
extract_DWT_SVD	30	65.259 s	47.959 s	
embed_AlHaj_DWT_SVD	6	38.962 s	16.668 s	
embed_DWT_SVD	6	15.756 s	9.044 s	

Figure 6 CPU time for proposed and Al-Haj's (Al-Haj and Ahmad, 2010) watermarking techniques.

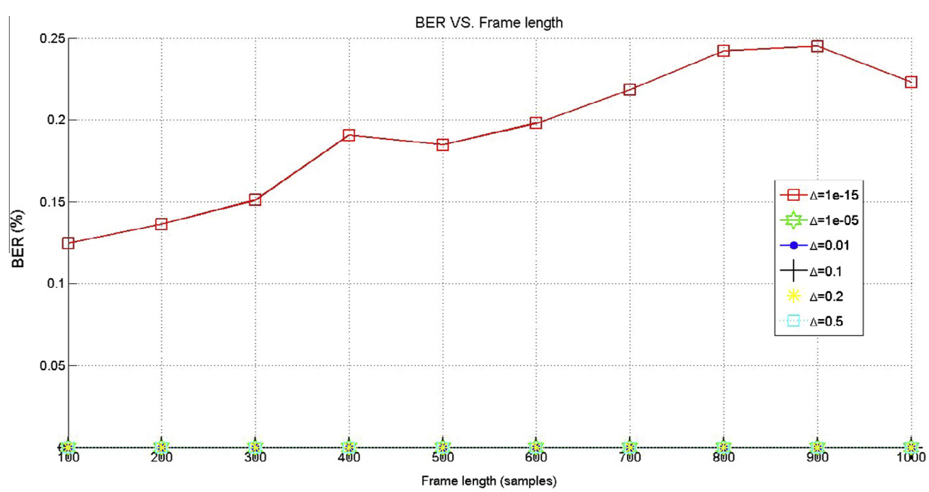

Figure 7 BER versus frame length for different quantization steps without any attack.

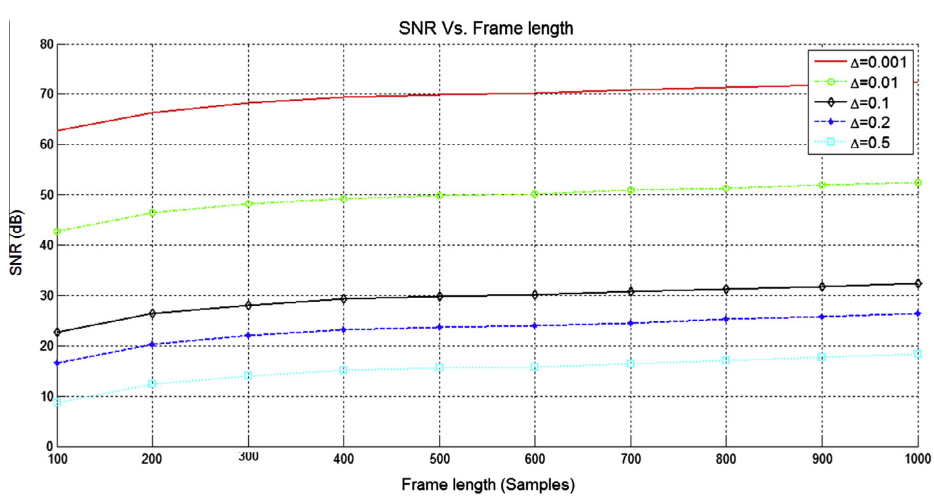

Figure 8 SNR versus frame length for different quantization steps.

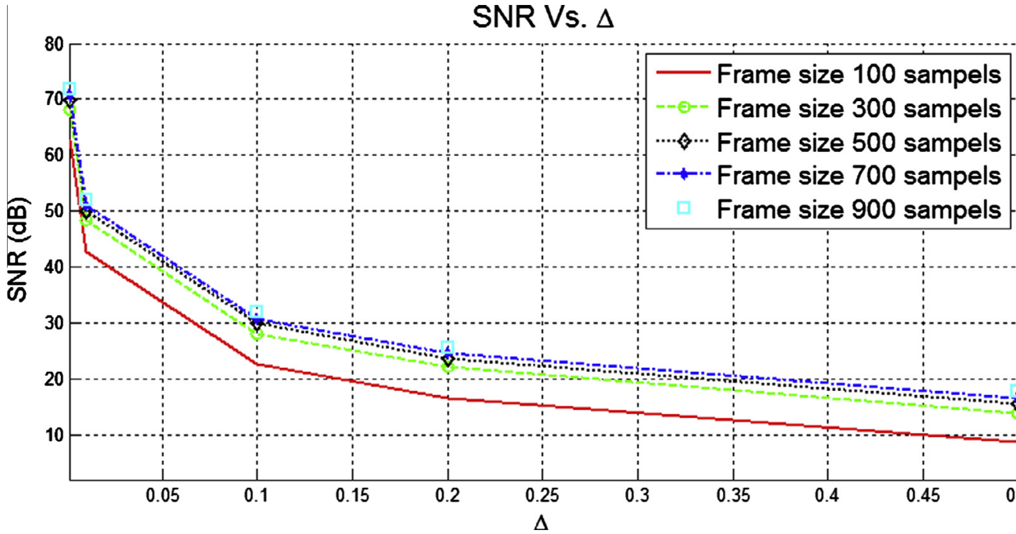

Figure 9 SNR versus quantization step for different frame lengths.

