Multimed Tools Appl
DOI 10.1007/s11042-011-0861-7

Speech authentication by semi-fragile speech watermarking utilizing analysis by synthesis and spectral distortion optimization

Bin Yan · Yin-Jing Guo

Abstract This paper proposes an improved semi-fragile speech watermarking scheme by quantization of linear prediction (LP) parameters, i.e., the inverse sine (IS) parameters. The spectral distortion due to watermark embedding is controlled to meet the 'transparency' criterion in speech coding. A modified bit allocation algorithm combined with watermarking is developed to determine the quantization step so that the 'transparency' requirement is satisfied. Due to the statistical nature, the LP coefficients estimated from the watermarked speech signal are different from the watermarked LP coefficients even in the absence of attacks. This effect is the cause of increase in decoding error and minimum authentication length. To tackle this problem, an Analysis by Synthesis (AbS) scheme is developed to reduce the difference between the estimated LP coefficients and the watermarked ones. The watermark detection threshold and minimum authentication length are then derived according to the probability of error and the signal to noise ratio (SNR) requirements. Experimental results show that the proposed AbS based method can effectively reduce the difference between the watermarked IS parameter and the extracted IS parameter when there is no attacks. In addition, the modified bit allocation algorithm can automatically find the appropriate quantization step used in the odd-even modulation so that the transparency requirement is satisfied.

Keywords Speech authentication · Fragile watermark · Analysis by Synthesis (AbS) · Bit allocation · Spectral distortion

B. Yan (✉) · Y.-J. Guo
Department of Communication Engineering, School of Information and Electrical Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266510, People's Republic of China
e-mail: yanbinhit@hotmail.com

1 Introduction

In recent years, watermarking has proved itself as a powerful complementary tool to multimedia security, a realm that is governed by cryptography in the past. Starting from the initial copyright related applications, the application areas have been expanded to secure communication [9, 19, 32] and content authentication. More recently, more efforts have been exerted on authentication of speech signals [14, 40, 43].

The requirements for speech content authentication arise in several applications, especially in the open network environment. Streamed audio and speech content over Internet, such as VoIP, may be modified or replaced by attackers [22, 23, 42]. Speech evidence to the court may also be modified by third parties when it is stored in the database. In secure military commanding systems, every received command should be guaranteed to be the original one. Speech content authentication refers to the technique that is able to provide speech content integrity verification. Conventional speech authentication relies on authentication method in cryptography. A short digest of the speech signal, may it be the result of hashing or the result of feature extraction, is encrypted and combined in a secure way with the original multimedia data [42]. But the ordinary user has to decrypt the speech before using it. An authentication scheme that does not hinder the normal use of the speech signal is needed. Fragile watermarking has provided us such a mechanism [15, 26, 28]. A secure mark is generated first, which may be independent of the host speech or generated from the speech features. This mark is then embedded into the speech signal so that it is not perceived by normal users. At the receiving end, a detector is used to detect the hidden mark by authorized parties. If this mark is absent from the speech signal then the speech signal is highly likely to have been modified.

To design a watermark based speech authentication algorithm, several tradeoffs need to be considered. First, there is an obvious tradeoff between robustness and fragility. The embedded watermark should be robust to some unintentional signal processing attacks, such as constant amplitude scaling and weak noise addition. But in the case of removal attack or replacement attack, the absence of watermark should be identified with high probability. Second, there is a tradeoff between the authentication length and authentication reliability. Short authentication length is preferred in order to precisely locate the segments being attacked. But long segment is preferred for improving the detection performance. How to find the appropriate authentication length from detection performance requirement is thus highly relevant to the watermarking system design problem.

This paper proposes an improved semi-fragile speech watermarking scheme by quantizing the linear prediction (LP) parameters, i.e., the inverse sine (IS) parameters. This scheme is a semi-fragile one since it is robust to amplitude scaling and semi-fragile to noise addition. There are basically two new ingredients in this improvement. The first ingredient is a modified bit allocation algorithm. Bit allocation is combined with watermarking to determine the appropriate quantization steps so that the 'transparency' requirement is satisfied. The second ingredient is an Analysis by Synthesis (AbS) watermark embedding scheme. Using an AbS loop outside of the original watermark embedder, we can effectively reduce the difference between the estimated LP coefficients and the watermarked ones. This reduction may help us to reduce the minimum authentication length for a given performance design requirement.

This paper is organized as follows. We first motivate our work by reviewing available works on speech watermarking in Section 2. In Section 3 we review the basic watermark embedding and detection algorithm. The proposed AbS based watermark embedding scheme is described in Section 4. The modified bit allocation algorithm is presented in Section 5. The watermark based authentication system design problem is discussed in Section 6. In Section 7, we present and discuss the experimental results. Finally we conclude the paper in Section 8.

2 Review of related work on speech watermarking

In the theory of watermarking, mark embedding schemes fall into two groups: spread spectrum based non-informed embedding and QIM (quantization index modulation) based side-informed embedding [1, 7]. However, when combined with special characteristic of the speech signal production and perception process, we may have a variety of algorithms. In this section, we first categorize various mark embedding algorithms in speech watermarking. Then the current advances in speech authentication based on watermarking are reviewed. Finally, we highlight the differences between our approach and the previous works.

2.1 Hidden mark embedding in speech signals

Currently available techniques for hidden mark embedding in speech signal can be roughly classified into the following categories:

Phase modification These schemes use the fact that the human auditory system is less sensitive to absolute phase than to relative phase and amplitude. In [6, 38, 46], all-pass filters were used to alter phase information. The filter coefficients are chosen as secret key to provide security. Sakaguchi proposed to invert the polarity of the excitation pulses to embed information [39]. The polarity inversion actually changes the phase of some speech segments by 180 degrees. So it is a special case of phase modification. Recently, Unoki et al. proposed a new audio watermarking framework based on cochlear delay characteristics of the human auditory system [43]. This characteristics is utilized in watermarking by modulating the group delays of two first order all-pass filters according to the watermark bits.

Linear spread spectrum In [3, 4], the watermark information is modulated using DS/BPSK (Direct Sequence Spread Spectrum/Binary Phase Shift Keying). The modulated watermark signal is embedded into the residual signal of the speech after inverse filtering. This is equivalent to shaping of the watermark signal using vocal tract filter to provide certain perceptual masking.

Parametric modeling In image and audio processing, there is no universal signal generation model. However, the speech production model is available in speech processing. The human articulatory system can be modeled as an all-pole filter driven by stimulating signal. This model has provided us an efficient tool in speech processing coding and recognition. Gurijala et al. proposed to modify the autoregressive (AR) model parameters indirectly to embed robust watermark [17]. The autocorrelation sequence is modified according to watermark sequence. The distortion is controlled in time domain using an ℓ_∞ norm. During watermark detection process, the original speech signal is required for the purpose of robust watermarking. Hatada et al. proposed to embed watermark using vector quantization of line spectrum pair (LSP) parameters [20]. The VQ codebook is modified to embed different watermark bits. Pre-selected elements of the codewords are increased or decreased by a small amount. It has been pointed out by Hatada that the extracted LSP parameters from watermarked speech signal are different from the modified LSP parameters. They choose to embed watermark in those LSPs with the smallest differences. Guard bits are also added to correct errors.

Bitstream domain scheme Low bit rate compression algorithms retain only perceptually relevant component of the speech signal hence impose severe degradation on the watermark. Embedding the mark directly into the bitstream during compression or after compression is a way to bypass the compression attack [27]. Geiser et al. designed a scheme to hide watermark into ACELP coded speech data [16]. They modified the codebook searching strategy to attain joint speech coding and watermarking. Singh et al. combined watermarking with Data Encryption Standard (DES) in G.729 coded stream to provide both security and protection of streamed speech [41].

2.2 Review of speech authentication algorithms based on watermarking

The speech watermarking schemes reviewed above mainly aim at copyright protection. So they are robust in nature. In authentication applications, a fragile watermark is preferred. Wu and Kuo presented a fragile watermarking scheme by quantization of DFT coefficients in log scale [44]. The distortion is controlled by the SNR of contemporary speech coders such as GSM speech coder. Wu also pointed out that their scheme shares the same problem as other quantization based schemes, i.e., not robust against amplitude scaling. In [31], Lu et al. combined watermarking with CELP (Code Excited Linear Prediction) speech coding process for authentication of compressed speech by CELP type coders. The excitation codebook is divided into three parts to reduce distortion. Their authentication process is based on statistical detection so that the probability of detection error is controlled under prescribed level. Recently, this algorithm is further improved by incorporating vector quantization (VQ) [30]. In [2], the authors modified the position of excitation pulses according to the content dependent watermark to provide attack identification. The last two authentication schemes are applicable only to compressed speech. In [40], Saraswathi modified the Mel-frequency cepstrum (MFCC) in speech segments having low intensity for speech authentication. However, the modified MFCCs need to be sent to the receiver, which makes this scheme essentially a non-blind one. Non-blind watermark detection has limited application in speech authentication since we need to store and transmit side information.

