Research Journal of Applied Sciences, Engineering and Technology 11(7): 746-758, 2015
DOI: 10.19026/rjaset.11.2037
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: April 14, 2015 Accepted: May 10, 2015 Published: November 05, 2015

Research Article

A Method to Improve Voice Quality in Speech Watermarking at the Receiver Using the Adaptive Filters

[1]Shervin Shokri, [1]Mahamod Ismail, [1]Nasharuddin Zainal and [2]Vahideh Hashemzadeh
[1]Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan, Malaysia
[2]Payam-E Noor University of Mashhad, Khorasan Razavi, Iran

Abstract: This study presents a new subject in audio-speech watermarking, voice quality, and then proposes a watermarking scheme to improve the voice quality at the receiver. In recent years, we are witnessing rapid development in applications of watermarking in digital communication systems. Voice quality is considered as a main topic to design the systems and device that are used in speech telecommunication (such as PSTN, VoIP, or mobile communication). Therefore these products inevitably should be having the acceptable voice quality to be successful in the market. Because the signal transmission through a communication channels has always been related with noise, noise removal techniques are very useful. Given this idea, this study presents a speech watermarking scheme with acceptable speech output quality for Human Hearing Systems (HHS) at the receiver. To improve voice quality at the receiver, a type of adaptive filter called Least Mean Square (LMS) is used in the proposed scheme. The proposed scheme is simulated and then the output voice quality of the system is evaluated by Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ) and segmental Signal-to-Noise Ratio (SNR_{seg}), as subjective and objectives methods. The experimental results for voice quality at the receiver have been compared with those of previous works. The measured scores for voice quality by the MOS and PESQ methods are 4.30 and 3.13, respectively and we also witnessed a significant increase in voice quality by SNR_{seg} measurement.

Keywords: Adaptive filters, speech watermarking, spread spectrum, voice quality

INTRODUCTION

In the early 1990s, watermarking as a modern technique was commenced for the hiding of data. Watermarking is the process of embedding private data, such as text or numbers, in a digital form within a media signal (video, image and audio). In the beginning, only images and videos were utilized, but audio signals began to be considered by researchers in 2000. Signal segmentation shows that audio and speech are narrow band signals and this feature has caused the watermarking process for audio and speech to be much more difficult than that for video and image. The left of Fig. 1 shows the objectives of the watermarking system that have previously been classified. These objectives are split into three categories, un-detectability (inaudibility), capacity (data rate) and robustness, as the vertices of a triangle (Cvejic and Seppanen, 2004; Chun-Shien, 2004). Signal quality is one of the important factors that has not been fully considered in audio-speech watermarking. Previous works, such as Hagmüller *et al*. (2004), Faundez-Zanuy *et al*. (2007), Hering *et al*. (2003), Coumou and Sharma (2008), Hofbauer and Hering (2007), Cheng and Sorensen (2005), Akhaee *et al*. (2010), Kondo (2012) and Zhang *et al*. (2012) and others, have shown that although these works are outstanding and have been done very well with regard to the watermark data and imperceptibility of the watermarked signal, they have not performed any special process to improve the voice quality of the received audio-speech signal at the output of the channel. Thus, the lack of attention in this area has motivated us to take a special look at the quality as an important topic in audio-speech watermarking and we have become convinced that the current triangle can be turned into a quadrilateral by adding a new topic, i.e., quality (Fig. 1).

To avoid confusion between the concepts of quality and acoustical invisibility or statistical un-detectability, we should define the meanings of these terms. Un-detectability is defined as an embedding process without sensitivity for the Human Sensory System (HSS), but quality is the estimation of the received signal from the effects channel. Figure 2 shows the area for inaudibility and voice quality measure.

Channel noise is a large challenge in digital signal transmission and communication systems. Always detrimental effects of noise on the emitted signal push designers to find a way to get rid or compensate for the damaging effects caused by noise. One of these ways is

Corresponding Author: Shervin Shokri, Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan, Malaysia

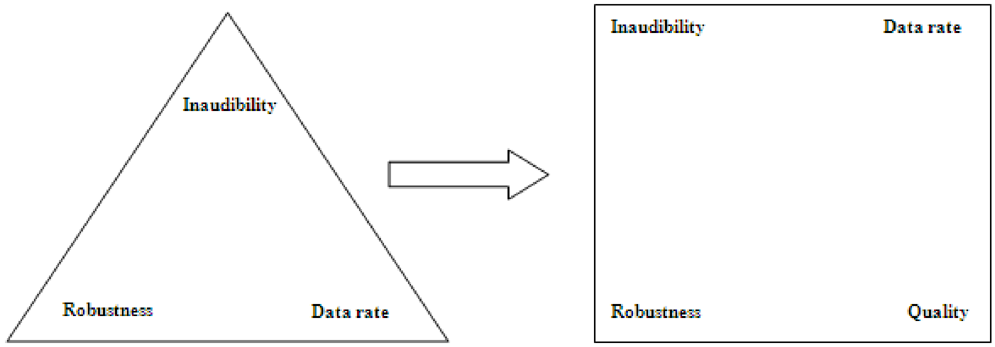


Fig. 1: A new topic for audio-speech watermarking

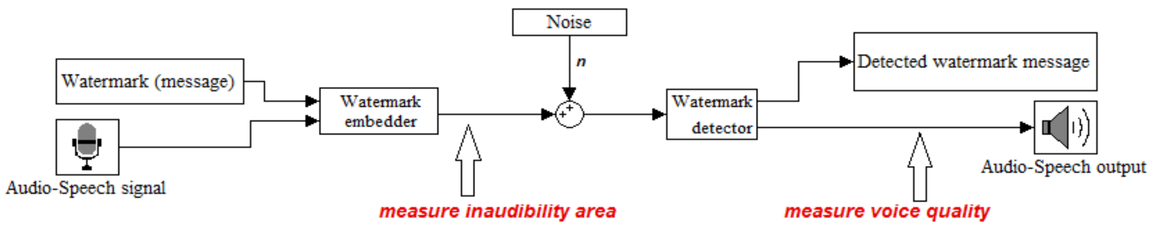


Fig. 2: Areas for measure inaudibility and voice quality

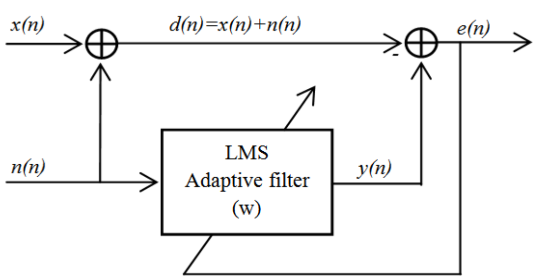


Fig. 3: Noise cancelation block diagram based on LMS algorithm

to build strong detectors, which, in some cases, results in them being very complex and expensive without any commercial value.

Because the watermarked data in audio-speech watermarking should be in an imperceptible form and the power of the watermark follows the audio-speech power in the embedding process, a limitation in adjusting the Signal-to Watermark Ratio (SWR) will appear before sending the signal to the channel. In this case, if we increase the power of the watermark signal the watermark is come robust against the channel noise but we lose the imperceptibility and also if we reduce the watermark power we lose the reliability. As noted above, it will push us to design a very high-power detection, which makes the system expensive.

