Steganography integrated into linear predictive coding for low bit-rate speech codec

Peng Liu[1,2] · Songbin Li[1,2] · Haiqiang Wang[3]

Abstract The extensive use of Voice over IP (VoIP) applications makes low bit-rate speech stream a very suitable steganographic cover media. To incorporate steganography into low bit-rate speech codec, we propose a novel approach to embed information during linear predictive coding (LPC) process based on Matrix Embedding (ME). In the proposed method, a mapping table is constructed based on the criterion of minimum distance of Linear-Predictive-Coefficient-Vectors, and embedding position and template are selected according to a private key so as to choose the cover frames. The original speech data of the chosen frames are partially encoded to get the codewords for embedding and then the codewords that need to be modified for embedding are selected according to the secret bits and ME algorithm. The selected codeword will be changed into its best replacement codeword according to the mapping table. When embedding k ($k > 1$) bits into $2^k - 1$ codewords, the embedding efficiency of our method is k times as that of LPC-based Quantization Index Modulation method. The performance of the proposed approach is evaluated in two aspects: distortion in speech quality introduced by embedding and security under steganalysis. The experimental results demonstrate that the proposed approach leads to a better performance with less speech distortion and better security.

✉ Songbin Li
  lisb@dsp.ac.cn

  Peng Liu
  liup@dsp.ac.cn

  Haiqiang Wang
  haiqianw@usc.edu

[1] Haikou Laboratory, Institute of Acoustics, Chinese Academy of Sciences, Haikou 570105, China

[2] National Network New Media Engineering Research Center, Beijing 100190, China

[3] University of Southern California, Los Angeles, CA, USA

1 Introduction

Information hiding is the technique of covert communication via public channels. The basic model of information hiding system dates back to the prisoners' problem [24] and the subliminal channel [25, 26] proposed by Simmons. Digital multimedia signals, like video, images, and audio, are prevalent in our daily life, which expose a new avenue of covert communication, namely, steganography. The Voice over IP (VoIP) applications are widely used in real-time communication on the Internet. The network traffic of VoIP stream is gigantic, and its growth rate has already been larger than that of Time Division Multiplexing systems [27]. Ubiquitousness of VoIP stream makes itself a very suitable carrier for information hiding.

In order to reduce redundancy in original speech stream, several low bit-rate speech compression algorithms, such as G.723.1 and G.729, were specially proposed for VoIP by International Telecommunication Union (ITU), leading to the extensive use of low bit-rate speech codecs in VoIP applications. Therefore, it is essential to study information hiding techniques in low bit-rate compressed speech streams.

However, it remains a challenging problem to embed information in low bit-rate VoIP streams. Firstly, as mentioned before, redundancy in low bit-rate codec has already been eliminated via advanced codec, but steganographic methods rely highly on such redundancy to embed information. Secondly, few redundancy as there are, embedding information will introduce discernible distortion in low bit-rate compression systems, which makes embedding process vulnerable under corresponding steganalysis. Finally, VoIP based steganographic algorithms are different from image or video based steganographic algorithms, the former has to fulfill real-time requirement of VoIP communications [6].

Existing steganography algorithms in low bit-rate VoIP streams can be divided into three categories. Algorithms in the first category embed information by directly modifying bit plane in the compressed speech streams. For example, both Hui et al.'s and Liu et al.'s algorithms [8, 15] conducted Least Significant Bit (LSB) replacement method in VoIP streams encoded by G.729 codec. In [9, 11, 28], LSB replacement steganography algorithms were applied in G.723.1 and G.711 compressed streams, respectively. Huang et al. [7] and Roselinkiruba et al. [23] embedded information in inactive frames of VoIP bit streams. These algorithms have the same shortcoming, i.e. embedding process is independent of speech encoding process, and thus result in discernible distortion in temporal domain. In order to reduce distortion introduced in embedding process, algorithms in the second category propose to embed information in the prediction process of the Long Term Predictor (LTP) in the speech codec, and those in the third category conduct embedding in the Short Term Predictor (STP). For example, the pitch modulation based steganography algorithms [6, 33] embedded secret information when the encoder estimates the pitch of the speech sub-frame, and the Quantization Index Modulation (QIM) based steganography [32] hided the secret data during Vector Quantization (VQ) process of linear prediction coefficients.

Common low bit-rate VoIP speech codecs are mostly based on the Linear Predictive Coding (LPC) model. In addition, the linear predictive analysis-by-synthesis coding (ABS-LPC) model is widely used in low bit-rate codecs which aims to further minimize the distortion by decoding the encoded signal and choosing the codeword with the least error. Taking the advantage of ABS, embedding in the vector quantization process of LPC coefficients will

significantly decrease distortion in the follow-up encoding process. Therefore, this paper focuses on the LPC-based steganography in low bit-rate speech streams.

Image steganography based on VQ process has been well studied in recent years [1, 12, 17, 19]. For example, Chang et al. [1] proposed a data hiding algorithms which achieved high embedding capacity in VQ indices with neighboring correlations. Rahmani et al. [19] suggested a reversible data hiding scheme for VQ-compressed images based on search order coding. Comparatively, VQ-based steganography in low bit-rate speech streams is an emerging topic. Being the first approach to embed secrets into VQ process of linear prediction coding for low bit-rate speech codec, Xiao et al. [32] adopted the QIM steganography method [2] to conduct embedding by modifying the quantization vectors. The whole codebook was divided into two parts, representing '0' or '1', respectively. When a secret bit was embedded, the corresponding part of the codebook was used. On the decoding side, the secret bit was extracted by checking which part of codebook the codeword belongs to. The core problem in QIM steganography algorithm is to find an ideal codebook partition scheme so as to minimize the distortion. Chiang et al. [3] suggested a codebook partition algorithm based on codeword clustering. In [14, 30], the authors introduced partition algorithms based on a genetic algorithm. Lu et al. [16] proposed a method which used public codeword group to reduce the distortion. Xiao et al. [32] proposed an algorithm called Complementary Neighbor Vertices (CNV), which guaranteed that every codeword was separated from its nearest neighbor and achieved less distortion than any other partition methods.

As QIM hides data by modulating the codewords, the distribution characteristics of the codewords would be changed inevitably. Too many modifications introduced by QIM would make the steganography algorithm easily detected [13]. Therefore, it becomes an important issue to reduce the amount of necessary alterations under the same embedding rate. Matrix Embedding (ME) is a method of high embedding efficiency and imperceptibility, which was introduced by Crandall [4]. ME has gained wide attention after Westfeld gave the first implementation of matrix encoding in F5 algorithm [31]. Since higher embedding efficiency translates into better hiding security, ME becomes a good choice for secrets embedding. However, ME is based on scalar replacement, which cannot be applied to vector directly. Therefore, we propose a method that extends the scope of ME method to vectors by combining it with a mapping table, called MEL (Matrix Embedding in LPC). Extensive experimental results demonstrate that our approach outperforms other LPC based embedding algorithms with regard to less speech distortion and better security.