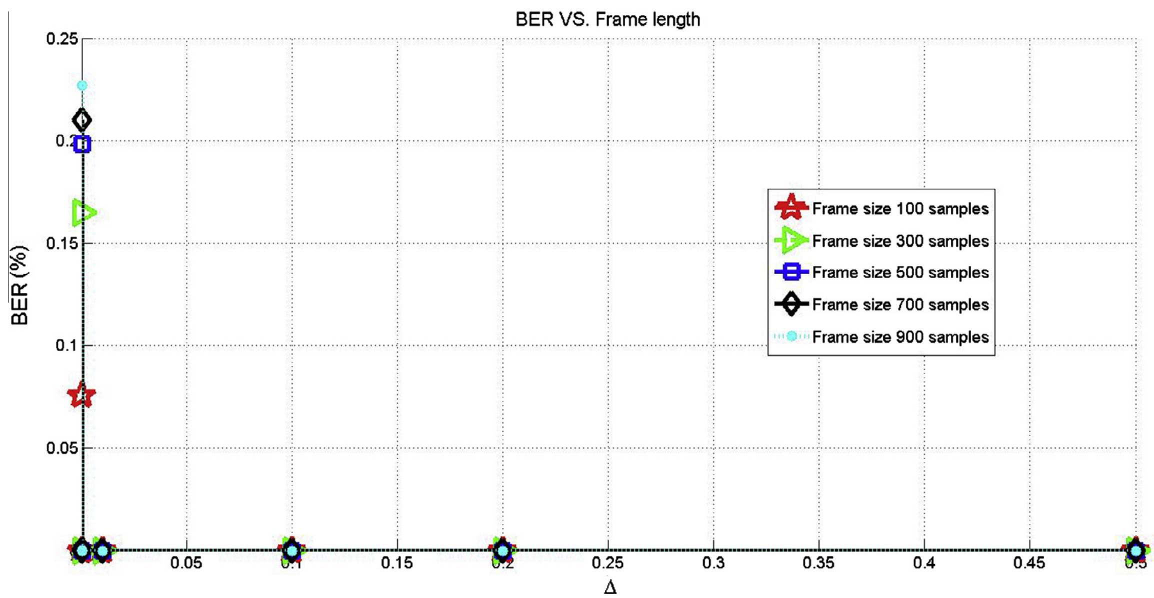

Figure 10 BER versus quantization step for different frame lengths.

For better and fairer comparison, the time was estimated by using MATLAB which is presented in Fig. 6. As seen, for embedding and the extraction functions, the proposed watermarking technique uses significantly less time than that in Al-Haj (Al-Haj and Ahmad, 2010).

It is clear that there is a direct relation between the quantization step, robustness, and imperceptibility. When the quantization step is increased, robustness is also increased but imperceptibility is decreased. Furthermore, selecting the frame length directly determines the capacity. When the frame length is decreased, the capacity is increased but imperceptibility and robustness are decreased. Therefore, the proposed blind digital speech watermark can be more robust against AWGN, filtering, resampling, and cropping attacks, when the frame length and quantization step are selected properly as they are application dependent.

Fig 7 shows BER in respect to the frame length. As seen, for different quantization steps, BERs were zero. The result may be due to the watermark bits being embedded in the high energy part of the speech signal which is not affected by frame length. Therefore, the watermarks can be extracted without any attack. However, to show that BER is not zero for all cases, very small quantization step was selected ($\Delta = 1 \times e^{-15}$). As seen, for very small quantization step, increasing the frame length increases BER. This situation may be due to quantizing of high energy Eigen-value (due to longer frame length) when small quantization step cannot embed and extract the watermark properly. The manipulation of an amount with high value and very small quantization step becomes useless. Therefore, there is a relationship between the quantization step and amount of Eigen value.