In addition, the channel noise will make the received audio-speech signal agonizing for human hearing. To solve this problem, Adaptive Noise Cancellation (ANC) is one of the major real-time methods available to remove the noise of a signal. These methods are based on adaptive filters and are widely used in many communication systems to feed the detector with a clean signal of the noise. Unique advantages of these filters are in simply design, no need the special configurations and grate power of them to remove the noise that needless us from the complex and expensive detectors. Figure 3 illustrates an ANC that uses the LMS filter (Poularikas and Ramadan, 2006). The advantages listed above gave us the courage to use these filters in the proposed scheme as our contribution in this study.

Our previous work (Shokri *et al*., 2012) proposed an algorithm for speech watermarking based on Spread Spectrum (SS) and then the voice quality was measured by the MOS method.

A new strategy for informed embedding has been defined to decrease the host interference in the watermark signal. Malvar and Florêncio (2003) showed an improvement in watermark robustness with interaction between the host and watermark signals and because their method is based on the SS technology, they called it Improved Spread Spectrum (ISS).

This study proposes a speech watermarking scheme based on ISS and uses the LMS filter as an adaptive filter to improve the voice quality in the received signal. The embedding process is conducted according to the proposed speech watermarking schemes and voice quality is measured by MOS, PESQ and SNR_{seg} as subjective and objective tests of the output signal. This study mainly focuses on improving the received voice quality by utilizing noise removal techniques and the output results show a dramatic improvement in voice quality.

SUBJECTIVE AND OBJECTIVE MEASURES

Voice quality is normally measured subjectively, where the most popular method is the Mean Opinion Score (MOS). This involves a panel of several listeners, usually placed in a soundproof room, listening to the audio recording under evaluation. They will then rate this according to the scale shown in Table 1. The scale of 1 to 5 was standardized by the International Telecommunications Union (ITU; P.800), a United Nations body responsible for telecommunications standardization (Rix, 2004; McLoughlin, 2009).

To estimate the MOS value, the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) was presented P.862, called Perceptual Evaluation of Speech Quality (PESQ), as a standard in February 2001 (Kondo, 2012). PESQ is classified as an objective method of estimating voice quality. It works by comparing the original and the degraded speech and giving a score that ranges from 0.5 (bad quality) to 4.5 (excellent quality) (Upadhyay and Karmakar, 2013; Hänsler and Schmidt, 2008).

In recent years, several objective quality algorithms have been developed. The Signal-to-Noise Ratio (SNR) is one of the famous methods which are widely used for objective measures. It is mathematically simple to calculate but requires both distorted and undistorted (clean) audio-speech samples. The SNR can be calculated as follows:

$$SNR = 10 \log_{10} \frac{\sum_{n-1}^{N} x_n^2}{\sum_{n-1}^{N} \{x(n) - \tilde{x}(n)\}^2} [dB] \tag{1}$$

where, N is the frame length (number of samples).

This classical definition of the SNR is not known to be well related to speech quality for a wide range of distortions. Thus, several variations on the classical SNR exist that exhibit much higher correlations with

Table 1: Applications in each class

Quality of the speech	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

subjective quality. It was observed that the classical SNR does not correlate well with voice quality because, even though voice is not a stationary signal, the SNR averages the ratio over the entire signal. Therefore, a way to address this issue is for the SNR to be calculated in short frames and subsequently averaged (McLoughlin, 2009; Kondo, 2012). This measure is called the segmental SNR and can be defined as:

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Lm}^{Lm+L-1} x^2(n)}{\sum_{n=Lm}^{Lm+L-1} \{x(n) - \tilde{x}(n)\}^2} [dB] \tag{2}$$

where, L is the frame length (number of samples) and M the number of frames in the signal ($N = ML$). The frame length is normally set between 20 and 30 ms.

PROPOSED SPEECH WATERMARKING SCHEME

The proposed scheme is based on the SS technology and the simple basic frequency masking approach. Figure 4 shows the block diagram for the embedding process (encoder) as an emitter. The emitter can be divided into three major sections. First, the error-control coding is employed by the channel coding to increase the reliability of the system. The next level involves spreading the watermark signal over the available frequency band. Finally, the watermark is embedded in the speech signal by utilizing perceptual methods. The ISS and Linear Prediction Coefficient (LPC) filters are utilized for spectral shaping of the watermark spectrum. The shaped watermark is then

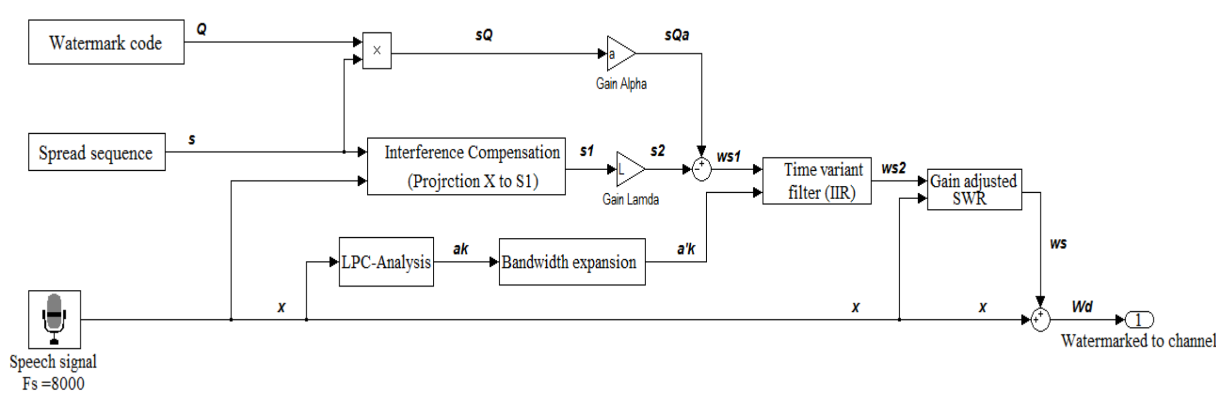


Fig. 4: Embedding blocks diagram

embedded in the speech signal. The resulting signal, which is called the watermarked signal, will be transferred to the channel after being converted into analog.

The process of watermark embedding will be explained in continue:

In the watermark code box, the payload data have been encoded with the BCH code. BCH codes are a type of cyclic linear block code that is permitted to encode a large selection of block data and is widely used in channel coding. The BCH code checks the error numbers at the receiver side. If the number of errors is within the feasible range, then the errors will be corrected; otherwise, the errors will only be detected (Shokri *et al*., 2012). In the next step, a synchronization sequence is added to the watermark signal for system synchronization. The message coding output (Q) is called the watermark code.