The rest of the paper is organized as follows. Section 2 describes the information hiding and extracting algorithms of the MEL method. Experimental results are discussed in Section 3. Finally, Section 4 concludes with a summary.

2 Information hiding and extracting

This paper proposes to embed information through the replacement of the codewords associated to the filter coefficients of the Linear Predictive Coding by other codewords from the same codebook, under a minimum distance criterion. The block diagram is shown in Fig. 1.

In the embedding process, a mapping table is constructed based on the criterion of minimum distance of Linear- Predictive-Coefficient-Vectors regained from the codewords before and after embedding. Secondly, embedding position and template are selected according to the private key so as to get the frames for embedding. Thirdly, the original speech data of the chosen frames are partially encoded to get the codewords for embedding. Then, codewords are translated into binary sequences based on the mapping table,
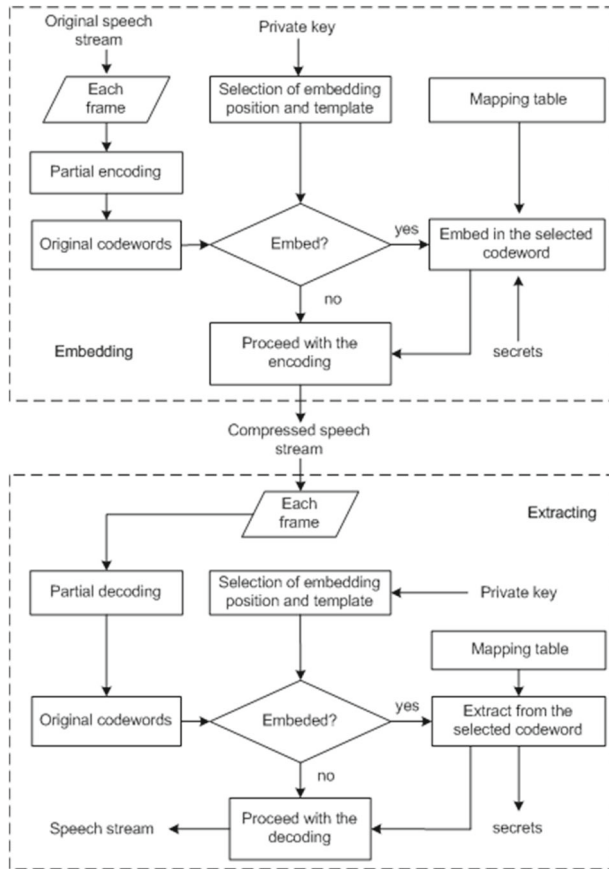
Fig. 1 Block diagram of the proposed approach

and codewords that need to be modified for embedding are selected through ME and changed according to the mapping table. Finally, compressed speech stream is achieved by proceeding with the encoding.

In the decoding process, the original speech frames are partly decoded to obtain the original codewords. Then decide whether this point is embedded or not according to the embedding position and template. If so, secret information is extracted according to the mapping table.

2.1 Construction of the mapping table

Linear predictive coding is one of the most useful speech analysis techniques for encoding good quality speech at a low bit rate. It is used mostly in speech processing and audio signal processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. The *m-th* order LPC filter can be described as follows:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{m} a_i z^{-i}}. \tag{1}$$

Let $\mathbf{a} = \{a_1, a_2, \cdots, a_m\}$ denotes the Linear Predictive Coefficient Vector (LPCV) of the LPC filter. Usually, the LPC coefficients are converted to Line Spectral Pair (LSP) coefficients or Line Spectral Frequency (LSF) coefficients for quantization, which can be transformed to each other.

The basic idea of LPC-based speech coding is to quantify the LPCV into a codeword set using the fixed codebooks of speech codec. In order to enhance the compression efficiency, the codeword set is eventually represented by a LPC quantization index set $\mathbf{x} = \{x_1, x_2, \cdots, x_i, \cdots, x_n\}$, where x_i points to the corresponding codeword of the *i-th* fixed codebook. As for decoding, the LPCV $\mathbf{a}$ can be generated based on quantization index set $\mathbf{x}$ and the fixed codebooks. Suppose that the decoding process from $\mathbf{x}$ to $\mathbf{a}$ can be denoted by a function f, namely, $\mathbf{a} = f(x_1, x_2, \cdots, x_n)$, which is defined by codecs. For example, in G.729 codec, the quantization is achieved through three fixed codebooks after the LPC coefficients transforming into the LSF coefficients. Each LSF coefficient s_i can be obtained from the sum of two codebooks

$$s_i = \begin{cases} C1_i(x_1) + C2_i(x_2) & i = 1, \cdots, 5 \\ C1_i(x_1) + C3_{i-5}(x_3) & i = 6, \cdots, 10 \end{cases} \qquad (2)$$

where $C1$, $C2$, and $C3$ are the fixed codebook indices [10]. At last, the LPCV is obtained by the transformation of LSF coefficients.

Assume that we need to modify the *j-th* quantization index in $\mathbf{x}$ from $\{x_1, x_2, \cdots, x_j, \cdots, x_n\}$ into $\{x_1, x_2, \cdots, x_j', \cdots, x_n\}$, $1 \leq j \leq n$. In order to quantify the difference of LPCV before and after embedding, a loss function is defined as

$$\Psi\left(\mathbf{x}, j, x_j'\right) = \left\| f\left(x_1, x_2, \cdots, x_j, \cdots, x_n\right) - f\left(x_1, x_2, \cdots, x_j', \cdots, x_n\right) \right\| \qquad (3)$$

where $\Psi\left(\mathbf{x}, j, x_j'\right)$ reflects the impact on the carrier speech introduced by the information embedding process. The smaller the $\Psi\left(\mathbf{x}, j, x_j'\right)$ is, the safer the steganography becomes. Where $\|*\|$ stands for the Euclidean norm.

Suppose the codebook of the *j-th* quantization index x_j is denoted as C. For any index $x_j \in C$, there exists an index $y \in C$ and $y \neq x_j$ satisfying that

$$y = \operatorname*{arg\,min}_{x_j' \in C, x_j' \neq x_j} \Psi\left(\mathbf{x}, j, x_j'\right). \qquad (4)$$

We define y as the best replacement choice for x_j and further denote $\{x_1, x_2, \cdots, y, \cdots, x_n\}$ as the best-replacement-set for $\{x_1, x_2, \cdots, x_j, \cdots, x_n\}$ when modifying the *j-th* quantization index. Each quantization index set has n different best-replacement-sets at most.

Quantization index space consists of all possible quantization index. The mapping table can be constructed by two steps. Firstly, go through every point in quantization index space and get all the corresponding best-replacement-sets. Secondly, assign each point with a label of '0' or '1' and make sure that each point has different labels with the points corresponding to its best-replacement-sets. This process involves four steps:

1. Let V be the set of points in quantization index space and E be the set of edges. Initialize a graph $G = (V, E)$ with isolated points belonging to V.
2. For every point $P \in V$, insert undirected edges between P and the points corresponding to its best-replacement-sets into E.
3. Delete duplicated edges between any two points.
4. Color the graph with 2 colors, so as two points of an edge have different color, and each color corresponds to a label of '0' or '1' .