Special patterns in the watermark signal can also be designed and utilized to identify the type of the attacks or partially recover the speech content. In [35], a cyclic pattern is created to identify whether the attack is deletion, insertion or replication. In [29], partial reconstruction data are stored in the LSBs of excitation signals in the G.723.1 speech codec. With the aid of this data, the original speech content may be reconstructed after attacks.

In some security related applications, other application specific factors also need to be considered and incorporated into the authentication algorithms. In a series of papers [11–14], Faundez-Zanuy et al. successfully incorporated watermarking based speech authentication into telephonic recording used in social security monitoring and security enhanced speaker verification or speaker identification applications. Another successful application specific design is used in air traffic control [18, 21, 25], where the characteristic of VHF (very high frequency) radio channel is fully explored when designing the authentication algorithms.

2.3 How is this work different from previous schemes?

Our work described in this paper is different from previous schemes in the following aspects:

1. The application of this work is speech authentication, so the proposed watermarking scheme is fragile in nature. A fragile watermark should be robust against amplitude scaling which is common operation in digital speech processing. It is also expected that a fragile watermark should be partially robust against noise addition and fragile to other attacks. In addition, the original speech signal is not available in speech authentication applications, so blind decoding/detection of watermark is required. The scheme presented in this paper is able to satisfy the above requirements.
2. We choose to quantize the LP coefficients instead of the time domain signal or residual signal after inverse filtering. One reason for this is that the LP coefficients are more perceptually relevant than residual signal. In addition, the LP coefficients are invariant to constant amplitude scaling.
3. An analysis by synthesis (AbS) scheme is employed to effectively reduce the difference between the watermarked LP coefficients and the estimated LP coefficients from watermarked speech. This reduction in difference also helps to reduce the quantization steps so as to reduce the distortion introduced by watermarking.
4. The distortion introduced by watermarking LP parameters is controlled according to the 'transparency' requirement in speech coding. A modified greedy algorithm is designed to find the appropriate quantization steps.

This work is an improvement of our previous preliminary work presented in [45]. The first improvement is the use of AbS loop to reduce the difference between the watermarked LP coefficients and the extracted LP coefficients from watermarked speech. The second improvement is the use of modified bit allocation algorithm to optimize the spectral distortion.

Before describing the proposed improvements, we would like to review the basic semi-fragile speech watermarking scheme in [45]. The existing problems in the basic algorithm also motivate our new improvements in this paper.

3 Review of the basic watermark embedding and detection scheme

The basic embedding scheme without AbS loop and modified bit allocation is similar to our previous work that were presented in a conference [45]. Even though the current paper focus on improvements of the original work in [45], we review the basic algorithm here for completeness.

The structure of the watermark embedder is shown in Fig. 1. The host speech signal is first segmented into small frames each having L_f samples: $\mathbf{c} = \{\mathbf{c}_j, j = 1, \cdots, N_f\} = \{c_{i,j}, i = 1, \cdots, L_f, j = 1, \cdots, N_f\}$, where N_f is the number of frames. The binary watermark sequence $\mathbf{w}$ is also segmented into N_f frames, with each frame containing P bits, where P is the order of the linear predictor used in LPA (linear prediction analysis). For embedding, we only need to focus on one frame, say the j–th frame $\mathbf{c}_j$. Let c_i be the speech sample with index i within the j-th frame, where we have suppressed the frame index for clarity. Let $\mathbf{R}$ be the short-term autocorrelation matrix and $\mathbf{r} = (r_1, r_2, \cdots, r_p)^T$, where $r_\eta = \mathcal{L}\{c_i c_{i-\eta}\}$ and $\mathcal{L}(\cdot)$ denotes the time average of one realization of the WSS ergodic stochastic process. The estimated LP coefficients are $\mathbf{a} = -\mathbf{R}^{-1} \cdot \mathbf{r}$, where $\mathbf{a} = \{a_i\}_{i=1}^P$. In practical implementation, this matrix inversion is usually realized by fast algorithm: the Levinson–Durbin recursion [5]. Using the estimated LP coefficients, we may get the residual signal of this frame

$$e_i = c_i + \sum_{k=1}^{P} a_k c_{i-k}, \quad i = 1, \cdots, L_f, \tag{1}$$

where P is order of the AR model. To embed watermark bit, we need to quantize these model parameters. However, these estimated LP coefficients cannot be quantized directly because the pole locations of AR model can't be controlled by LP coefficients. So stability of the watermarked AR model can't be guaranteed. In speech coding, the LP coefficients are usually transformed into LAR (log area

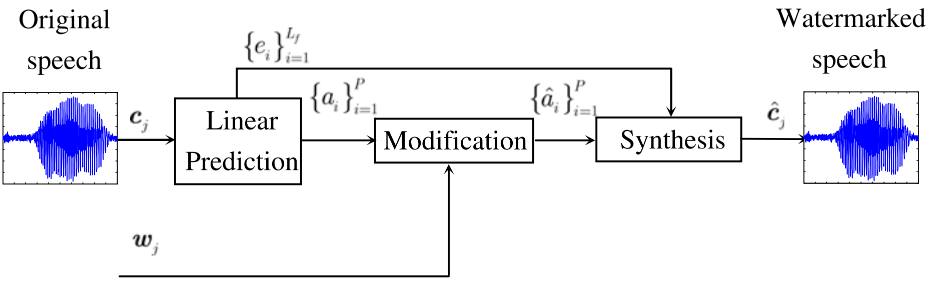


Fig. 1 Structure of the watermark embedder described in Table 1

Table 1 Basic fragile watermark embedding algorithm

1. Variable definitions.
 $\mathbf{c}$: The host speech signal where $\mathbf{c} = \{c_i\}_{i=1}^{L}$.
 P: Order of the linear predictor.
 $\mathbf{w}$: The watermark to embed where $\mathbf{w} = \left\{w_{i,j}, i = 1, \cdots, P, \text{and } j = 1, \cdots, N_f\right\}$
 Δ_i: The quantization step for the i-th inverse sine parameter.
2. Divide the host speech $\mathbf{c}$ into non-overlapping frames, which produces
 $\mathbf{c} = \left\{\mathbf{c}_j, j = 1, \cdots, N_f\right\} = \left\{c_{i,j}, i = 1, \cdots, L_f, j = 1, \cdots, N_f\right\}$.
3. Embed watermark.
 For $j \leftarrow 1, \cdots, N_f$
 $\{a_i\}_{i=1}^{P} \leftarrow \text{LPA}(\mathbf{c}_j)$.
 $e(n) \leftarrow c(n) + \sum_{k=1}^{P} a(k)c(n-k), n = 1, \cdots, L_f$.
 $\{\kappa_i\}_{i=1}^{P} \leftarrow \text{LPC2RC}\left(\{a_i\}_{i=1}^{P}\right)$.
 $g_i \leftarrow \frac{2}{\pi} \sin^{-1}(\kappa_i), i = 1, \cdots, P$.
 $\hat{g}_i \leftarrow \lfloor \frac{g_i + (1 - w_{i,j})\Delta_i}{2\Delta_i} \rfloor \times 2\Delta_i + w_{i,j}\Delta_i, i = 1, \cdots, P$.
 $\hat{\kappa}_i \leftarrow \sin\left(\frac{\pi}{2}\hat{g}_i\right), i = 1, \cdots, P$.
 $\left\{\hat{a}_i\right\}_{i=1}^{P} \leftarrow \text{RC2LPC}\left(\left\{\hat{\kappa}_i\right\}_{i=1}^{P}\right)$.
 $\hat{c}(n) \leftarrow -\sum_{k=1}^{P} \hat{a}(k)\hat{c}(n-k) + e(n), n = 1, \cdots, L_f$.
 End
4. Concatenate all the synthesized output frames $\hat{\mathbf{c}}_j$ into output stream $\hat{\mathbf{c}}$.
Note:
 LPA$(\cdot)$ is the linear prediction analysis.
 LPC2RC$(\cdot)$ convert the LP coefficients to the reflection coefficients (RC).
 $\lfloor \cdot \rfloor$ round to the nearest integer towards $-\infty$.
 RC2LPC$(\cdot)$ convert the reflection coefficients back to the LP coefficients.