SS is one of the most popular techniques in watermarking. SS can generate the conditions to embed watermarks in any frequency or time domain (Davarynejad *et al*., 2010). With this technique, a spread watermark signal $v(n)$ is achieved by spreading the bits of Q_m over a set of samples (N_b) of sequence $s(n)$ (as the vector $s = [s(0),s(1),\ldots\ldots s(N_b-1)]^T$) . The $s(n)$ signal is presented by the vector s, which is made by the PN sequence (in $\{-1,+1\}$). The $v(n)$ signal is given by:

$$v(n) = Q(n)s(n) \tag{3}$$

$$v(n) = \sum_{m=0}^{M-1} a_m s(n - mN_b) \tag{4}$$

where, the a_m symbol in {-1, 1} is given by $a_m = 2k-1$. Therefore, the vector direction will be adjusted by the a_m values (Shokri *et al*., 2013):

$$v_m = \begin{cases} +s & if \quad a_m = +1 \\ -s & if \quad a_m = -1 \end{cases} \tag{5}$$

Embedding the watermark in the speech signal with maximum energy and minimum perceptual distortion are the main aims of data embedding. The watermarked signal is achieved by embedding the spread watermark sequence $v(n)$ into the speech signal as mentioned above. Adding the watermark in the speech signal without any terms will create a large interference in the watermark signal; therefore, the embedding can be better controlled by utilizing ISS techniques in terms of temporal energy (Hering *et al*., 2003). This can be achieved by projecting the speech signal over the spreading watermark (Malvar and Florêncio, 2003; Hagmüller *et al*., 2004; Zhang *et al*., 2012). The linear form of ISS embedding can be formulated as follows:

$$w[n] = x[n] + \mu(\tilde{x}, Q)s[n] \tag{6}$$

where,

$$\tilde{x} \triangleq \frac{\langle x, s\rangle}{\|s\|} \tag{7}$$

$\tilde{x}$ is the projection of vector x on vector s:

$$\langle x, s\rangle = \frac{1}{N}\sum_{I=0}^{N-1} x_i s_i \text{ and } \|x\| \triangleq \langle x, x\rangle \tag{8}$$

The function $\mu(\tilde{x}, Q)$ is a linear function of the speech signal (x). Vector s has the N signal sample and the bit rate is 1/N/bits/sample. The following equations were derived based on Fig. 4:

$$s_1 = \tilde{x}s \tag{9}$$

$$s_2 = \lambda s_1 \tag{10}$$

$$w_{s1} = \alpha Qs - \lambda\tilde{x}_k s = (\alpha Q - \lambda\tilde{x})s \tag{11}$$

The parameters α and λ are used to control the distortion level and removal of the carrier distortion on the detection statistics. In the traditional SS, these parameters are set to $\alpha = 1$ and λ = 1. To decrease the perceptual distortion, a Linear Prediction Coefficient (LPC) is utilized to estimate the spectrum of the speech signal by vocal formant coefficients (a_k) of the speech signal $x(n)$ (Fig. 5 and 6). The spectral of the spread watermark $v(n)$ is made like similar to the speech signal by passing through a time variant filter (IIR), which is created from coefficients a_k (Fig. 4) (Kotnik *et al*., 2009; Zölzer, 2011; Ramamurthy and Spanias, 2010). The vocal transfer function and LPC transfer function are defined as follows:

$$H(z) = \frac{g}{A(z)} = \frac{g}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{12}$$

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \tag{13}$$

The LPC order p is expressed as:

$$p = 2 + \frac{F_s}{1000}, \; F_s = \text{sample frequency} \tag{14}$$

Thus, for telephone frequency sampling (8000 kHz), the LPC order is 10.

The bandwidth expansion technique is utilized to avoid interference between the watermark and speech signal. In this technique, the filter coefficients (a_k) are adjusted by the γ factor. This factor can create a small gap between two signals to protect the watermark signal

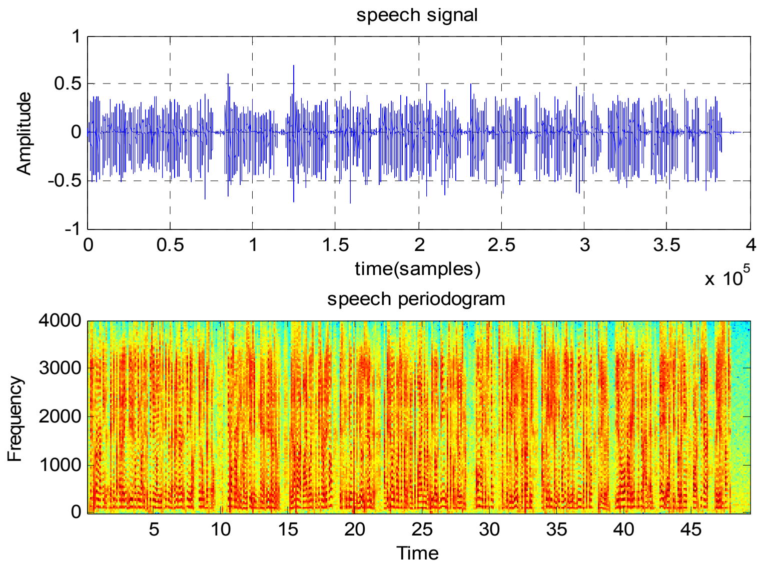


Fig. 5: Speech signal and spectrogram

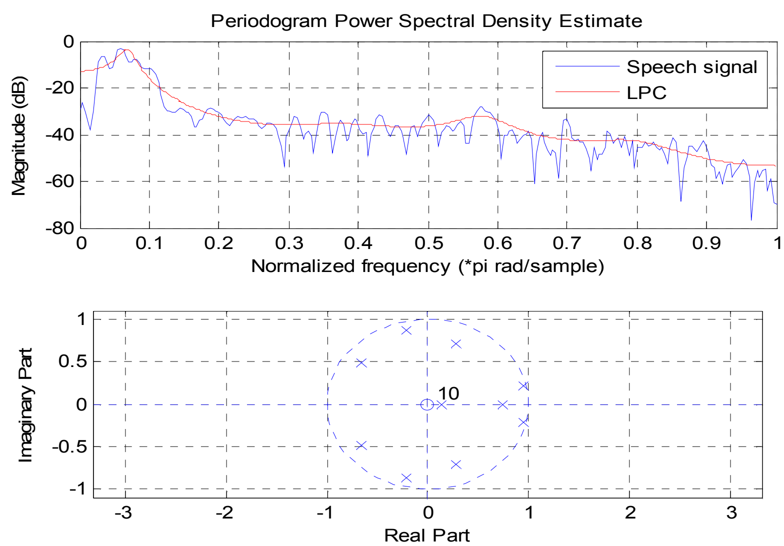


Fig. 6: Top: Speech signal and LPC cover; bottom: z plane of the polls

from speech formants (Faundez-Zanuy *et al*., 2007; Ramamurthy and Spanias, 2010):

$$a'_k = a_k \gamma^k \quad 0 \le k \le order \tag{15}$$

The best value for the γ factor usually ranges from 0.90 to 0.97. In this case, by adjusting the γ factor to 0.90, a bandwidth expansion appears at the spectral peaks by moving all the poles to the center of the unit circle (Fig. 7 and 8) (Ramamurthy and Spanias, 2010).