For example, for a codec with three fix codebooks, we use $(0, 0, 0)$ as the origin to establish a three dimensional Cartesian coordinate space. Quantization index set $\{x_1, x_2, x_3\}$ can be seen as a point (x_1, x_2, x_3) in this space. According to graph theory [29], the points with two identical coordinates in $G(V, E)$ are 2-colorable. So we make the origin as a starting point, and color the points on the x-axis direction, as shown in Fig. 2a. Then we color along the y-axis direction with the colored points as starting points, as shown in Fig. 2b. Finally do it along the z-axis direction, as shown in Fig. 2c.

Since the crux of the mapping algorithm is to calculate the value of loss function. We will take G.729 and G.723.1 codecs as examples to introduce the solving process.

– Loss Function in G.729

In G.729, the difference between the predicted and computed LSF coefficients is quantized into three codewords and a flag L_0 that defines which MA (Moving Average) predictor to use. The order of the linear prediction filters is 10. The LSP coefficients **s** for the current frame m can be obtained from the weighted sum of previous quantizer outputs $\mathbf{q}^{m-k}$ and the current quantizer output $\mathbf{q}^m$ by

$$s_i^m = \left(1 - \sum_{k=1}^{4} c_{i,k}\right) q_i^m + \sum_{k=1}^{4} c_{i,k} q_i^{m-k} \quad i = 1, \cdots, 10 \tag{5}$$

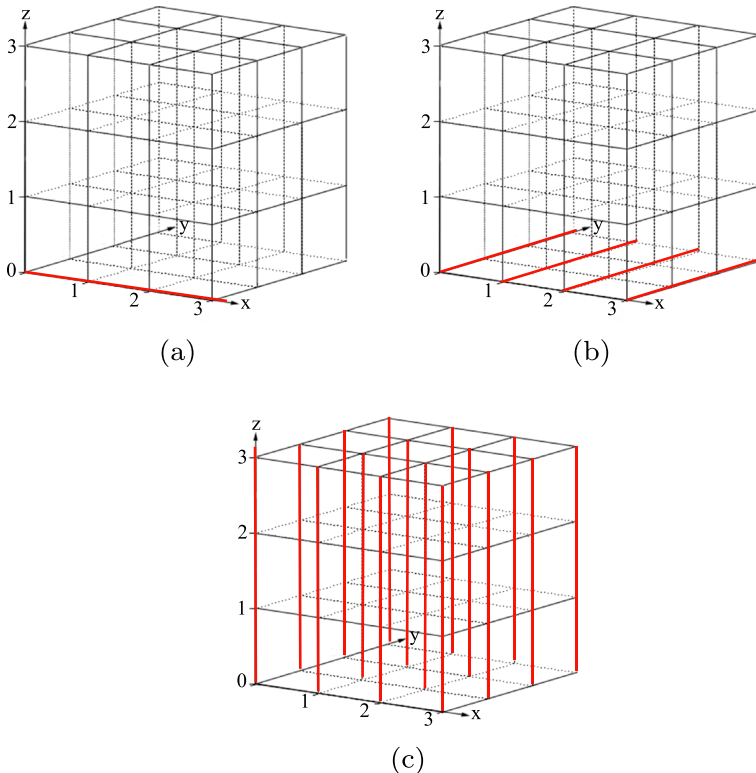(a)   (b)

(c)

Fig. 2 Illustration of coloring process

where $c_{i,k}$ is the coefficient of the switched MA predictor. Since the weighted sum of previous quantizer outputs $\mathbf{q}^{m-k}$ will not be changed by the replacement of the current quantizer output $\mathbf{q}^m$, the LSF coefficients $\mathbf{s}^{m'}$ obtained from the embedded codewords can be calculated by

$$s_i^{m'} = \left(1 - \sum_{k=1}^{4} c_{i,k}\right) q_i^{m'} + \sum_{k=1}^{4} c_{i,k} q_i^{m-k} \quad i = 1, \cdots, 10 \tag{6}$$

where $\mathbf{q}^{m'}$ denotes the coefficients obtained from the embedded codewords.

Since LPC coefficients and LSF coefficients can be transformed to each other, we use $\mathbf{s}^m$ here to calculate the distortion. Hence, the calculation of the distance of LPC coefficient vectors are equivalent to calculating $\|\mathbf{s}^m - \mathbf{s}^{m'}\|$, where $\mathbf{s}^{m'}$ denotes the LSF coefficient vector obtained from the embedded codewords. $\|\mathbf{s}^m - \mathbf{s}^{m'}\|$ can be calculated by

$$\left\|\mathbf{s}^m - \mathbf{s}^{m'}\right\| = \sqrt{\sum_{i=1}^{10} (s_i^m - s_i^{m'})^2}$$

$$= \sqrt{\sum_{i=1}^{10} \left[\left(1 - \sum_{k=1}^{4} c_{i,k}\right)\left(I_i^m - I_i^{m'}\right)\right]^2}. \tag{7}$$

One thing to note is that there are two MA predictors in G.729 codec, so we should calculate a mapping table for each predictor and choose the corresponding mapping table according to L_0 for embedding.

- Loss Function in G.723.1

In G.723.1, the order of the linear prediction filters is 10. The codewords are mapped to the residual coefficients $\mathbf{e}$ according to the codebooks. The LSP coefficients $\mathbf{p}$ can be obtained by

$$p_i = e_i + p_{DC} + \bar{p}_i \quad i = 1, \cdots, 10 \tag{8}$$

where p_{DC} is the long-term direct current component, $\bar{\mathbf{p}}$ denotes the predicted coefficients.

Since LPC coefficients and LSP coefficients can be transformed to each other, we use $\mathbf{p}$ here to calculate the vector distance. Hence, the calculation of the distance of LPC coefficients are equivalent to calculating $\|\mathbf{p}^m - \mathbf{p}^{m'}\|$, where $\mathbf{p}^{m'}$ denote the LSP coefficients obtained from the embedded codewords. $\|\mathbf{p}^m - \mathbf{p}^{m'}\|$ can be calculated by

$$\left\|\mathbf{p}^m - \mathbf{p}^{m'}\right\| = \sqrt{\sum_{i=1}^{10} (e_i - e_i')^2} \tag{9}$$

where e_i' denotes the i-th coefficient obtained from the embedded codewords.