Fig. 8 shows SNR versus the frame length. As seen, whenever the size of the frame is increased, SNR is also increased. This situation is due to less distortion being induced by the watermark bits inside the watermarked speech signal. Further-

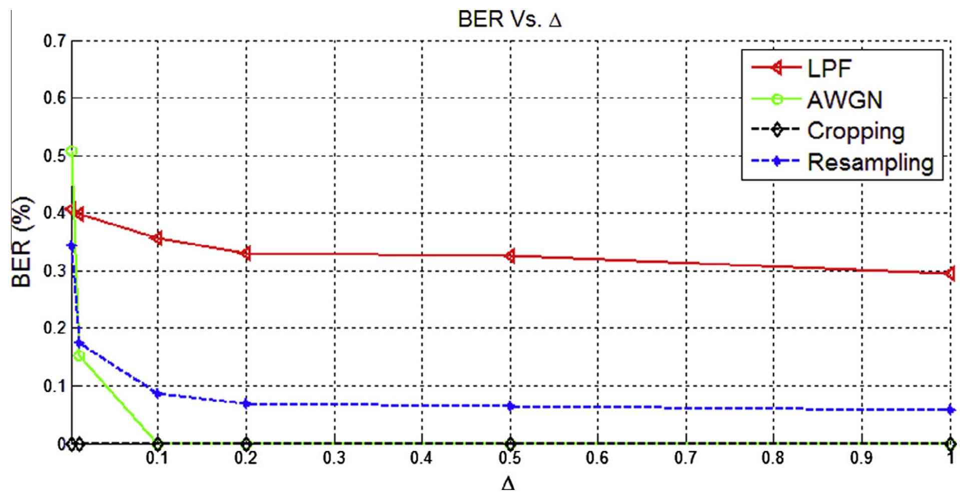

Figure 11 BER versus quantization step for frame length = 200.

more, selecting less quantization steps can improve the quality of the speech signal.

Fig. 9 shows SNR versus the quantization step. It is clear that increasing the quantization step can decrease SNR which affects the quality of the speech signal. Furthermore, all SNR values were more than 20 dB which is the minimum value to satisfy the criteria of the International Federation of the Phonographic Industry (IFPI) Katzenbeisser and Petitcolas, 2000. Fig. 10 shows robustness in terms of BER with respect to the quantization step for different frame lengths. It is obvious that the proposed watermarking technique has perfect ability in watermark extraction ability for various frame sizes with different quantization steps when no attack happens. Furthermore, it must be stated that BER is not zero for all cases. As seen, for very small quantization step, all BERs are not zero. It shows that very small quantization step is not effective for large frame size because such step cannot change Eigen-value properly but instead causes errors in the process of watermark extraction.

Fig. 11 shows BER in respect to the quantization step under different attacks. It is obvious that whenever the quantization step is increased, BER is decreased. Therefore, selecting a high quantization step can improve the robustness but degrade the quality. Fig. 11 also shows that selecting a high quantization step does not improve the robustness using the low-pass filter. Much energy of the wavelet is removed by using the low-pass filter which directly affects the first Eigen-value which carries the watermark.

8. Conclusion and future works

In this paper, a new blind digital speech watermarking technique is proposed by using quantization of Eigen-value in DWT domain. The experimental setup shows that this technique is robust against different attacks when the quantization step is high and the frame length selected is long. Otherwise, the watermark bits are degraded. Furthermore, embedding the watermark with detail coefficients can improve the imperceptibility but as a consequence, it degrades the robustness. Studying new and robust transforms can be the future works in this field. Using adaptive quantization steps and synchronization techniques can further improve the process of watermark extraction.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments to improve the quality of this paper.

References

Phadikar, Amit., 2013. Multibit quantization index modulation: a high-rate robust data-hiding method. J. King. Saud Univ-Comput. Inf. Sci. 25 (2), 163–171.

Vivekananda Bhat, K., Sengupta, Indranil, Das, Abhijit, 2010. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. Digit. Signal. Process 20 (6), 1547–1558.

Kent, R.D., Read, C., 2002. Acoustic Analysis of Speech, 2nd ed. Delmar, New York.

Hofbauer, K., 2009. Speech watermarking and air traffic control (Diss. Ph. D. dissertation). Graz University of Technology, Graz, Austria.

Faundez-Zanuy, Marcos., Hagmüller, Martin., Kubin, Gernot., 2006. Speaker verification security improvement by means of speech watermarking. Speech Commun. 48 (12), 1608–1619.

Faundez-Zanuy, Marcos., Hagmüller, Martin., Kubin, Gernot., 2007. Speaker identification security improvement by means of speech watermarking. Pattern Recogn. 40 (11), 3027–3034.

Nematollahi, Mohammad Ali, Al-Haddad, S.A.R., 2013. An overview of digital speech watermarking. Int. J. Speech Technol. 78, 1–18.

Al-Shoshan, Abdullah, I., 2006. Speech and music classification and separation: a review. J. King Saud Univ.

Mallat, S., 1989. A theory for multi-resolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11 (7), 674–693.

Andrews, H., Patterson, C., 1976. Singular Value Decomposition (SVD) image coding. IEEE Trans. Commun. 42 (4), 425–432.