In the final step to embed the watermark in speech, a variable gain (λ_G) is applied to obtain the desired Signal-to-Watermark Ratio (SWR) (Deshpande and Prabhu, 2009):

$$w_s = (\alpha Q - \lambda \widetilde{X} a'_k) s \lambda_G \tag{16}$$

After spreading and shaping, the watermark signal can be embedded in speech signal by simple addition:

$$w[n] = x[n] + w_s[n] \tag{17}$$

Figure 9 shows the group delay between a_k and a_k' coefficients during the LPC filter. This delay is typically considered in the embedding process. The simulations show that the delay is reduced after bandwidth expansion. This delay is shown in Eq. (18) (Faundez-Zanuy *et al*., 2006):

$$w[n] = x[n - M] + w_s[n] \tag{18}$$

Here, M is a delay that is practically set to 100 ms (Faundez-Zanuy *et al*., 2006, 2007). Figure 10 and 11

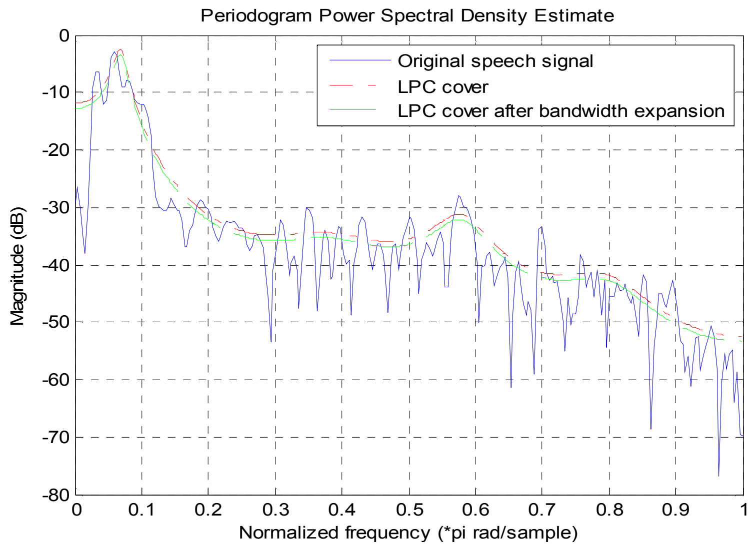


Fig. 7: Bandwidth expansion

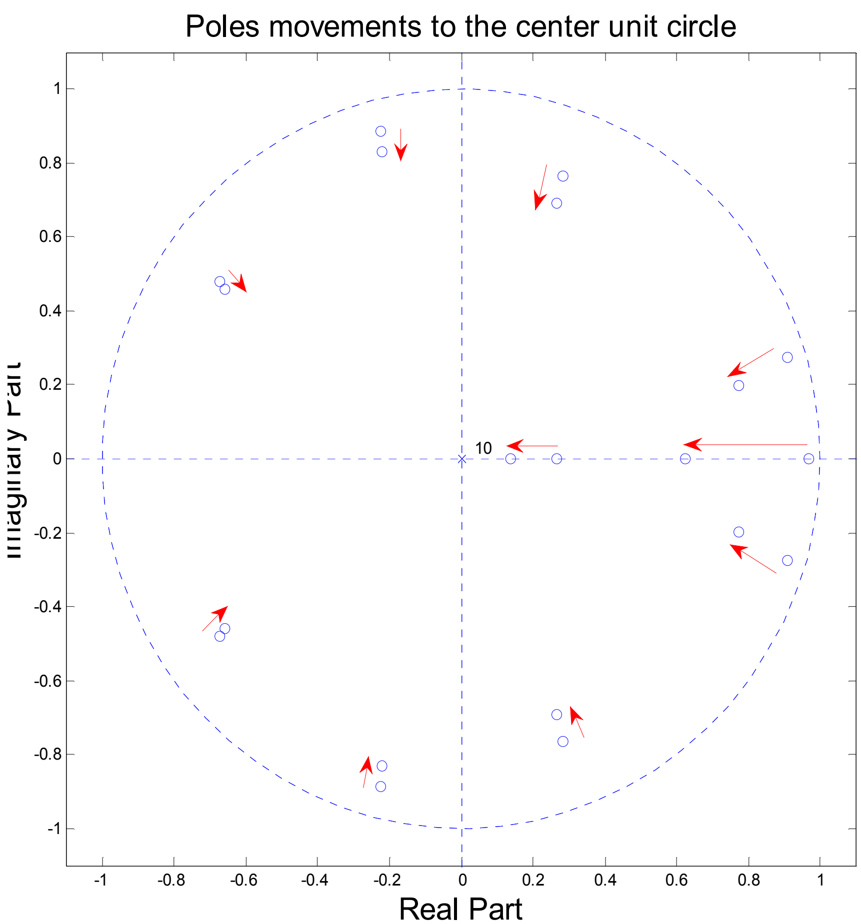


Fig. 8: Z plane of the polls in bandwidth expansion by moving the poles to the center of the unit circle

show the simulation results of the embedding process. Figure 10 shows the Power Spectral Density (PSD) of the watermark, speech and watermarked signal. Figure 11 shows the waveforms of the original signal and the watermarked signals with and without spectral shaping. The real dynamic channel model is not considered in this study. Therefore, the static channel model was simulated by Adding White Gaussian Noise (AWGN)