2.2 Embedding and extracting algorithms

The constructed mapping table can be denoted by mapping function $Map(\mathbf{x}, j, t)$, where t represents the operation type. When $t = 0$, $Map(\mathbf{x}, j, 0)$ returns the binary sequences corresponds to $\mathbf{x}$. By this time, j is not used. When $t = 1$, $Map(\mathbf{x}, j, 1)$ returns the codeword which is used to replace the j-th codeword in $\mathbf{x}$. Once the mapping table is constructed, the embedding algorithm of MEL method involves the following procedure:

1. Map the codewords for embedding into binary sequences according to the mapping table. Suppose x_j^m represents the j-th codeword in the m-th frame. $Q(x_j^m)$ denotes the binary value which x_j^m corresponds to. Thus, $Q(x_j^m)$ can be obtained by $Map(\mathbf{x}^m, j, 0)$, where $\mathbf{x}^m$ is the codeword set of the m-th frame.
2. Determine the position that needs to be modified for embedding by ME [31] . Given a set of $n = 2^k - 1, (k > 1)$ embedding coefficients, ME method can embed k bits by modifying 1 bit at most.
3. Suppose that we need to modify x_j^m. Modify it to $Map(\mathbf{x}^m, j, 1)$.

Table 1 shows the implementation of MEL method in G.723.1 codec, where $n = 3, k = 2$, and the codeword set is $\{x_1^m, x_2^m, x_3^m\}$. Without loss of generality, let $Q_1 = Q(x_1^m)$, $Q_2 = Q(x_2^m)$, $Q_3 = Q(x_3^m)$ The symbol $\oplus$ means XOR operation.

The extracting algorithm of MEL method involves two steps. Firstly, determine the codewords used for embedding. Secondly, map the codewords into binary sequences according to the mapping table using the mapping function $Map(\mathbf{x}^m, j, 1)$. The time complexity of MEL method is $O(1)$, which can satisfy the real-time requirement of VoIP communications. For example, when $n = 3$ and $k = 2$, it only needs twice XOR operations and once table-mapping operation to embed two secret bits into three cover vectors. As for extracting the two secret bits, one table-mapping operation is enough.

2.3 Embedding positions and templates

2.3.1 Embedding positions selection

In LPC-based speech coder, the original speech stream is encoded frame by frame using LPC analysis. Therefore, the compressed speech stream can be seen as a filter sequence.

Table 1 An implementation of secrets mapping in G.723.1

Secrets to embed	Embedding conditions	Embedding rule
$w = 00$	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 0$	No change
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 0$	Change x_1^m to $Map(\mathbf{x}^m, 1, 1)$
	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 1$	Change x_2^m to $Map(\mathbf{x}^m, 2, 1)$
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 1$	Change x_3^m to $Map(\mathbf{x}^m, 3, 1)$
$w = 01$	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 0$	Change x_2^m to $Map(\mathbf{x}^m, 2, 1)$
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 0$	Change x_3^m to $Map(\mathbf{x}^m, 3, 1)$
	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 1$	No change
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 1$	Change x_1^m to $Map(\mathbf{x}^m, 1, 1)$
$w = 10$	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 0$	Change x_1^m to $Map(\mathbf{x}^m, 1, 1)$
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 0$	No change
	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 1$	Change x_3^m to $Map(\mathbf{x}^m, 3, 1)$
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 1$	Change x_2^m to $Map(\mathbf{x}^m, 2, 1)$
$w = 11$	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 0$	Change x_3^m to $Map(\mathbf{x}^m, 3, 1)$
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 0$	Change x_2^m to $Map(\mathbf{x}^m, 2, 1)$
	$Q_1 \oplus Q_3 = 0, Q_2 \oplus Q_3 = 1$	Change x_1^m to $Map(\mathbf{x}^m, 1, 1)$
	$Q_1 \oplus Q_3 = 1, Q_2 \oplus Q_3 = 1$	No change

Each coefficient of LPC filter needs to be encoded by Vector Quantization (VQ). Split vector quantization is commonly used to improve the coding efficiency, which contains the procedure of VQ several times. After each VQ process, one codeword is obtained. All of the codewords in the compressed speech stream constitute a steganography space.

Let matrix A denote the steganography space. Suppose the number of frames contained in compressed speech stream is n and each frame contains m codewords. Then matrix A can be represented as below

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix}_{n \times m}. \tag{10}$$

Usually, there is no need to embed in all of the cover frames. Therefore, the selection of embedding positions is needed. Since the length of speech stream is indeterminate, in order to ensure the embedding rate, we partition matrix A into blocks with length of p and embed in each sub-block according to the embedding rate. Each sub-block is $p \times m$ order matrix, where p is a preset value. So we can get the partitioned matrix $A = [\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_m}]^T$.

Firstly, select i cover frames from each sub-block according to the desired embedding rate. So we can get the matrix for embedding as

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{i,1} & b_{i,2} & \cdots & b_{i,m} \end{bmatrix}_{i \times m}. \tag{11}$$

Secondly, partition matrix $B = [\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_m}]^T$ into blocks. Each sub-block is a $q \times m$ order matrix, where q is a preset value. Each sub-block is called the Smallest Embedding Unit (SEU). We use matrix U to denote a SEU. The acquisition process of a SEU is shown in Fig. 3.

2.3.2 Embedding templates selection

In order to further enhance the security of the steganography algorithm, we design eight embedding templates in this paper. Each embedding template corresponds to a selection approach of the embedding sequence from matrix U. The black dots in Fig. 4 denote the starting position for each embedding sequence. The arrows represent the direction of extraction. Besides, the dashed lines are used to indicate the connecting relations of arrows. After the embedding template is determined, we can get one or more codeword sequences for embedding. For example, when $q = 7$ and $m = 3$, we can get the SEU as below:

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{2,3} \\ \vdots & \vdots & \vdots \\ u_{7,1} & u_{7,2} & u_{7,3} \end{bmatrix}_{7 \times 3}. \tag{12}$$

The embedding codeword sequences correspond to different selection templates are shown in Table 2. The parameters p and q, embedding position, and template can be selected by a private key using a pseudorandom number generator [31] or other algorithms, which is not the focus of this paper.
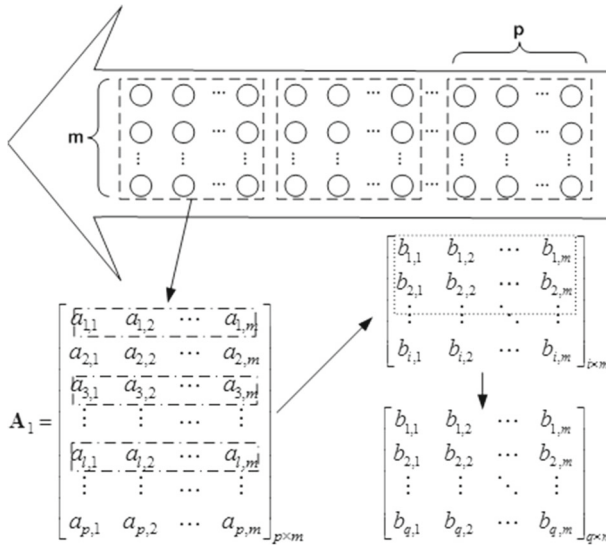
Fig. 3 The acquisition process of a Smallest Embedding Unit

3 Results and discussion

The performance of the proposed algorithm is evaluated in two aspects: distortion in speech quality introduced by embedding and security under steganalysis. This paper focuses on the LPC-based steganography in low bit-rate speech streams. The comparison is conducted between the proposed algorithm and LPC-QIM method [32].