Biglieri, E., Yao, K., 1989. Some properties of singular value decomposition and their applications to digital signal processing. Signal. Process. 18 (3), 277–289.

Al-Haj, Ali, Ahmad, M., 2010. Digital audio watermarking based on the discrete wavelets transform and singular value decomposition. Eur. J. Sci. Res 39 (1), 6–21.

Tempest, W. (Ed.), 1985. The Noise Handbook. Academic Press.

Schroeder, M.R., Atal, B.S., Hall, J.L., 1979. Optimizing digital speech coders by exploiting masking properties of the human ear. J. Acoustical Soc. Am. 66 (6), 164.

Djebbar, F., Abed-Maraim, K., Guerchi, D., Hamam, H., 2010. Energy based text-in speech spectrum hiding using speech mask properties. ICSRA Mai.

Zwicker, E., Fastl, H., 1990. Psychoacoustics. Springer Verlag, Berlin.

Khan, K., 1984. Cryptology and the origins of spread spectrum. IEEE Spectr. 21, 70–80.

Cox, I.J., Kilian, J., Leigton, F.T., Shamoon, T., 1996. Secure spread spectrum watermarking for multimedia. IEEE Trans. Image Process. 6, 1673–1687.

Cheng, Q., Sorensen, S., 2005. Spread spectrum signaling for speech watermarking. United States Patent, No. US6892175B1.

Li, X., Yu, HH., 2000. Transparent and robust audio data hiding in cepstrum domain, Proc. IEEE International Conference on Multimedia and Expo, (ICME 2000) (New York, USA, 2000).

Gopalan, K., 2005. Audio Steganography by Cepstrum Modification. Proc. of the IEEE 2005 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05),(Philadelphia, USA, March 2005)

Gopalan, K., 2009 .A unified audio and image steganography by spectrum modification, IEEE International Conference on Industrial Technology (ICIT'09), (Gippsland, Australia, 10-13 Feb. 2009), pp.1–5.

Gurijala, A., Deller, J.R., Seadle Jr., M.S., Hansen, J.H.L., 2002. Speech watermarking through parametric modeling. In: Proceedings of international conference on spoken language processing (ICSLP).

Hatada, M., Sakai, T., Komatsu, N., Yamazaki, Y., 2002. Digital watermarking based on process of speech production. Proc. SPIE 4861, 258–267.

Bin Yan, Yin-Jing Guo., 2011. Speech authentication by semi-fragile speech watermarking utilizing analysis by synthesis and spectral distortion optimization. Springer Science + Business Media, LLC 2011.

Hofbauer, K., Kubin, G., 2006. High-rate data embedding in unvoiced speech. In: Proc. Int. Conf. Spoken Language Processing (INTERSPEECH), Pittsburgh, PY, USA, September 2006, pp. 241–244.

Geiser, B., Vary, P., 2008. High rate data hiding in ACELP speech codes. In: Proceedings of ICASSP 2008, pp. 4005–4007.

Singh, J., Garg, P., De, A.N., 2009. A combined watermarking and encryption algorithm for secure VoIP. Inf. Secur. J. Global Perspect. 18, 99–105.

Aoki, N., 2008. A Technique of Lossless Steganography for G.711 Telephony Speech, International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'08), Harbin, China, 2008, pp.608–611

Huang, Y.F., Tang, S., Yuan, J., 2011. Steganography in inactive frames of VoIP streams encoded by source codec. IEEE Trans. Inf. Forensics Secur. 6 (2), 296–306.

Costa, M.H.M., May 1983. Writing on dirty paper. IEEE Trans. Inf. Theory 29 (3), 439–441.

Chen, B., Wornell, Gregory W., 2001. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans. Inf. Theory 47 (4), 1423–1443.

Yan, B. et al, 2005. Speech Authentication by Semi-Fragile Watermarking. Knowledge-Based Intelligent Information and Engineering Systems. Springer, Berlin Heidelberg.

EBU, SQAM–Sound Quality Assessment Material, <http://tech.ebu.ch/publications/sqamcd>.

Hofbauer, Konrad, Petrik, Stefan, Hering, Horst, 2008. The ATCOSIM corpus of non-prompted clean air traffic control speech. LREC.

Katzenbeisser, S., Petitcolas, F.A.P., 2000. Information Hiding Techniques for Steganography and Digital Watermarking. Artech House Norwood Mass, USA.

Aldhaheri, Rabah W., Al-Saadi, Fuad E., 2004. Robust text-independent speaker recognition with short utterance in noisy environment using SVD as a matching measure. J. King. Saud Univ.-Comput. Inf. Sci. 17, 25–44.

Muhammad, Ghulam, 2010. Noise-robust pitch detection using autocorrelation function with enhancements. J. King. Saud Univ.-Comput. Inf. Sci. 22, 13–28.