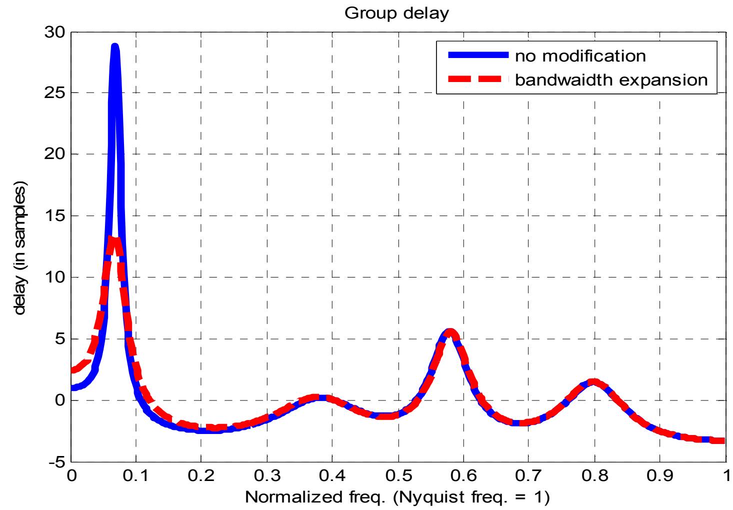


Fig. 9: Group delay for speech signal with and without bandwidth expansion

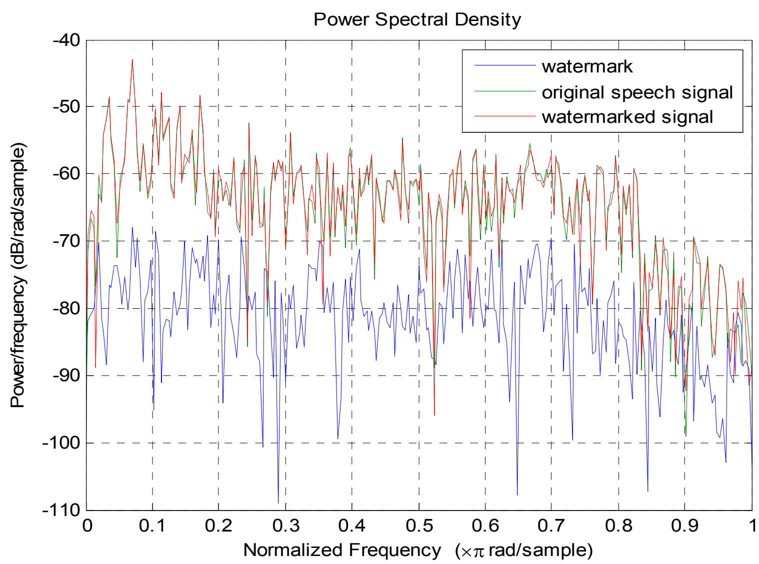


Fig. 10: Power spectrum density for watermark, speech and watermarked signal

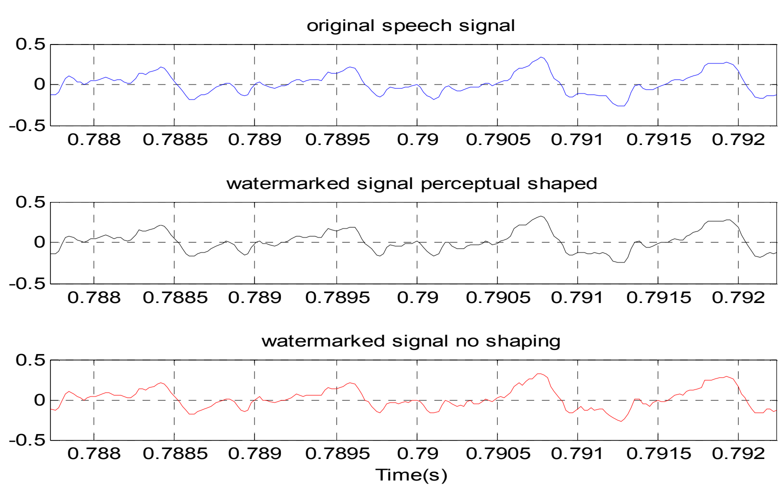


Fig. 11: Original Speech, watermarked shaped and watermarked without shaping signal

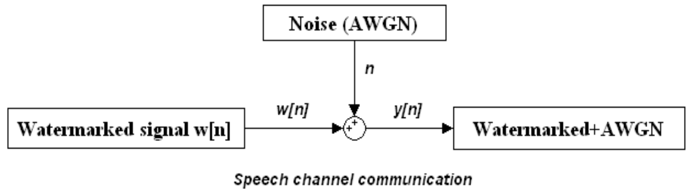


Fig. 12: Communication channel

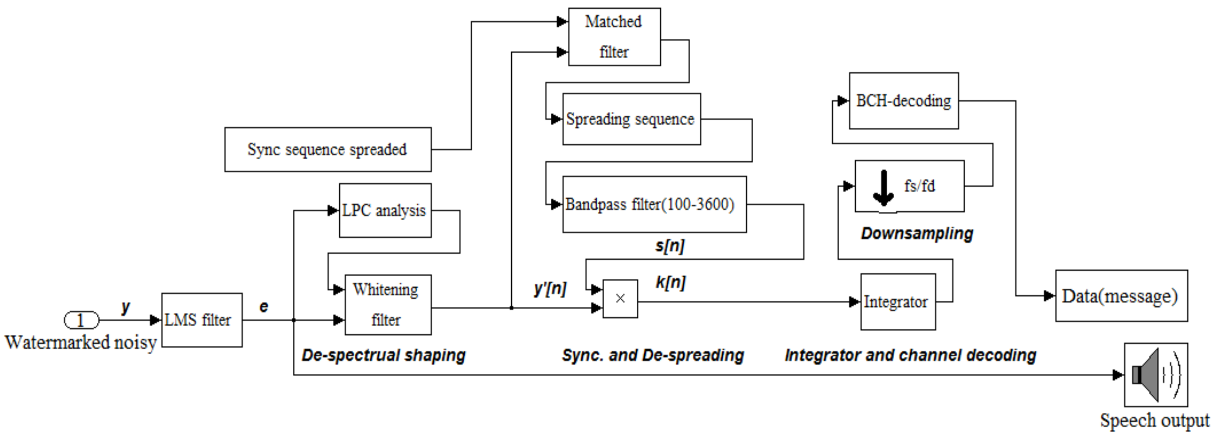


Fig. 13: Extraction (decoder) blocks diagram

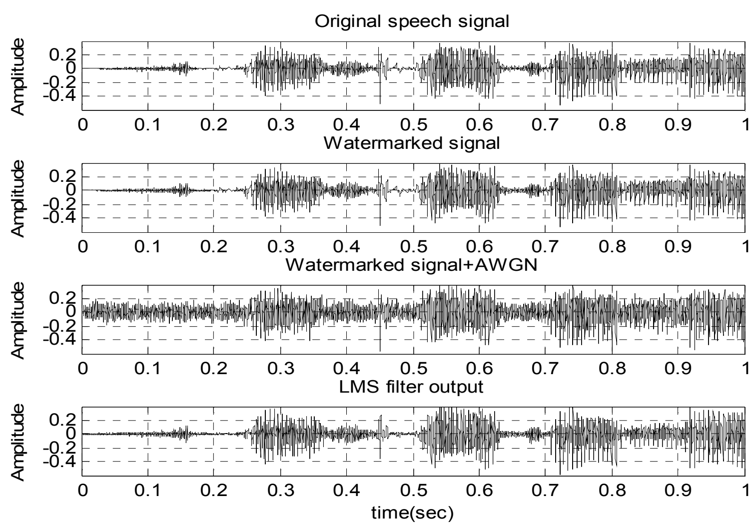


Fig. 14: Comparison between the original, watermarked, watermarked+AWGN in SNR = 20d and output signal from LMS filter

to the transmitted watermarked signal. The watermarked signal (w_d) passes through the AWGN channel and the noise (n) will act as environmental noise or hacker attacks. A schematic block diagram of the noise's effect on the watermarked signal is shown in Fig. 12:

$$y[n] = w[n]+n[n] \tag{19}$$

Figure 13 shows the process of extracting the embedded watermark from the speech signal at the receiver side (decoder).