3.1 Introduction of test samples

The same dataset used in [13] was used to conduct the experiments. These samples are classified into four groups, Chinese Speech Man (CSM), Chinese Speech Woman (CSW),
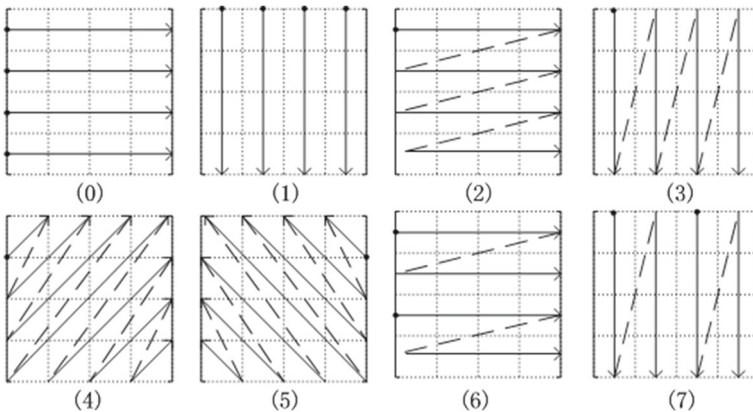

Fig. 4 Eight selection templates for embedding in SEU

Table 2 The embedding codeword sequences corresponding to different selection templates

Selection template	Corresponding codeword sequences
0	$(u_{1,1}, u_{1,2}, u_{1,3}), \cdots, (u_{7,1}, u_{7,2}, u_{7,3})$
1	$(u_{1,1}, u_{2,1}, \cdots, u_{7,1}), \cdots, (u_{1,3}, u_{2,3}, \cdots, u_{7,3})$
2	$(u_{1,1}, u_{1,2}, u_{1,3}, u_{2,1}, u_{2,2}, \cdots, u_{7,1}, u_{7,2}, u_{7,3})$
3	$(u_{1,1}, u_{2,1}, \cdots, u_{7,1}, u_{1,2}, u_{2,2}, \cdots, u_{6,3}, u_{7,3})$
4	$(u_{1,1}, u_{2,1}, u_{1,2}, u_{3,1}, \cdots, u_{3,3}, u_{6,1}, u_{5,2}, \cdots, u_{6,3}, u_{7,3})$
5	$(u_{1,3}, u_{2,3}, u_{1,2}, u_{3,3}, \cdots, u_{3,1}, u_{6,3}, u_{5,2}, \cdots, u_{6,1}, u_{7,1})$
6	$(u_{1,1}, u_{1,2}, \cdots, u_{4,2}, u_{4,3})(u_{5,1}, u_{5,2}, \cdots, u_{7,2}, u_{7,3})$
7	$(u_{1,1}, u_{2,1}, \cdots, u_{6,2}, u_{7,2})(u_{1,3}, u_{2,3}, \cdots, u_{6,3}, u_{7,3})$

English Speech Man (ESM), and English Speech Woman (ESW). The duration of each sample in these groups is 10 seconds and each speech is stored in PCM format. The detailed information is shown in Table 3.

We choose two speech codecs, G.729 and G.723.1, to implement the embedding. Before the experiments, we firstly prepare the stego files of the two methods with the Embedding Rate (ER) of 20, 40, 60, 80 and 100 %. We use template 0 as the selection for MEL. For G.729, the parameter p is set to 1000 and q is set to 5. For G.723.1, the parameter p is set to 333 and q is set to 3. In order to keep the same data embedding rate, as for the embedding process of LPC-QIM method, two of the three embedding positions in each frame are chosen randomly for embedding. The secret messages are chosen from the 'Etymology' section in Wikipedia [5]. After the above processing, we obtain 80 groups of stego speech files and 8 groups of cover files.

3.2 Influence of steganography on speech quality

Speech quality evaluation methods can be classified into two categories: subjective methods and objective methods. The most common approach to conduct subjective quality assessment is the so-called Mean Opinion Score (MOS) method, which is defined by the International Telecommunication Union (ITU) in the ITU-T P.800 [18]. In this method, a group of listeners are asked to score the general quality of a given set of speech samples and the average result indicates the quality of signals. This method is such widely recognized that even the after-coming objective methods turn their evaluation results into MOS scores. In this paper, we take MOS method to conduct the subjective test.

Objective speech quality evaluation methods can be mainly categorized as time-domain parameter based methods, spectrum-distance based methods, hearing-model based methods,

Table 3 The information of test samples

Samples	Content	Number	Format	Sampling frequency	Digitalizing bit
CSM	Reading	500	PCM	8KHz	16bit
CSW	Reading	532	PCM	8KHz	16bit
ESM	Report	818	PCM	8KHz	16bit
ESW	Report	824	PCM	8KHz	16bit

judgment-model based methods. Among them, the methods based on time-domain parameter are the simplest and most straight-forward ones and the methods based on hearing-model have the best evaluation performance. Consequently, our paper uses these two kinds of methods to conduct the objective speech quality evaluation. Signal-to-Noise Ratio (SNR) and Peak Signal-to-Noise Ratio (PSNR) are two representative time-domain parameter based evaluation methods. We apply the PSNR because PSNR can intuitively compare the differences between the original signal and the embedded signal, compared to SNR. The most widely used hearing-model based methods are proposed in ITU-T P.862 [22] and ITU-T P.563 [20]. ITU-T P.862 defines an end-to-end speech quality assessment named Perceptual Evaluation of Speech Quality (PESQ), which is used to evaluate the quality of experience of communication system while taking end-to-end network delay, noise and packet-loss into consideration. ITU-T P.563 is the newest single-end speech quality evaluation method. In a practical application, listeners will never have the knowledge from both ends and single-end test is conducted in the subjective test. In coordinate with the subjective test, single-end test method followed P.563 criterion is applied in objective evaluation.

3.2.1 Subjective test

In this test, 28 randomly selected subjects with normal hearing took necessary training at first. The overall 88 groups of speech data have 58,828 samples with a total time length over 163 hours, which is too much for the subjects. Thus, in our subjective test, we choose 30 samples from each group with total time duration of 26,400 seconds. We conducted the test in a quiet laboratory and let the subjects score these samples using the five-point absolute category rating listening quality scale [18] where 1 denotes the lowest perceived audio quality, and 5 denotes the highest perceived audio quality.

Tables 4 and 5 show the Mean Opinion Score - Listening Quality Subjective (MOS-LQS) assessment results for the stego speech files and the cover speech files with data embedding by MEL method, processed by G.729 and G.723.1 respectively. The group with 0 % embedding rate denotes the results of the cover files. Table 4 shows that the absolute value of Change Rate (CR) of MOS-LQS value increases with the increasing of embedding rate. It can be seen that the distortion introduced by MEL method is 2.44 % at most.