ratio) or IS (inverse sine) parameters to guarantee stability and to reduce the spectral sensitivity. The LP coefficients are first transformed into reflection coefficients (RC) $\{\kappa_i\}_{i=1}^{P}$ and then into the IS parameters $\{g_i\}_{i=1}^{P}$ by $g_i = \frac{2}{\pi}\sin^{-1}(\kappa_i), i \in \{1, 2, \cdots, P\}$. The modification to the IS parameters are based on odd-even modulation [26]

$$\hat{g}_i = \left\lfloor \frac{g_i + (1 - w_i)\Delta_i)}{2\Delta_i} \right\rfloor \times 2\Delta_i + w_i\Delta_i, \quad i \in \{1, 2, \cdots, P\} \tag{2}$$

where $\lfloor \cdot \rfloor$ denotes rounding towards $-\infty$. Δ_i is the quantization step for the i-th IS parameter. $w_i \in \{0, 1\}$ is the watermark bit to be embedded. The modified IS parameters $\{\hat{g}_i\}_{i=1}^{P}$ are inversely transformed into LP coefficients $\hat{a} = \{\hat{a}_i\}_{i=1}^{P}$. The residual signal from the LP analysis stage (1) is used to synthesize the watermarked speech signal $\hat{c}_i = -\sum_{k=1}^{P} \hat{a}_k\hat{c}_{i-k} + e_i$. This basic watermark embedding algorithm is summarized in Table 1.

At the receiving end, the watermark bit is extracted by minimum distance decoding [1]:

$$\tilde{w}_i = \left\lfloor \frac{\tilde{g}_i}{\Delta_i} + \frac{1}{2} \right\rfloor \pmod 2, \quad i \in \{1, 2, \cdots, P\} \tag{3}$$

where $\{\tilde{g}_i\}_{i=1}^{P}$ are the IS parameters estimated from the received speech signal. A detection statistic based on the embedded bits and extracted bits is then calculated

Table 2 Basic fragile watermark extraction and detection algorithm

1. Variable Definitions.
$\tilde{\mathbf{c}}$: the watermarked and possibly attacked signal, where $\tilde{\mathbf{c}} = \{\tilde{c}_i\}_{i=1}^{L}$.
2. Divide the received speech signal $\tilde{\mathbf{c}}$ into non-overlapping frames, which produces:
$\tilde{\mathbf{c}} = \{\tilde{\mathbf{c}}_j, j = 1, \cdots, N_f\}$
$= \{\tilde{c}_{i,j}, i = 1, \cdots, L_f, j = 1, \cdots, N_f\}$.
3. Extract the watermark.
For $j \leftarrow 1, \cdots, N_f$
$\{\tilde{a}\}_{i=1}^{P} \leftarrow \mathrm{LPA}(\tilde{\mathbf{c}}_j)$.
$\{\tilde{\kappa}\}_{i=1}^{P} \leftarrow \mathrm{LPC2RC}\left(\{\tilde{a}_i\}_{i=1}^{P}\right)$.
$\tilde{g}_i \leftarrow \frac{2}{\pi} \sin^{-1}(\tilde{\kappa}_i), i = 1, \cdots, P$.
$\tilde{w}_{i,j} = \lfloor \frac{\tilde{g}_i}{\Delta_i} + \frac{1}{2} \rfloor \pmod 2, \ i = 1, \cdots, P$.
End
4. Watermark detection.
Calculate the detection statistic $D = N_f \times P - \sum_{i=1}^{P} \sum_{j=1}^{N_f} w_{i,j} \oplus \tilde{w}_{i,j}$.
Return ‘watermark detected’ if D is greater than some threshold $\mathcal{T}$.
Return ‘no watermark’ otherwise.
Note:
$\oplus$ is exclusive or operation.

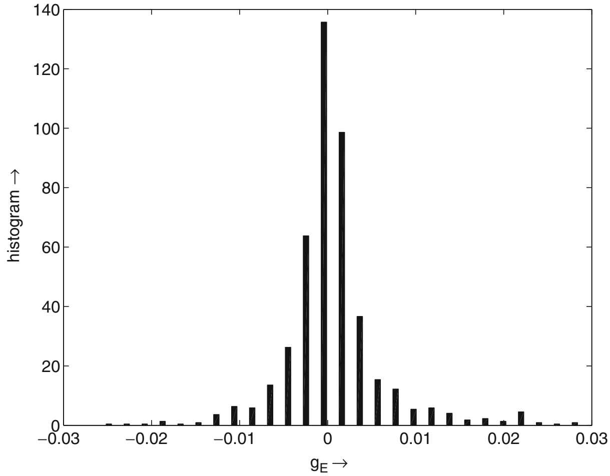
$$D = N_f \times P - \sum_{i=1}^{P} \sum_{j=1}^{N_f} w_{ij} \oplus \tilde{w}_{ij} \ .$$

Note that we have used both the frame index j and the index i within one frame since we need more than one frame in authentication stage. This statistic is then compared to a predefined threshold $\mathcal{T}$. We conclude that the speech signal is not modified when the detection statistic is above this threshold, i.e. $D > \mathcal{T}$. We have summarized the watermark extraction and authentication algorithm in Table 2. For detailed implementation and testing of this algorithm we refer the readers to [45].

4 Analysis by synthesis based embedding scheme

The estimated LP coefficients $\{\tilde{a}_i\}_{i=1}^{P}$ are different from the modified LP coefficients $\{\hat{a}_i\}_{i=1}^{P}$ even when there is no attack. It is pointed out in [8] that the estimated $\{\tilde{a}_i\}_{i=1}^{P}$ is multivariate Gaussian distributed around $\{\hat{a}_i\}_{i=1}^{P}$. But the conversion from LP coefficients to IS parameters is nonlinear, so the set of IS parameters is no longer Gaussian distributed. To visualize the statistic of the difference in terms of IS parameters, We plot the histogram of the difference between $\tilde{g}_1$ and $\hat{g}_1$ in Fig. 2. This effect may actually destroy some watermark bits and cause decoding error even in the absence of attacks. A similar problem is encountered in [20] when the authors tried to embed watermark by vector quantization of LSP parameters. They chose to embed the watermark in those LSP parameters that have the smallest variation and used guard bits to further increase reliability. In our preliminary fragile watermarking scheme presented in [45], we have modeled the difference as random noise and used statistical detection in authentication process. However, all these schemes aim not to reduce the difference but to combat the effect of this difference by choosing appropriate embedding position or by using statistical detection. We present in this

Fig. 2 Histogram of the difference between estimated inverse sine parameter $\tilde{g}_1$ and watermarked inverse sine parameter $\hat{g}_1$, the quantization step Δ is 0.015



section a framework to reduce the difference when there is no attack. We employ an Analysis by Synthesis (AbS) approach. The basic structure of our AbS method is illustrated in Fig. 3. The basic algorithm consists of the following steps.

Step 1 Embed the watermark into the speech signal using the basic embedding algorithm described in Table 1 and obtain the watermarked speech signal $\hat{\mathbf{c}}$.

Step 2 If the difference between estimated IS parameter $\tilde{g}$ from watermarked signal and watermarked IS parameter $\hat{g}$ is lower than some threshold then stop and return the watermarked speech signal $\hat{\mathbf{c}}$. Otherwise, go to Step 3.

Step 3 Analyze the synthesized speech $\hat{\mathbf{c}}$ again using LP analysis and get the new residual signal $\{\hat{e}_i\}_{i=1}^{L_f}$. Use the new residual signal $\{\hat{e}_i\}_{i=1}^{L_f}$ and the modified LP coefficients $\{\hat{a}_i\}_{i=1}^{P}$ to re-synthesize the watermarked speech signal $\hat{\mathbf{c}}$.

Step 4 Go to Step 2 to check the loop condition.

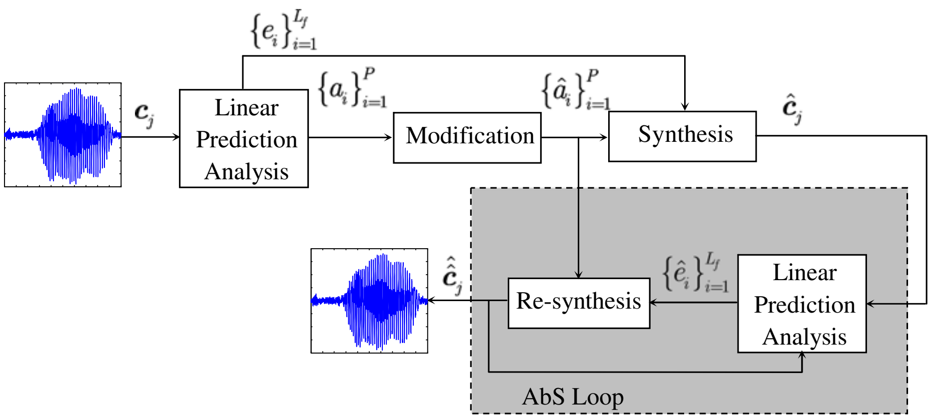


Fig. 3 Structure of the watermark embedder with Analysis by Synthesis (AbS) loop

The effectiveness of this approach will be demonstrated in the simulation experiment in Section 7.