Our contribution to improve the voice quality is included at the receiver input. The received signal (y_n) is needed to remove the noise, so an LMS filter is placed in the first step of the receiver. The main objective of the noise cancellation is the estimation of the noise signal by subtracting it from the noisy signal and hence obtaining the noise-free signal (Jebastine and Rani, 2012; Rahman *et al*., 2011). Figure 14 shows the process of embedding and noise removing the watermark signal:

$$e[n] = d[n]-y[n] \tag{20}$$

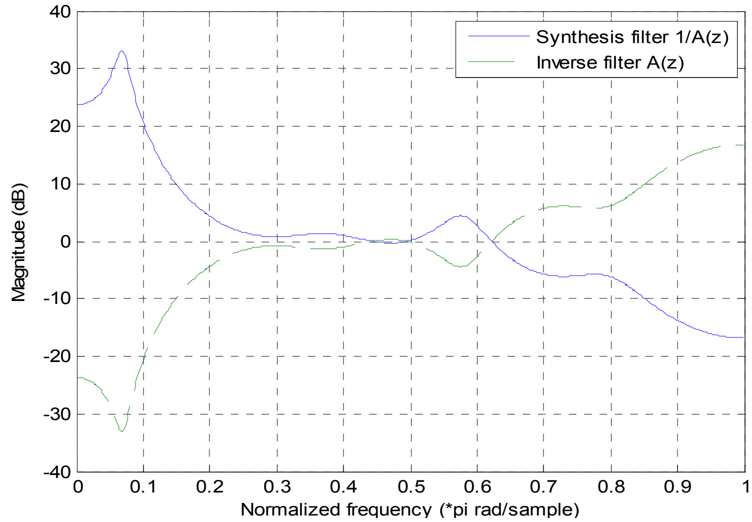


Fig. 15: Whitening filter output

Figure 14 shows a large similarity between the LMS output and the watermarked signal, which is fed the result of the detection process with a clean signal as much as possible. At the next step, a whitening filter is used to undo the spectral shaping from the incoming signal at the receiver side. The whitening filter is an inverse filter $A(z)$, which calculates the prediction error. The incoming signal $y[n]$ is passed through the LPC filter again to extract the coefficients. $A(z)$ then utilizes the coefficients to undo the spectral shaping. Note that for the bandwidth expansion, the zeros are also returned to their place by the whitening filter. After inverse LPC filtering, the speech signal becomes the periodic pulse. Of course, the spectral shape of the signal at the receiver side is not equal to the original speech signal, but the spectral shape of the signal is expected to be similar. The whitening filter output is shown in Fig. 15 (Shokri *et al*., 2012).

Because the receiver does not know the source of the message, a blind detector is used at the receiver side. For synchronization, a special synchronization sequence that is known by the transmitter and receiver is added to the payload data (in the encoder). De-spreading will start when the impulse response from the reverse of the spread synchronization sequence is sensed by the matched filter (Hagmüller *et al*., 2004).

De-spreading is performed after synchronization to yield the payload data by multiplying the signal by the spread sequence. The PN sequence $s(n)$ is passed through the BPF (100-3400) filter and then multiplied by the $y'[n]$ signal:

$$k_r[n] = y'[n]s[n] \tag{21}$$

With consideration to the above channel noise model, sufficient statistics are calculated as:

$$k_r = \langle y', s \rangle / \|s\| = \alpha Q + (1-\lambda)\tilde{x} + n_r,\ n_r \cong 0 \tag{22}$$

where, n_r is a very low value of the noise that remains in the signal. Therefore, for $\lambda \approx 1$in Eq. (22), more influence of $\tilde{x}$ is reduced or removed from k_r. A simple integrator is used in the detector by knowing the data bits and the length of them. Bit detection is conducted by integrating over the period of one data bit and quantizing the result to 1 or -1 (Ferrera *et al*., 2010; Savoj and Razavi, 1999; Bo *et al*., 2001; Hagmüller *et al*., 2004). For one received data bit, k' is:

$$k' = sign\left(\sum_{n=0}^{bit-length} k_r[n]_i \right) \tag{23}$$

where, i is the current bit interval. At this step, the process to reduce the sampling rate to the binary symbol rate is performed via down-sampling **[x↓]**. Finally, the BCH decoder is used for error correction as much as possible (Shokri *et al*., 2012).

SIMULATION AND EXPERIMENT RESULTS

In this section, practical and analytical tests are considered to evaluate the voice quality. For test preparation, Wavepad Sound Editor Masters edition version 5.33 is used for speech recording. The software is set to an 8 kHz sample rate in mono channel. A Philips SHM3100U/97 In-ear Earphone microphone is also used for voice recording. The technical data for the microphone is as follows:

- Wired
- In-the-ear
- Ear Bud Design
- 50 mW of Max Power Input
- Built-in Microphone
- 12 Hz-20000 Hz Headset Frequency Response
- 80 Hz-15000 Hz Microphone Frequency Response
- 106 dB/mW Headset Sensitivity at 1 kHz

The pre-emphasis for the speech signal is done by a first-order filter whose transfer function is $H(z)=1-0.95z^{-1}$. A 30 ms hamming window in the LPC filter is used with the 2/3 (20 ms) overlapped frames and a message signal (watermark) is converted into binary with a length of 520 bits. During the process of receiving the watermarked signal, an AWGN at 20 dB is used as channel noise. The MOS technique, as a subjective technique, is used to measure voice quality in the proposed scheme. Forty participants in two different groups participated in the practical test and their averaged votes are utilized to measure the quality of the processed signal. The perceptual quality of the signal is estimated in five aspects: the audio-speech signal input (original), the watermarked signal in terms of shaping or not shaping and the watermarked signal at the receiver with or without using the LMS filter. The experimental results are shown in Table 2 and 3.

Table 2: MOS score for group 1

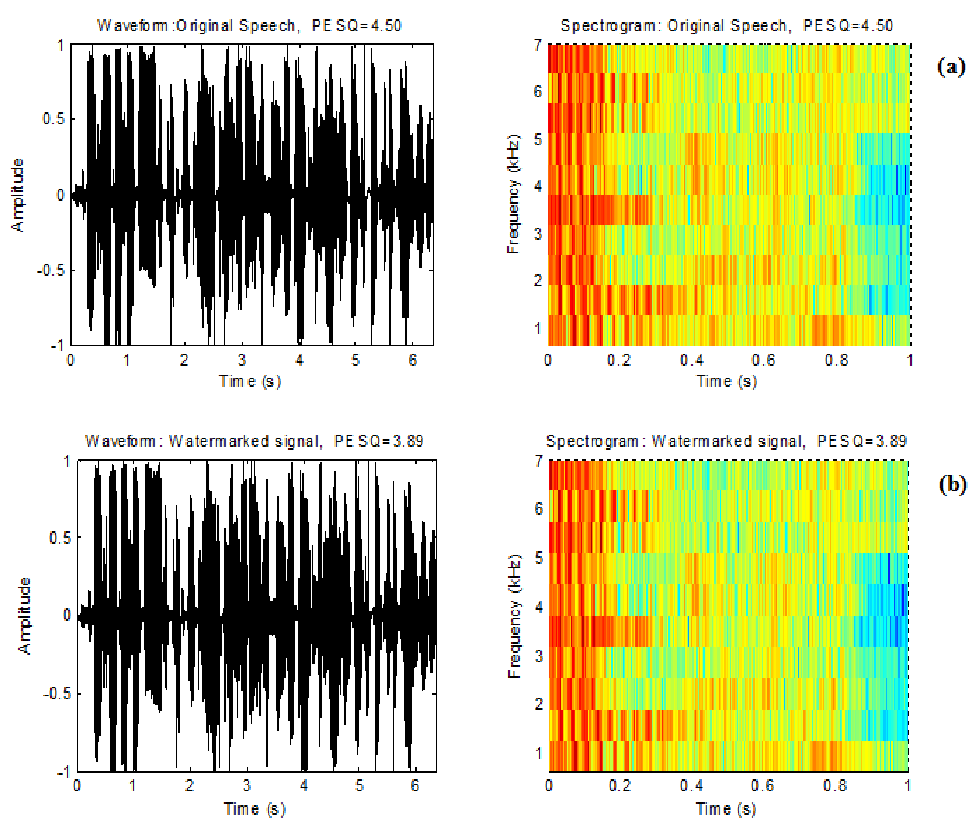
MOS Score	1	2	3	4	5	Mean
Original signal	-	-	-	-	20	5
Watermarked no shaping	-	-	3	17	-	3.85
Watermarked shaping	-	-	-	7	13	4.65
Receiving Signal	-	5	15	-	-	2.75
LMS filterOutput	-	-	-	13	7	4.35