Table 5 shows that the absolute value of change rate of MOS-LQS value generally increases with the increasing of embedding rate, but the trend is not as obvious as that in Table 4. This is because the frame length in G.723.1 is longer than G.729. 10-second speech sample can be coded into 1000 frames by G.729 codec but only 333 frames by G.723.1. Under the same embedding rate, fewer frames translate into fewer alterations which will

Table 4 MOS-LQS test results of samples processed by G.729

ER	CSM		CSW		ESM		ESW	
(%)	Value	CR (%)	Value	CR (%)	Value	CR (%)	Value	CR (%)
0	3.375	0	2.97738	0	2.87024	0	2.90119	0
20	3.36724	−0.23	2.96875	−0.29	2.86106	−0.32	2.89887	−0.08
40	3.36353	−0.34	2.95803	−0.65	2.85905	−0.39	2.88523	−0.55
60	3.34564	−0.87	2.95118	−0.88	2.84670	−0.82	2.87334	−0.96
80	3.33619	−1.15	2.94076	−1.23	2.82690	−1.51	2.85825	−1.48
100	3.31189	−1.87	2.91456	−2.11	2.80021	−2.44	2.83591	−2.25

Table 5 MOS-LQS test results of samples processed by G.723.1

ER	CSM		CSW		ESM		ESW	
(%)	Value	CR (%)	Value	CR (%)	Value	CR (%)	Value	CR (%)
0	3.41786	0	3.05477	0	2.93929	0	2.94643	0
20	3.41991	0.06	3.03644	−0.60	2.93841	−0.03	2.94584	−0.02
40	3.41478	−0.09	3.04499	−0.32	2.94135	0.07	2.94466	−0.06
60	3.41239	−0.16	3.04377	−0.36	2.93753	−0.06	2.94466	−0.06
80	3.41239	−0.16	3.03919	−0.51	2.93870	−0.02	2.94142	−0.17
100	3.40966	−0.24	3.04102	−0.45	2.93694	−0.08	2.93965	−0.23

lead to less distortion. The results shown in Tables 4 and 5 clearly indicate that the distortion introduced by MEL method is small.

Tables 6 and 7 illustrate the comparison between MEL and LPC-QIM method based on MOS-LQS values, processed by G.729 and G.723.1 respectively. In order to compare the two methods, the MOS-LQS test results of LPC-QIM method are used as benchmark. Table 6 shows that the MOS-LQS values increase in every group, which means the distortion introduced by MEL method is less than that of LPC-QIM. Table 7 shows that the MOS-LQS values increase in 19 out of 20 groups. The Observable differences are not obvious. This is because the linear predictive analysis-by-synthesis coding (ABS-LPC) model is used in low bit-rate codecs, which will minimize the distortion by decoding the encoded signal. Due to its higher embedding efficiency, the MEL method is better than the LPC-QIM method in that lower average probability of changing for each bit when embedding secret message into cover vectors, especially when the number of frames is higher. From the analysis above, it is clear that MEL method has less impact on the quality of cover speech than the LPC-QIM method.

3.2.2 Objective test

PSNR is an approximation to human perception of reconstruction quality. It is an expression for the ratio between the maximum possible power of a signal and the power of distorting noise. The signal in this case is the original speech data, and the noise is the error introduced by compression and steganography. To make the plot much clearer, we show only part of the samples. PSNR test results are shown in Figs. 5 and 6. The embedding rate of the stego

Table 6 The change in MOS-LQS value on samples processed by G.729

ER	%Change in MOS-LQS				
(%)	CSM (%)	CSW (%)	ESM (%)	ESW (%)	AVERAGE (%)
20	0.15	0.28	0.26	0.04	0.18
40	0.36	0.37	0.33	0.29	0.34
60	0.38	0.51	0.31	0.53	0.43
80	0.38	1.15	0.42	0.77	0.68
100	0.63	1.22	0.45	1.32	0.91

Table 7 The change in MOS-LQS value on samples processed by G.723.1

ER	%Change in MOS-LQS				
(%)	CSM (%)	CSW (%)	ESM (%)	ESW (%)	AVERAGE (%)
20	0.06	0.03	0.05	0.12	0.07
40	0.09	0.14	0.12	0.15	0.13
60	0.25	0.19	0.18	0.23	0.21
80	−0.02	0.21	0.26	0.21	0.17
100	0.18	0.43	0.45	0.38	0.36

files is 100 %. It is clear in the figures that the PSNR of MEL method is larger than that of the LPC-QIM method.

In order to evaluate the noise introduced by the message embedding process for different embedding rates, we calculate the average PSNR for each situation, as shown in Figs. 7 and 8. We can see that the average PSNR of MEL method is clearly higher than that of the LPC-QIM method for every embedding rate. Thus, we can safely draw the conclusion that the MEL method has less time domain distortion than the LPC-QIM method according to the experimental results.

We used the Mean Opinion Score - Listening Quality Objective (MOS-LQO) value to assess the quality of the stego speech samples followed ITU-T P.563 criterion. The automated MOS-LQO test software [20] provided by ITU-T is used to conduct the test. Tables 8 and 9 show the MOS-LQO test results for the cover speech files and the stego speech files with data embedding by MEL method, processed by G.729 and G.723.1 respectively. The group with 0 % embedding rate denotes the results of the cover files. Table 8 shows that the absolute value of change rate of MOS-LQO value increases with the increasing of embedding rate. It can be seen that the distortion introduced by MEL method is 4.44 % at most.

Table 9 shows that the absolute value of change rate of MOS-LQO value generally increases with the increasing of embedding rate, but the trend is not so obvious as that in
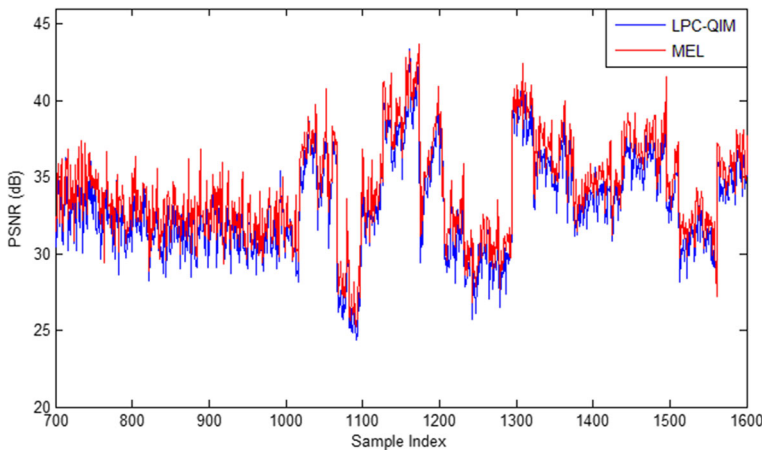