5 Quantization step optimization by modified bit allocation

In algorithms described above, only one quantization step was used in odd-even modulation to watermark all the IS parameters. This quantization step was designed by trial and error and subjective listening test [45]. However, due to the differences in IS parameters' contribution to distortion, different quantization steps should be used for different IS parameters. We should design the quantization steps for quantizers in odd-even modulation so that the distortion introduced by watermarking is not perceptible. On the other hand, if the step size is too small, then the semi-fragility is not retained. So the step size should be large enough to achieve semi-fragility and it should be small enough so that there is no perceptual difference between original and watermarked speech. In this section, we design a greedy algorithm to determine the quantization steps in watermarking. This algorithm is a modification of the conventional bit allocation technique in speech coding [5]. Speech distortion here is measured by the objective distortion measure: spectral distortion (SD). SD is the difference between the power spectrum density (PSD) of two autoregressive signals. Let $\{a_i\}_{i=1}^{P}$ and $\left\{\hat{a}_i\right\}_{i=1}^{P}$ be the un-watermarked and watermarked LP coefficients respectively. Then their corresponding AR signal PSDs are given as [24]:

$$S(\omega) = \frac{1}{\left|1+\sum_{\ell=1}^{P} a_\ell e^{-j\omega\ell}\right|^2}, \quad \hat{S}(\omega) = \frac{1}{\left|1+\sum_{\ell=1}^{P} \hat{a}_\ell e^{-j\omega\ell}\right|^2}, \tag{4}$$

where we have assume that the driving white Gaussian noise of the AR processes have unit variance. The spectral distortion between these two signals is defined by

$$SD = \sqrt{\frac{1}{2\pi}\int_0^{2\pi}\left[10\log_{10} S(\omega) - 10\log_{10}\hat{S}(\omega)\right]^2 d\omega}. \tag{5}$$

In speech coding, it is found from subjective listening test that the SD between the original speech signal and the compressed speech signal should satisfy the following 'transparency' requirements: (1) The average value of the SD should be less than 1dB, (2) the percentage of outliers with spectral distortion greater than 2dB should be less than 2%, and (3) there should be no outliers with spectral distortion greater than 4dB. In the following algorithm, we will use this criterion to determine the watermark quantization steps. Since we use uniform quantization to quantize each IS parameter and we know that the IS value is within the range of $[-1, 1]$, so the quantization step determination problem is equivalent to determining how many bits we should use to represent each IS parameter. For each of the P IS parameters, we design a quantizer $\mathcal{Q}_\ell$, where $\ell \in \{1, \cdots, P\}$. Suppose that b bits are used to quantize an IS parameter, then the quantization step is

$$\Delta = \frac{IS_{\max} - IS_{\min}}{2^b},$$

where $[IS_{\min}, IS_{\max}]$ is the range of the uniform quantizer. This range is chosen according to the statistical property of each IS parameter. The distribution of higher order IS parameters are well approximated by Gaussian distribution. So the range of these quantizers are chosen as

$$IS_{\max} = \mu + 4\sigma,\ IS_{\min} = \mu - 4\sigma,$$

where μ and σ^2 are mean and variance of the Gaussian distribution respectively. For Gaussian distribution, approximately 99.99% of the probability mass falls into this range. If this range is too small, say $\pm 2\sigma$ around the mean μ, then the quantization step will be too small for fixed b, resulting in low robustness and high fragility. There is no appropriate parametric distribution that fits the histogram of the lower order IS parameters. So the minimum and maximum value of lower order ISs are used as the quantizer's range. A histogram plot of the IS parameters obtained from the training speech signals is shown in Fig. 4.

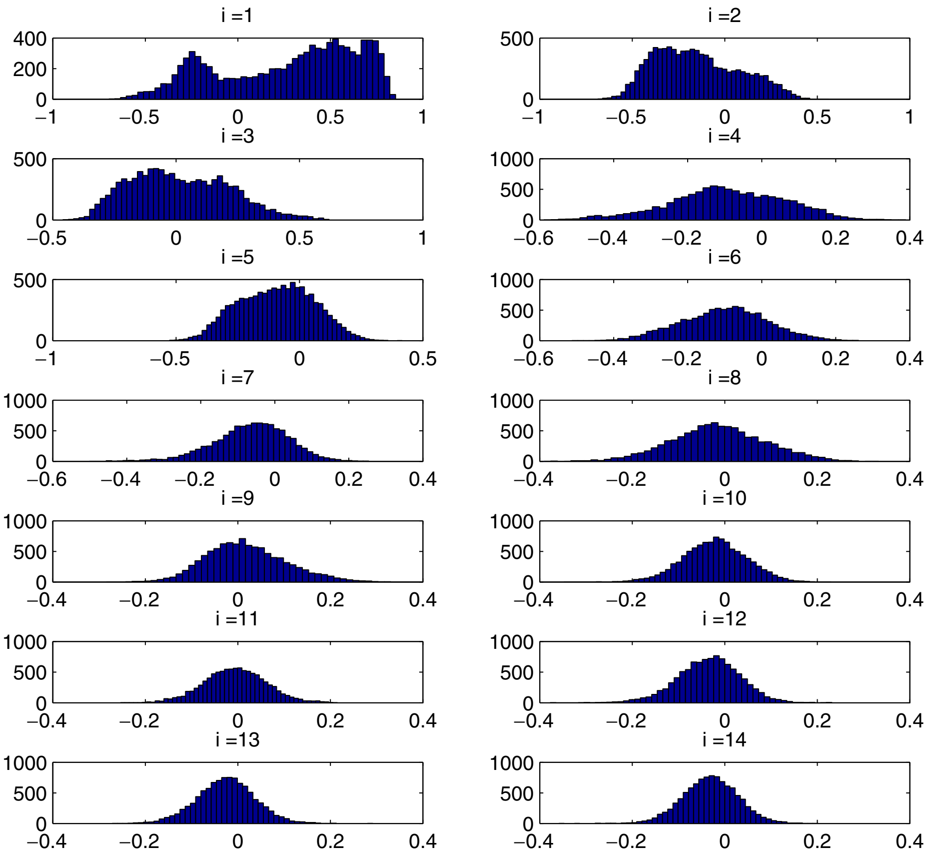


Fig. 4 Histogram of the IS parameters obtained from the training speech signals. The distribution of higher order IS parameters are well approximated by Gaussian distribution. There is no appropriate parametric distribution that fits the histogram of the lower order IS parameters

The statistical property of the watermark sequence used in the training algorithm is also very important. In the watermarking algorithm described in Table 1, we use two quantizers to quantize the IS parameters. In addition, we use the watermark bit to decide which quantizer to use. If the watermark bit is zero, then we use a mid-tread quantizer. Otherwise we use a mid-rise quantizer. In conventional speech coding, either the mid-tread or the mid-rise quantizer is used. Bit allocation schemes that are optimal for mid-tread quantizers are not optimal for mid-rise quantizers. The relative frequency of 0's and 1's in the training sequence determines the relative frequency of using mid-tread and mid-rise quantizers. So training watermark sequence with unrealistic relative frequency may degrade the performance of the bit allocation algorithm. It is appropriate to assume that $\Pr\{w=0\} = \Pr\{w=1\}$ in most cases. So we generate the watermark training sequence that consists of equal number of 1's and 0's. Each bit in the sequence is generated independent of others.

Suppose that there are totally N_b bits to allocate, then the modified bit allocation algorithm consists of the following steps.

Step 1 Generate the training watermark sequence $w_{i,j} \in \{0,1\}^{P\times N_f}$ that satisfies $\Pr\left\{w_{i,j}=0\right\} = \Pr\left\{w_{i,j}=1\right\}$. Initialize the number of allocated bit $m=0$. Set $r_{m,\ell}=0$, where $r_{m,\ell}$ denotes the number of bits allocated to the ℓ-th IS parameter after m bits have been allocated.