Table 3: MOS score for group 2

MOS Score	1	2	3	4	5	Mean
Original signal	-	-	-	-	20	5
Watermarked no shaping	-	-	5	15	-	3.75
Watermarked shaping	-	-	-	6	14	4.7
Receiving Signal	-	5	15	-	-	2.75
LMS filterOutput	-	-	-	15	5	4.25

Table 4: Performance comparisons with method 1 and method 2 (SNR = 20 dB)

	Proposed method		Method 1		Method 2	
Method measure method	MOS	PESQ	MOS	PESQ	MOS	PESQ
Original signal	5	4.50	5	4.50	5	4.50
Watermarked signal	4.65	3.89	4.55	3.82	4.60	3.53
Output signal	4.30	3.13	2.65	1.81	2.70	1.83



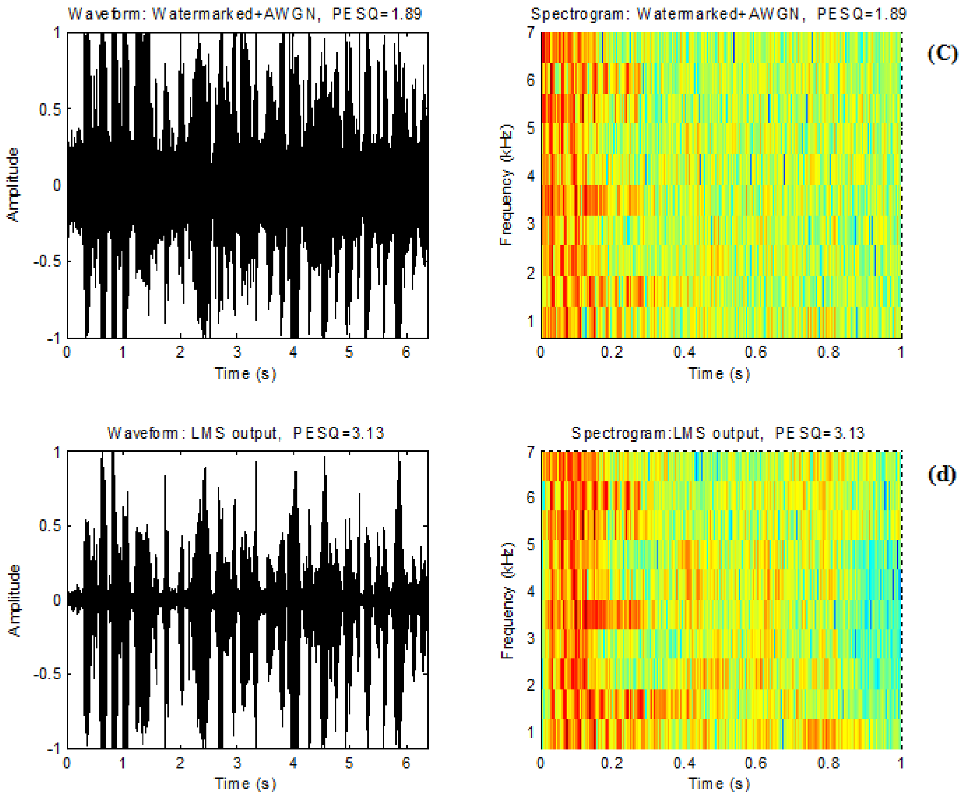


Fig. 16: PESQ measurment technique; (a) The original speech signal, (b) The watermarked signal (this measure is for inaudibility), (c) The watermarked signal and channal noise (W_d+AWGN), (d) Watermarked or the signal output at the receiver for human hearing (this measure is for voie quality)

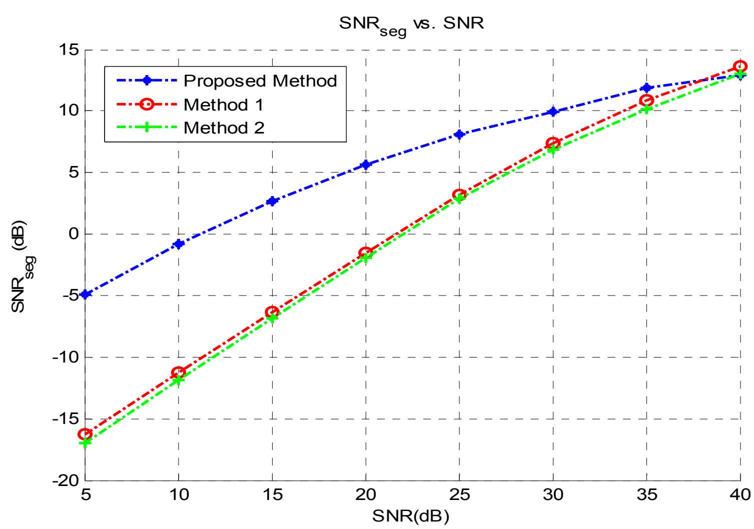


Fig. 17: Objective quality measures obtained with SNR_{seg}

To estimate the MOS measurement, PESQ is used as an objective method. The results of PESQ tests carried out to measure voice quality show an improvement at the receiver. The output results are shown in Fig. 16. To compare the proposed scheme with other works, the performance of our scheme is compared with methods in Hagmüller *et al.* (2004) as method 1 and Zhang *et al.* (2012) as method 2. The performance of the methods and the corresponding measurements of the imperceptibility and quality of the signal are listed in Table 4.

The output results show us that the imperceptibility of the watermarked signal for all methods is in an acceptable range, but the voice quality in method 1 and method 2 is troublesome for human hearing. Looking at the output results, this problem has largely been resolved via the proposed method.

The SNR_{seg} in the range of 0-40 dB input SNR for an AWGN channel is investigated as an objective method to measure voice quality. The instrumental audio-speech quality obtained with the different input SNR ranges is plotted in Fig. 17 and this figure shows improved voice quality at the receiver in comparison with method 1and method 2.