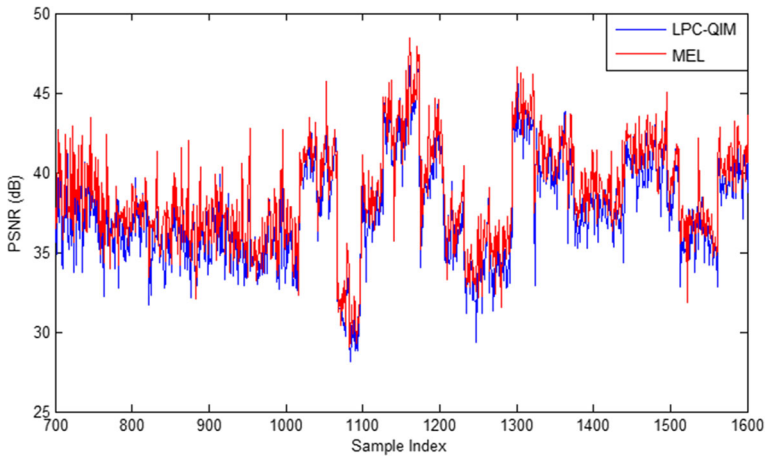**Fig. 5** PSNR test results of samples processed in G.729

Fig. 6 PSNR test results of samples processed in G.723.1

Table 8. This is because, under the same embedding rate, fewer frames translate into fewer alterations which will lead to less distortion. The results shown in Tables 8 and 9 clearly indicate that the distortion introduced by MEL method is small.

Tables 10 and 11 illustrate the comparison between MEL and LPC-QIM method based on MOS-LQO values, processed by G.729 and G.723.1 respectively. In order to compare the two methods, the MOS-LQO test results of LPC-QIM method are used as benchmark. Table 10 shows that the MOS-LQO values increase in every group, which means the distortion introduced by MEL method is less than that of LPC-QIM. Table 11 shows that the
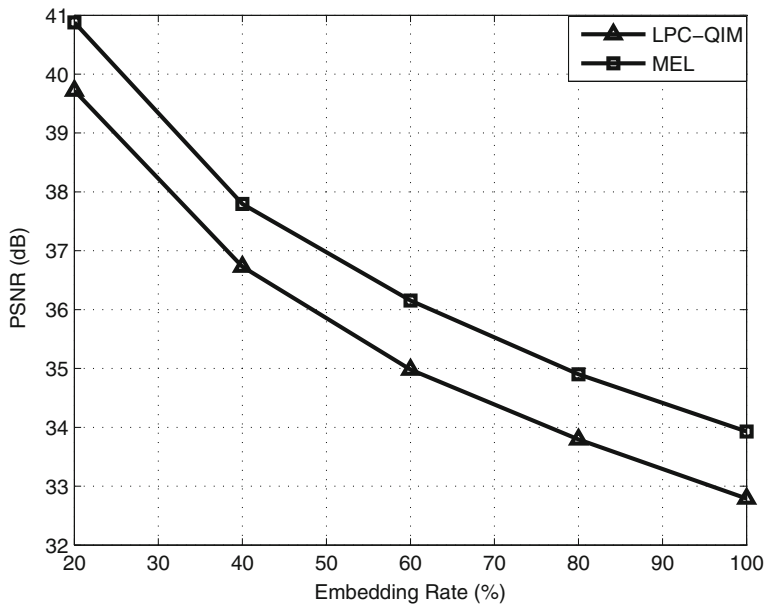

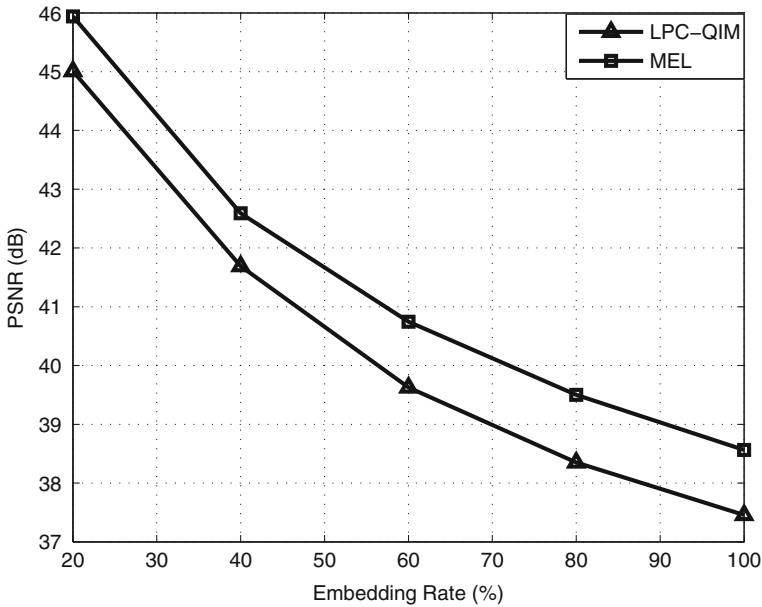**Fig. 7** Average PSNR test results of samples processed in G.729

Fig. 8 Average PSNR test results of samples processed in G.723.1

MOS-LQO values increase in 18 out of 20 groups. This is because the ABS-LPC model is used in low bit-rate codecs. It will minimize the distortion introduced by embedding. Due to its higher embedding efficiency, the MEL method is better than the LPC-QIM method especially when the number of frames is higher. From the analysis above, it is clear that MEL method has less impact on the quality of cover speech than the LPC-QIM method.

3.3 Security analysis

To evaluate the security of the proposed steganography algorithm, we employ the effective steganalysis method proposed in [18]. It combines the Derivative Mel-Frequency Cepstral Coefficients-based feature of speech sample together with the Support Vector Machine to detect audio steganography. Figure 9 shows the detection accuracy on the speech samples

Table 8 MOS-LQO test results of samples processed by G.729

ER	CSM		CSW		ESM		ESW	
(%)	Value (%)	CR (%)	Value (%)	CR (%)	Value (%)	CR (%)	Value (%)	CR (%)
0	3.28891	0	2.95443	0	2.76324	0	2.86536	0
20	3.26178	−0.82	2.93882	−0.53	2.75566	−0.27	2.84106	−0.85
40	3.24752	−1.26	2.90321	−1.73	2.73654	−0.97	2.82974	−1.24
60	3.20346	−2.60	2.87576	−2.66	2.71979	−1.57	2.80133	−2.23
80	3.18253	−3.23	2.85234	−3.46	2.69786	−2.37	2.78436	−2.83
100	3.16028	−3.91	2.82322	−4.44	2.67731	−3.11	2.75489	−3.86

Table 9 MOS-LQO test results of samples processed by G.723.1

ER	CSM		CSW		ESM		ESW	
(%)	Value (%)	CR (%)	Value (%)	CR (%)	Value (%)	CR (%)	Value (%)	CR (%)
0	3.35346	0	2.96471	0	2.78193	0	2.71836	0
20	3.35178	−0.05	2.94682	−0.60	2.78389	0.07	2.71643	−0.07
40	3.34113	−0.37	2.95514	−0.32	2.78271	0.03	2.71462	−0.14
60	3.33755	−0.47	2.94775	−0.57	2.77918	−0.10	2.71776	−0.02
80	3.32852	−0.74	2.93813	−0.90	2.78082	−0.04	2.70825	−0.37
100	3.32771	−0.77	2.95102	−0.46	2.78533	0.12	2.70675	−0.43

processed by G.729. We can see that the detection accuracy of MEL is significantly lower than that of LPC-QIM method.