Step 2 Increase the number of allocated bits by one, i.e., $m=m+1$. If m exceeds N_b , then stop.

Step 3 For $\ell=1,2,\cdots,P$ do Step 4.

Step 4 Increase the number of allocated bit for the ℓ-th IS parameter by one: $r_{m,\ell} = r_{m-1,\ell}+1$ and $r_{m,k}=r_{m-1,k}$, for $k\neq\ell$. Then design the uniform quantizers using step size

$$\Delta_k = \frac{IS_{\max,k} - IS_{\min,k}}{2^{r_{m,k}}},\ k=1,2,\cdots,P.$$

The training watermark sequence $\left\{w_{i,j}, i=1,\cdots,P,\ j=1,\ldots,N_f\right\}$ was then embedded into the training speech sequence TS using watermarking algorithm described in Table 1 with quantizer set $\left\{\mathcal{Q}_{m,k}\right\}_{k=1}^{P}$. Calculate the average spectral distortion $D_{m,\ell}$.

Step 5 Find the quantizer index that has the minimum spectral distortion:

$$\ell_{\min} = \arg\min_{\ell\in\{1,2,\cdots,P\}} D_{m,\ell}.$$

Then the m-th bit is allocated to the quantizer for the $\ell_{\min}$-th IS parameter, i.e.,

$$r_{m,\ell_{\min}} = r_{m-1,\ell_{\min}}+1,$$

and

$$r_{m,k} = r_{m-1,k},\ \forall k\neq\ell_{\min}.$$

Then go to Step 2 and repeat the process to allocate the next bit.

When the algorithm terminates, we get the final bit allocation scheme for each IS parameter $r_{N_b,k}, k=1,2,\cdots P$. This bit allocation scheme is treated as design parameter and is made known to both the embedder and the decoder. The training sequence should be selected to reflect the true statistic of the ensemble of all possible speech signals.

6 Authentication system parameter design

Speech authentication is based on the test statistics D defined in Table 2: $D = N_f \times P - \sum_{i=1}^{P}\sum_{j=1}^{N_f} w_{i,j} \oplus \tilde{w}_{i,j}$. Authentication system performance is described in terms of probability of false alarm (P_{FA}) and probability of missed detection (P_{MD}). If the received speech is not modified but we find that D is blow certain threshold $\mathcal{T}$, then false alarm occurs. On the other hand, if the speech is modified but we find that D is above the threshold $\mathcal{T}$, then missed detection occurs. We want our system to have low P_{FA} and low P_{MD}. But these two performance indices are related to each other. One can't reduce them simultaneously for a given design. Usually these two performance indices are given by the end users as performance requirement. The designer of the system need to find the appropriate design parameters such as authentication length and detection threshold $\mathcal{T}$ according to the system performance requirement [15, 45].

The improvement of the fragile watermarking system reported in this paper is focused on reducing difference between modified IS parameters and extracted IS parameters and designing quantization steps based on spectral distortion optimization. The watermark message decoding and watermark detection process are basically the same as that described in our previous system [45]. So we can use the relationship derived in that paper as described in (6) and (7).

$$N_{\mathrm{w}}^{*} = \left[\frac{Q^{-1}(1-\beta)\sqrt{g(\mathrm{SNR})\cdot(1-g(\mathrm{SNR}))} - 0.5Q^{-1}(\alpha)}{g(\mathrm{SNR}) - 0.5}\right]^2, \tag{6}$$

$$\mathcal{T}^{*} = \frac{N_{\mathrm{w}}^{*}}{2} + \frac{\sqrt{N_{\mathrm{w}}^{*}}}{2} Q^{-1}(\alpha) \;, \tag{7}$$

where $Q^{-1}(\cdot)$ is the inverse function of the Q function which is defined as the right tail probability of the Gaussian distributed random variable: $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})dt$. α and β are the design requirements on P_{FA} and P_{MD} respectively: $P_{\mathrm{FA}} = \alpha$, $P_{\mathrm{MD}} = \beta$. These design requirements are usually determined by the applications. SNR is the signal to noise ratio in additive noise channel [45], and $g(\cdot)$ is the functional relationship between SNR and tamper assessment function $\mathrm{TAF} = 1 - D/(N_f \times P)$ as found in [45]. Using (6) and (7), we can find the required authentication length N_w^* and threshold $\mathcal{T}^*$ from the design requirement on P_{FA} and P_{MD}.

7 Experimental results

7.1 Implementation

The improved algorithm is implemented and tested in MATLAB. The speech signals used in the experiments are taken from international event news distributed over

Internet. A segment of the testing signal is shown in Fig. 5. The sampling rate of the signal is 8 kHz. Each frame consists of 160 samples hence corresponds to a 20 ms segment of speech signal. This segment length results in an achieved embedding rate of $50P$. The testing signal has totally 20,000 frames. The number of frames involved in the authentication process is 20. So we have 1,000 authentication experiments. We use a linear predictor of order 14 in the experiments which is common in speech processing. Before the LP analysis we also perform the offset removal and pre-emphasis operations. The offset removal filter in this implementation is

$$H_{OR}(z) = \frac{1 - z^{-1}}{1 - 0.998z^{-1}}$$

and the pre-emphasis filter is $H_{PE}(z) = 1 - 0.86z^{-1}$. The watermark bit sequence was generated by thresholding a sequence of numbers uniformly distributed between [0,1] using threshold 0.5. So the number of 1's and 0's are roughly equal. Thus the requirement that $\Pr\{w = 0\} = \Pr\{w = 1\}$ as discussed in Section 4 are met.

The filtering operation in watermark embedding is realized by transposed direct form II structure due to its numerical stability [33]. Both the basic embedding algorithm and the AbS based embedding algorithm are based on frame processing. In speech analysis and synthesis, usually we use overlapping frames to suppress the abrupt change of parameters between frames. But in watermark embedding, frame overlapping may destroy the embedded watermark. So non-overlapping framing is needed. But the modification to the speech signal using non-overlapping frames, if not implemented appropriately, may results in audible artifacts [10]. To mitigate this

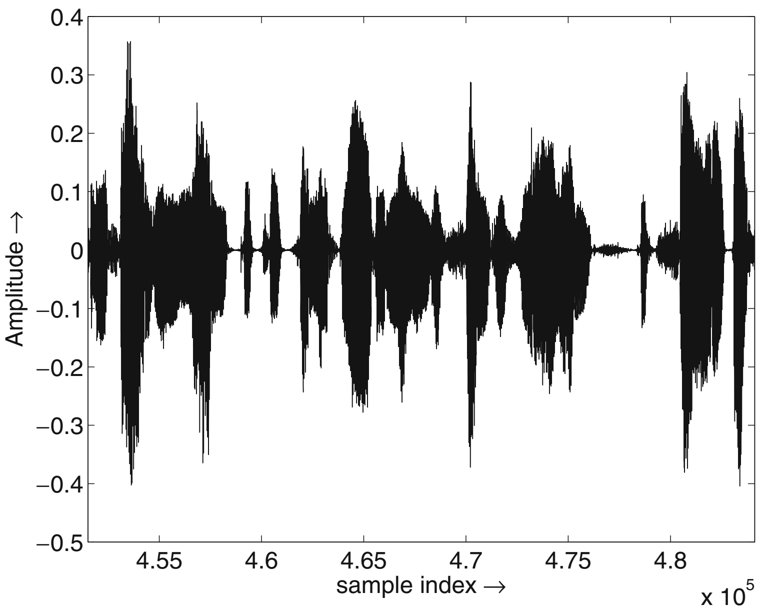


Fig. 5 A segment of the testing speech signal. This signal is taken from FM radio program of international event news

artifacts, when filtering the current frame, the initial state of the filter is set to the final state of the filter used in last frame. After all frames are watermarked, de-emphasizing the overall synthesized speech is also needed. De-emphasizing is done by using the inverse filter of $H_{PE}(z)$. For more information about implementation of non-overlapping frame based speech processing, we refer the readers to books related to realization of AbS based speech coders, such as [10, 36, 37].

Spectral distortion between two AR models in (5) are calculated numerically. Given parameters of two AR models, we first calculate their frequency response and sample the frequency axis. The number of points used in frequency sampling is 256. But when calculating the spectral distortion, only samples between 4 and 100 are used. This index range corresponds to frequency range from 125 to 3,100 Hz, since our speech signal is sampled at 8 kHz in time domain. The frequency range [125, 3,100] are chosen since the human auditory system (HAS) is insensitive to spectral distortion of speech signals outside of this range [8, 34]. We have also experimented with frequency sampling rate that is larger than 256, the results are essentially the same.