CONCLUSION

In this study, we have introduced a new topic for audio-speech watermarking, quality, and we have also proposed a speech watermarking scheme for improving the voice quality at the receiver side. In this study, the quality of an original speech signal and a watermarked signal was evaluated in the process of watermarking. To improve the voice quality, adaptive filters have been introduced as a noise cancelation filter at the receiver. The proposed scheme has used LMS filter as an adaptive filter to remove or reduce the noise effects in the received signal. A subjective and two objective methods have been investigated to measure voice quality in the proposed scheme. As a subjective method, the P. 800 (MOS) standardization by the International Telecommunications Union (ITU) was used as a practical technique to measure voice quality. The practical test was performed by employing 40 participants organized into two different groups. The experimental results show that the inaudibility of this algorithm after watermark shaping is near excellent (≈4.65), which is extraordinary. The average score at the receiver side shows that the MOS score before using the LMS filter is close to fair (≈2.70), but, after using the LMS filter, the average score improved to between good and excellent (≈4.30). The PESQ technique (ITU-T recommendation P. 862) as an objective measure of voice quality was employed to estimate the MOS scores and the output results are largely confirmed by the MOS scores. Objective output results in SNR_{seg} also show an improved speech quality in a different range of SNR. We strongly believe that adaptive filters in audio-speech watermarking can have more interesting roles; for example, they can increase the reliability of the system. Therefore, for further work, the system reliability will be considered in relation to the functions of the adaptive filters. We also hope that the presentation of this study will inspire other researchers to come up with even better methods for speech watermarking.

REFERENCES

Akhaee, M.A., N.K. Kalantari and F. Marvasti, 2010. Robust audio and speech watermarking using Gaussian and Laplacian modeling. Signal Process., 90(8): 2487-2497.

Bo, X., L. Shen and W. Chang, 2001. Sign Correlation Detector for Blind Image Watermarking in the DCT Domain. In: Shum, H.Y., M. Liao and S.F. Chang (Eds.), Advances in Multimedia Information Processing-PCM. LNCS 2195, Springer-Verlag, Berlin, London, pp: 780-787.

Cheng, Q. and J.S. Sorensen, 2005. Spread spectrum signaling for speech watermarking. Patent Number: 6892175.

Chun-Shien, L., 2004. Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property. IGI Global eBook, Idea Group Publishing, Hershey PA, ISBN: 1591401933, pp: 350.

Coumou, D.J. and G. Sharma, 2008. Insertion, deletion codes with feature-based embedding: A new paradigm for watermark synchronization with applications to speech watermarking. IEEE T. Inf. Foren. Sec., 3(2): 153-165.

Cvejic, N. and T. Seppanen, 2004. Channel capacity of high bit rate audio data hiding algorithms in diverse transform domains. Proceeding of the IEEE International Symposium on Communications and Information Technology (ISCIT, 2004), pp: 84-88.

Davarynejad, M., C.W. Ahn, J. Vrancken, J. Van den Berg and C.A. Coello, 2010. Evolutionary hidden information detection by granulation-based fitness approximation. Appl. Soft Comput., 10(3): 719-729.

Deshpande, A. and K.M.M. Prabhu, 2009. A substitution-by-interpolation algorithm for watermarking audio. Signal Process., 89(2): 218-225.

Faundez-Zanuy, M., M. Hagmüller and G. Kubin, 2006. Speaker verification security improvement by means of speech watermarking. Speech Commun., 48(12): 1608-1619.

Faundez-Zanuy, M., M. Hagmüller and G. Kubin, 2007. Speaker identification security improvement by means of speech watermarking. Pattern Recogn., 40(11): 3027-3034.

Ferrera, M., Y. Park, L. Razzari, B.E. Little, S.T. Chu, R. Morandotti, D.J. Moss and J. Azaña, 2010. On-chip CMOS-compatible all-optical integrator. Nat. Commun., 1: 29.

Hagmüller, M., H. Horst, A. Kröpfl and G. Kubin, 2004. Speech watermarking for air traffic control. Watermark, 8(9): 10.

Hänsler, E. and G. Schmidt, 2008. Speech and Audio Processing in Adverse Environments. Springer, Berlin, Heidelberg, ISBN: 354070602X, pp: 736.

Hering, H., M. Hagmuller and G. Kubin, 2003. Safety and security increase for air traffic management through unnoticeable watermark aircraft identification tag transmitted with the VHF voice communication. Proceeding of the 22nd Digital Avionics Systems Conference (DASC). Indianapolis, IN, USA.

Hofbauer, K. and H. Hering, 2007. Noise robust speech watermarking with bit synchronisation for the aeronautical radio. In: Furon, T. *et al*. (Eds.), IH, 2007. LNCS 4567, Springer-Verlag, Berlin, Heidelberg, pp: 252-266.

Jebastine, J. and B.S. Rani, 2012. Design and implementation of noise free audio speech signal using fast block least mean square algorithm. Signal Image Process. Int. J., 3(3): 39-53.

Kondo, K., 2012. Subjective Quality Measurement of Speech. Springer-Verlag, Berlin, Heidelberg, pp: 153, ISBN: 3642275052.

Kotnik, B., Z. Mezgec, J. Svečko and A. Chowdhury, 2009. Data transmission over GSM voice channel using digital modulation technique based on autoregressive modeling of speech production. Digit. Signal Process., 19(4): 612-627.

Malvar, H.S. and D.A.F. Florêncio, 2003. Improved spread spectrum: A new modulation technique for robust watermarking. IEEE T. Signal Proces., 51(4): 898-905.

McLoughlin, I., 2009. Applied Speech and Audio Processing: With Matlab Examples. Cambridge University Press, Cambridge, New York, ISBN: 0521519543, pp: 206.

Poularikas, A.D. and Z.M. Ramadan, 2006. Adaptive Filtering Primer with MATLAB. CRC Press, Boca Raton, FL, ISBN: 0849370434, pp: 240.

Rahman, M.Z.U., S.K. Mohedden, B.V. Rama Mohana Rao, Y. Jaipal Reddy and G.V.S. Karthik, 2011. Filtering non-stationary noise in speech signals using computationally efficient unbiased and normalized algorithm. Int. J. Comput. Sci. Eng., 3(3): 1106-1113.

Ramamurthy, K.N. and A.S. Spanias, 2010. MATLAB® software for the code excited linear prediction algorithm: The federal standard-1016. Syn. Lectures Algorithm. Software Eng., 2(1): 1-109.

Rix, A.W., 2004. Perceptual speech quality assessment-a review. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP, 2004), pp: 1056-1059.

Savoj, J. and B. Razavi, 1999. A CMOS interface circuit for detection of 1.2 Gb/s RZ data. Proceeding of the IEEE International Solid-State Circuits Conference, Digest of Technical Papers (ISSCC, 1999). San Francisco, CA, pp: 278-279.

Shokri, S., M. Ismail and N. Zainal, 2012. Voice quality in speech watermarking using spread spectrum technique. Proceeding of the International Conference on Computer and Communication Engineering (ICCCE, 2012). Kuala Lumpur, pp: 169-173.

Shokri, S., M. Ismail, N. Zainal and A. Shokri, 2013. Error probability in spread spectrum (SS) audio watermarking. Proceeding of the IEEE International Conference on Space Science and Communication (IconSpace). Melaka, pp: 169-173.

Upadhyay, N. and A. Karmakar, 2013. A multi-band speech enhancement algorithm exploiting iterative processing for enhancement of single channel speech. J. Signal Inform. Process., 4(2): 197-211.

Zhang, P., S.Z. Xu and H.Z. Yang, 2012. Robust audio watermarking based on extended improved spread spectrum with perceptual masking. Int. J. Fuzzy Syst., 14(2): 289-295.

Zölzer, U., 2011. DAFX: Digital Audio Effects. 2nd Edn., Wiley Chichester, West Sussex, England.