The attack experiments results for amplitude scaling and white noise addition are similar to that of our previous scheme [45], so they are omitted here to conserve space.

7.2 Inverse sine parameter estimation error reduction experiment

Let $g_E = \tilde{g} - \hat{g}$, then the change of the standard deviation of g_E with the number of iterations is shown in Fig. 6. Obviously the estimation errors are different for different IS parameters. If we define the relative drop as the ratio between the error

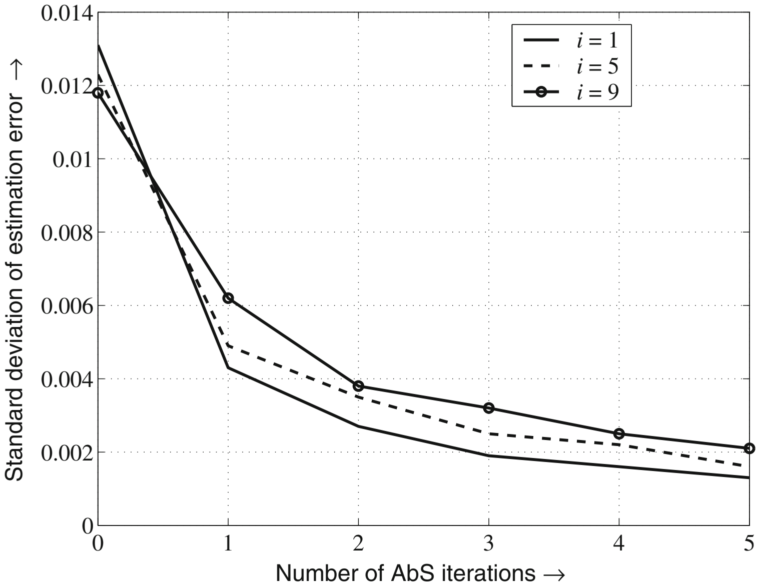


Fig. 6 The standard deviation of IS parameter estimation error with AbS loop, the first, fifth and ninth IS parameters' estimation errors are shown

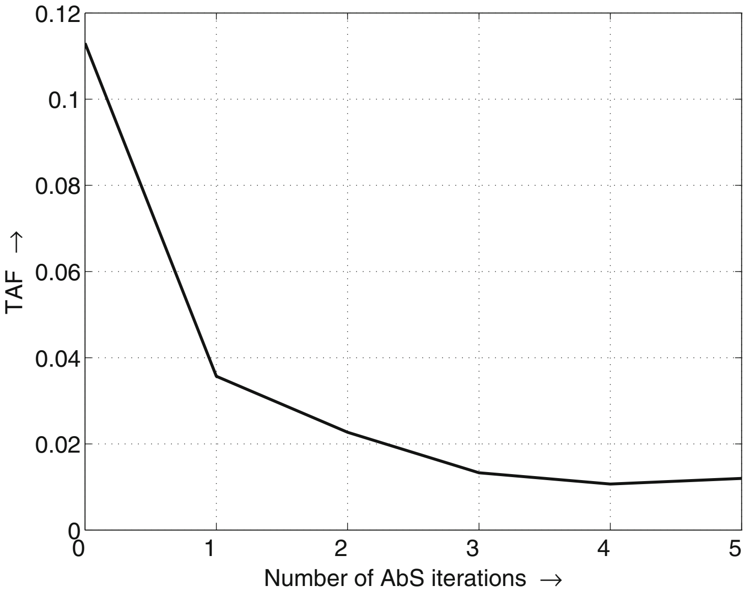


Fig. 7 The TAF value with AbS loop, no attacks. When the number of iterations exceed 3, the drop in TAF value is relatively small

reduction after one iteration and the error before this iteration, then the relative drop for the first IS parameter is roughly 70% after the first iteration. The relative drop for the fifth IS parameter is roughly 60 % after the first iteration. We also observe that the relative drop is the largest after the first iteration than after later iterations for all these three IS parameters. When the number of iterations is greater than 3, the relative drop tends to be the same. The TAF value vs. number of AbS iterations when there is no attack is shown in Fig. 7, which shows similar trend as in Fig. 6. The experimental results shown in Figs. 6 and 7 demonstrate that the proposed AbS algorithm can effectively reduce the estimation error of IS parameters and hence can effectively reduce the TAF value in the absence of attacks. It is observed from the experiments that when the number of iterations is greater than 3, the subjective quality of the synthesized speech is affected. However as we found above, when the number of AbS iterations is greater than 3, the estimation error reduction is small compared to the first and second iterations. Considering the appropriate tradeoff between subjective quality and estimation error reduction, the number of AbS iterations should be chosen as 1 or 2.

7.3 Modified bit allocation based SD optimization experiment

In our modified bit allocation scheme, the training algorithm combined with watermarking yields greater distortion than the original bit allocation algorithm used in speech coding. Figure 8 compares the average spectral distortion of training

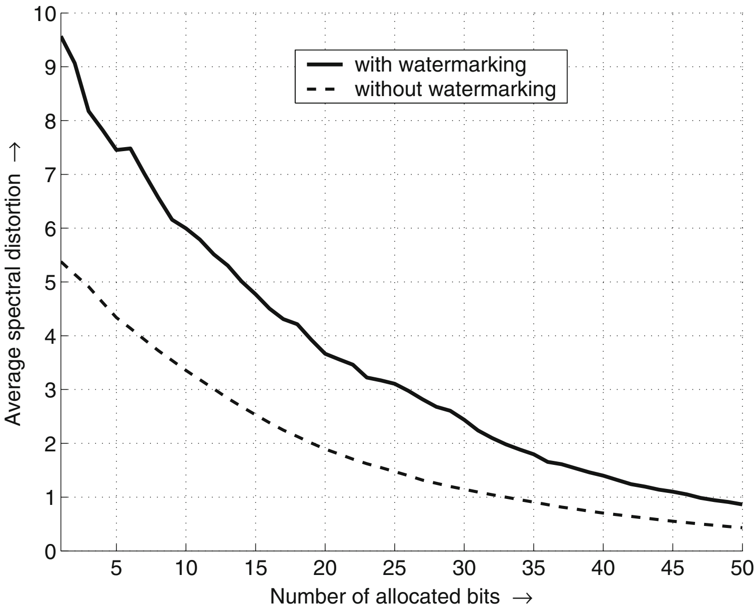


Fig. 8 Comparison of spectral distortion with and without watermarking

speech due to quantization with and without watermarking. The increase in distortion is actually the price we must pay for embedding information. As we can see from Fig. 8, watermarking needs roughly ten more bits to achieve the same SD as when there is no watermarking. Namely, in order to achieve the same spectral distortion, quantization watermarking requires smaller quantization steps than the case without watermarking. One may notice the fluctuation in the average spectral distortion curve with watermarking in Fig. 8. For example, when the number of allocated bits increases from 5 to 6, the average spectral distortion increases also, which is opposite to our expectation and the average trend of the curve. This fluctuation comes from the fluctuation of the training watermark sequence $\{w_{i,j}, i = 1, \cdots, P, \text{ and } j = 1, \cdots, N_f\}$. We have used a training watermark sequence to aid the allocation of each new available bit. But the training watermark sequences are different for different allocations. Namely, the watermark training sequence used to allocate the i-th bit is different from the training sequence used to allocate the $(i+1)$-th bit. So the fluctuation in the average spectral distortion curve reflects the fluctuation of the training watermark sequence.

Figure 9 shows how the percentage of outliers decrease with the increase of allocated bits. When the number of allocated bits is above 35, the percentage of outliers that are greater than 4 dB decreases to essentially zero. When the number of allocated bits is above 45, the percentage of outliers that are greater than 2 dB decreases to essentially zero.

According to the ‘transparency’ criterion and subjective listening test, when the number of allocated bits is greater than 65, no audible artifacts can be heard. When the total number of allocated bits is 65, the bit allocation scheme found by the above

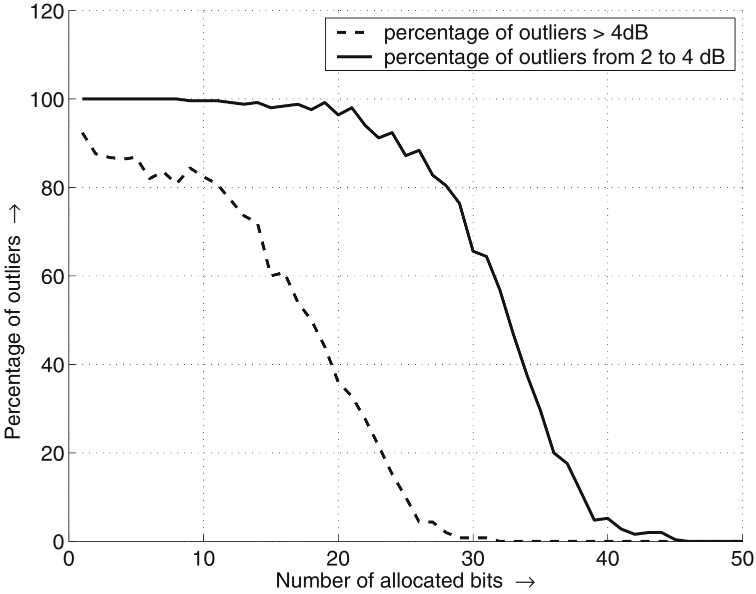


Fig. 9 The decrease of outliers' percentage with the increase of allocated bits. When the number of allocated bits is above 35, the percentage of outliers that are greater than 4 dB decreases to essentially zero. When the number of allocated bits is above 45, the percentage of outliers that are greater than 2 dB decreases to essentially zero

greedy algorithm is [7, 6, 5, 5, 5, 4, 5, 4, 4, 4, 4, 4, 4, 4]. Under this bit allocation scheme, several frames of the first IS parameters before, after quantization and estimated from watermarked speech are shown in Fig. 10. The time domain plot of

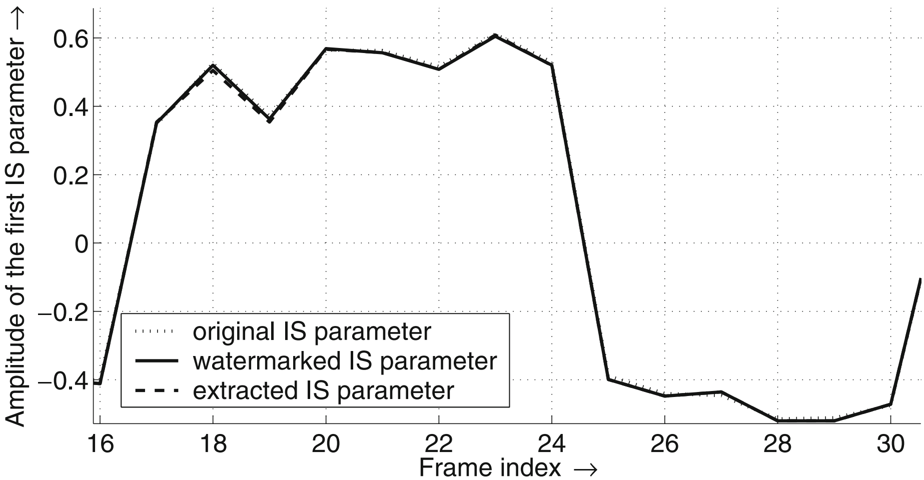


Fig. 10 The first IS parameter before watermarking, after watermarking and extracted from watermarked speech

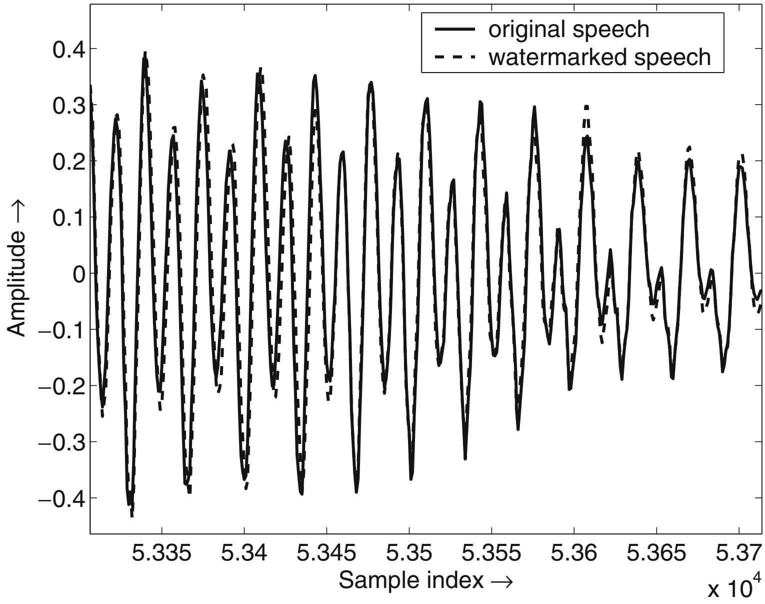


Fig. 11 The speech signals before and after watermarking

the speech signal before and after watermarking is shown in Fig. 11. The histogram of the spectral distortion is shown in Fig. 12. The average spectral distortion is found to be 0.83 and no outlier is found that is greater than 2dB.

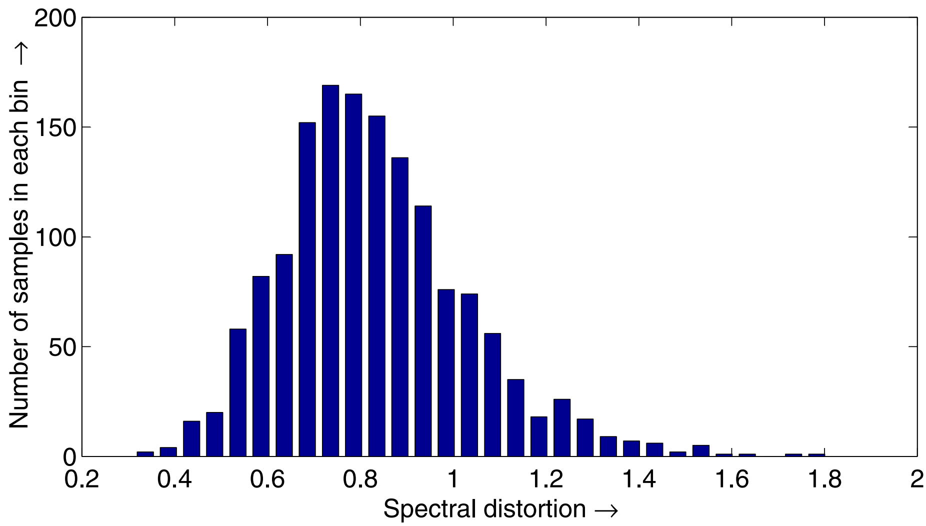


Fig. 12 The histogram of the spectral distortion. The number of bits used to quantize each IS parameter are 7, 6, 5, 5, 5, 4, 5, 4, 4, 4, 4, 4, 4, 4, respectively

8 Conclusion

In this paper, we propose two improvements to semi-fragile watermarking algorithm based on quantization of LP parameters. The experimental results show that the AbS based embedding algorithm can effectively reduce the difference between the watermarked LP parameters and the extracted LP parameters hence can reduce the minimum required authentication length. Considering the tradeoff between performance and perceptibility, only 1 or 2 iterations of the AbS loop are needed. The modified bit allocation scheme can automatically find the appropriate quantization steps for watermarking IS parameters hence can ease the design of the authentication system. Our improvement here combined with the basic authentication scheme reported in [45] is applicable to speech authentication applications where there is unknown constant amplitude scaling attack on the watermarked speech.

Several problems remain unexplored in our current work. For example, the mechanism of the error between the watermarked IS parameters and the estimated IS parameters is still not clear. In addition, the effectiveness of the AbS loop is only verified experimentally. One of our future research topics is to explore the theoretical explanation of the effectiveness of the AbS loop. Another possible research topic may be the use of non-uniform quantizers in watermark embedding, which may require joint optimization of the quantization steps and bit allocation schemes.

Acknowledgements This work is supported by research project of "SUST Spring Bud” under the grant number: 2009AZZ155. The work of the second author is also supported by the project of National Natural Science Foundation of China (NSFC) under project grant number: 61071087. The authors would like to thank the anonymous reviewers for their constructive comments and suggestions. We are indebted to the reviewers for their valuable time spent on the manuscript of this paper. The first author would like to thank Prof. Zhe-Ming Lu, Prof. Sheng-He Sun, Prof. Jeng-Shyang Pan and Prof. Xia-Mu Niu for their guidance and help in developing the basic algorithm upon which the extension in this paper is built.

References

1. Barni J, Bartolini F (2004) Watermarking systems engineering: enabling digital assets security and other applications. Marcel Dekker, New York
2. Chen OTC, Liu CH (2007) Content-dependent watermarking scheme in compressed speech with identifying manner and location of attacks. IEEE Transactions on Audio, Speech, and Language Processing 15(5):1605–1616
3. Cheng Q, Sorensen S (2001) Spread spectrum signaling for speech watermarking. In: IEEE international conference on acoustics speech and signal processing, vol 3, pp 1337–1340
4. Cheng Q, Sorensen S (2005) Spread spectrum signaling for speech watermarking. United States Patent, No. US6892175B1
5. Chu WC (2003) Speech coding algorithms: foundation and evolution of standardized coders. Wiley-Interscience, New Jersey
6. Ciloglu T, Karaaslan SU (2003) An improved all-pass watermarking scheme for speech and audio. In: International conference on multimedia and expo., vol 2, pp 1017–1020
7. Cox IJ, Miller ML, McKellips AL (1999) Watermarking as communication with side information. Proc IEEE 87(7):1127–1141
8. Deller JR, Hansen JHL, Proakis JG (1993) Discrete-time processing of speech signals. IEEE Press, Piscataway

9. Deng Z, Yang Z, Shao X, Xu N, Wu C, Guo H (2007) Design and implementation of steganographic speech telephone. In: Ip H, Au O, Leung H, Sun MT, Ma WY, Hu SM (eds) C PCM 2007: advances in multimedia information processing. Lecture notes in computer science, vol 4810. Springer, Berlin, pp 429–432
10. Dutoit T, Marques F (2009) Applied signal processing, 1st edn. A MATLAB-based proof of concept. Springer, New York
11. Faundez-Zanuy M (2010) Digital watermarking: new speech and image applications. In: Sol-Casals J, Zaiats V (eds) Advances in nonlinear speech processing. Lecture notes in computer science, vol 5933. Springer, Berlin, pp 84–89
12. Faundez-Zanuy M, Hagmüller M, Kubin G (2006) Speaker verification security improvement by means of speech watermarking. Speech Commun 48(12):1608–1619
13. Faundez-Zanuy M, Hagmüller M, Kubin G (2007) Speaker identification security improvement by means of speech watermarking. Pattern Recogn 40(11):3027–3034
14. Faundez-Zanuy M, Lucena-Molina JJ, Hagmüller M (2010) Speech watermarking: an approach for the forensic analysis of digital telephonic recordings. J Forensic Sci 55:1080–1087
15. Fei C (2006) Analysis and design of watermark-based multimedia authentication system. Ph.D. thesis, University of Toronto
16. Geiser B, Vary P (2008) High rate data hiding in ACELP speech codes.In: Proceedings of ICASSP 2008, pp 4005–4007
17. Gurijala A, Deller JR, Jr, MSS, Hansen JHL (2002) Speech watermarking through parametric modeling. In: Proceedings of international conference on spoken language processing (ICSLP)
18. Hagmüller M, Horst H, Kröpfl A, Kubin G (2004) Speech watermarking for air traffic control. In: Proceedings of 12th European signal processing conference. European Association for Signal, Speech and Image Processing, pp 1653–1656
19. Harjito B, Song H, Potdar V, Chang E, Miao X (2010) Secure communication in wireless multimedia sensor networks using watermarking. In: Proceedings of the 2010 4th IEEE international conference on digital ecosystems and technologies (DEST), pp 640–645
20. Hatada M, Sakai T, Komatsu N, Yamazaki Y (2002) Digital watermarking based on process of speech production. Proc SPIE 4861:258–267
21. Hofbauer K, Kubin G, Kleijn WB (2009) Speech watermarking for analog flat-fading bandpass channels. IEEE Transactions on Audio, Speech and Language Processing 17:1624–1637
22. Huang HC, Fang WC (2010) Metadata-based image watermarking for copyright protection. Simulation Modelling Practice and Theory 18(4):436–445
23. Huang HC, Chu SC, Pan J, Huang CY, Liao,BY (2011) Tabu search based multi-watermarks embedding algorithm with multiple description coding. Inf Sci 181(16):3379–3396
24. Kay SM (1988) Modern spectral estimation: theory and application. Prentice-Hall, Englewood Cliffs
25. Konrad H, Horst H, Gernot K (2005) Speech watermarking for the VHF radio channel. In: 4th Eurocontrol innovative research workshop. https://www.spsc.tugraz.at/basilic/Publications/2005/HHK05/
26. Kundur D (1999) Multiresolution digital watermarking: algorithms and implications for multimedia signals. Ph.D. thesis, Graduate Department of Electrical and Computer Engineering, University of Toronto
27. Lacy J, Quackenbusch S, Reibman AR, Shur D, Snyder J (1998) On combining watermarking with perceptual coding. In: IEEE ICASSP'98, Seattle (USA), vol 6, pp 3725–3728
28. Lin CC, Tsai WH (2004) Secret image sharing with steganography and authentication. J Syst Softw 73:405–414
29. Liu CH, Chen OTC (2004) A fragile speech watermarking scheme with recovering speech contents. In: The 47th IEEE international midwest symposium on circuits and systems, pp II165–II168
30. Liu JX, Lu ZM, Luo H A (2009) CELP-speech information hiding algorithm based on vector quantization. In: Proceedings of the fifth international conference on information assurance and security, pp 75–78
31. Lu ZM, Yan B, Sun SH (2005) Watermarking combined with CELP speech coding for authentication. IEICE Trans Inf Syst E88-D(2):330–334
32. Ma L, Wu Zj, Hu Y, Yang W (2007) An information-hiding model for secure communication. In: Huang DS, Heutte L, Loog M (eds) Advanced intelligent computing theories and applications. With aspects of theoretical and methodological issues. Lecture notes in computer science, vol 4681. Springer, Berlin, pp 1305–1314

33. Oppenheim AV, Schafer RW, Buck JR (1999) Discrete-time signal processing, 2nd edn. Prentice-Hall, Upper Saddle River
34. Painter T, Spanias A (2000) Perceptual coding of digital audio. Proc IEEE 88:451–515
35. Park CM, Thapa D, Wang GN (2007) Speech authentication system using digital watermarking and pattern recovery. Pattern Recogn Lett 28(8):931–938
36. Quatieri TF (2002) Discrete-time speech signal processing: principles and practice. Prentice Hall, Englewood Cliffs
37. Ramamurthy K, Spanias A (2010) MATLAB software for the code excited linear prediction algorithm: the federal standard-1016. Morgan and Claypool Publishers
38. Ruiz FJ, Deller JRJ (2000) Digital watermarking of speech signals for the national gallery ofthe spoken word. In: IEEE international conference on acoustics speech and signal processing, vol 3, pp 1499–1502
39. Sakaguchi S, Arai T, Murahara Y (2000) The effect of polarity inversion of speech on human perception and data hiding as an application. In: IEEE international conference on acoustics speech and signal processing, vol 2, pp 917–920
40. Saraswathi S (2010) Speech authentication based on audio watermarking. Int J Inf Technol 16(1):34–43
41. Singh J, Garg P, De AN (2009) A combined watermarking and encryption algorithm for secure voip. Information Security Journal: A Global Perspective 18:99–105
42. Stallings W (2005) Cryptography and network security principles and practices, 4th edn. Prentice Hall, Upper Saddle River
43. Unoki AM, Imabeppu K, Hamada D, Haniu A, Miyauchi R (2011) Embedding limitations with digital-audio watermarking method based on cochlear delay characteristics. Journal of Information Hiding and Multimedia Signal Processing 2(1):337–355
44. Wu CP, Kuo CCJ (2002) Fragile speech watermarking based on exponential scale quantization for tamper detection. In: Proceeding of the IEEE international conference on acoustic, speech and signal processing, pp 3305–3308
45. Yan B, Lu ZM, Sun SH, Pan JS (2005) Speech authentication by semi-fragile watermarking. In: Proceedings of international workshop on intelligent information hiding and multimedia signal processing. Lecture notes in computer science, vol 3683, pp 497–504
46. Yardyimci Y, Cetin AE, Ansari R (1997) Data hiding in speech using phase coding. In: Proceedings of Eurospeech 97, Rhodes, Greece, vol 3, pp 1679–1682

Bin Yan was born in Shandong Province, People's Republic of China in 1973. He received his Ph.D. degree from Harbin Institute of Technology in 2007. He is now with the Department of Communication Engineering, School of Information and Electrical Engineering, Shandong University of Science and Technology. His research interest include statistical signal processing, information hiding and multimedia security.

Yin-Jing Guo was born in Shandong Province, People's Republic of China in 1966. He received his Ph.D. degree from Beijing Institute of Technology in 2004. He is now with the Department of Communication Engineering, School of Information and Electrical Engineering, Shandong University of Science and Technology. His research interest include communication signal processing, multimedia security and communication networks.