Figure 10 shows the detection accuracy on the speech samples processed by G.723.1. We can see that the results are mostly concentrated in the vicinity of 50 %. This is because the number of alterations in the frames coded by G.723.1 is too small for effective detection. With the increasing of embedding rate, the detection rate of LPC-QIM method increases significantly higher than that of MEL method. Lower detection rate proves better security of MEL method.

3.4 Embedding efficiency analysis

Experimental results listed above prove that our method has better security and less influence on speech quality compared with the LPC-QIM method. Here we will discuss the underlying causes. Some evaluation parameters for scalar replacement are given in [31], which motivates us to define the corresponding parameters for vector replacement.

Definition 1 (*Hiding Rate*) Hiding rate (HR) denotes the average number of secret bits that embedded in each cover bit. When k secret bits are embedded into n cover bits, the hiding rate can be calculated by

$$\text{HR} = \frac{k}{n}. \tag{13}$$

It reflects the usage rate of the cover. The method with higher data embedding rate can carry more secret bits. For G.729 codec, an embedding throughput of 300 bits/s can be

Table 10 The change in MOS-LQO value on samples processed by G.729

	%Change in MOS-LQO				
ER (%)	CSM (%)	CSW (%)	ESM (%)	ESW (%)	AVERAGE (%)
20	0.27	0.52	0.41	0.18	0.35
40	0.72	0.32	0.32	0.42	0.45
60	0.34	0.77	0.27	0.66	0.51
80	0.26	1.30	0.28	0.83	0.67
100	0.83	1.31	0.47	1.41	1.01

Table 11 The change in MOS-LQO value on samples processed by G.723.1

ER (%)	%Change in MOS-LQO				
	CSM (%)	CSW (%)	ESM (%)	ESW (%)	AVERAGE (%)
20	0.32	−0.12	0.05	0.21	0.12
40	−0.01	0.10	0.18	0.19	0.12
60	0.34	0.13	0.21	0.28	0.24
80	0.06	0.18	0.22	0.16	0.13
100	0.07	0.77	0.59	0.43	0.47

achieved with a hiding rate of 100 %. However, for G.723.1 codec, only 99 bits/s can be achieved with a hiding rate of 100 %.

Definition 2 (*Codeword-Change Rate*) Codeword-change rate (CCR) denotes the average probability of changing for each bit when embedding secret message into cover codewords. When k' out of n cover codewords are changed for embedding, the codeword-change rate can be calculated by

$$CCR = \frac{k'}{n}. \tag{14}$$

It reflects the impact on the host by the information embedding process. The smaller CCR is, the safer steganography becomes. For example, when CCR is 0, the embedding process makes no impact on host. In this case, the steganography is absolutely safe.
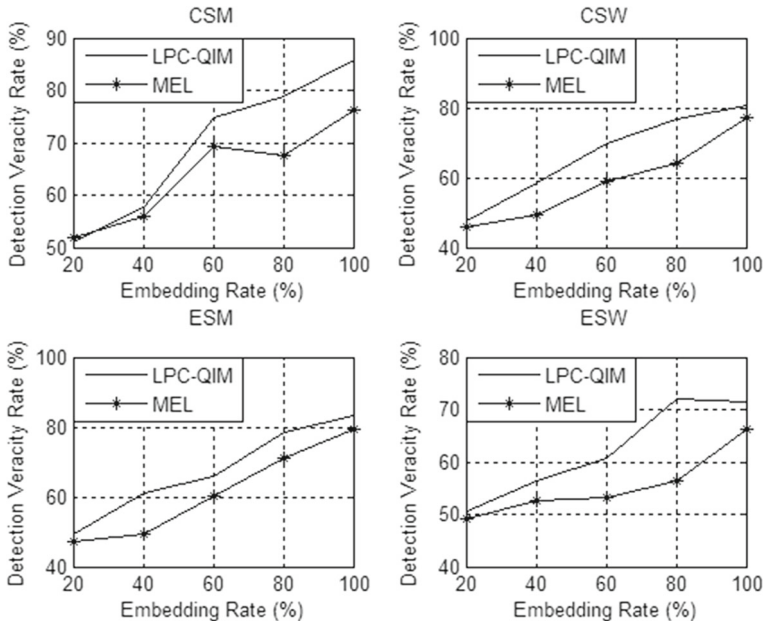

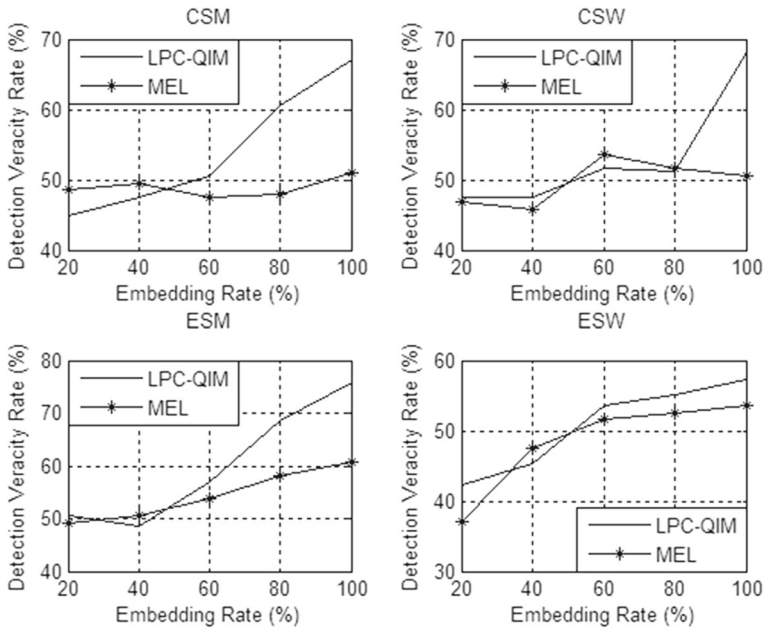**Fig. 9** Detection accuracy of samples processed in G.729

Fig. 10 Detection accuracy of samples processed in G.723.1

Definition 3 (*Embedding Efficiency*) Embedding efficiency (EE) denotes the expected number of secret bits embedded with one embedding change. When k secret bits are embedded by changing k' cover codewords, the embedding efficiency can be calculated by

$$\text{EE} = \frac{k}{k'}. \tag{15}$$

Table 12 shows the comparison of the two methods with the maximum hiding rate. Table 13 shows the comparison of the two methods with the same hiding rate. From Table 12 and Table 13, we can see that:

1. The codeword-change rate of the MEL method is less than that of LPC-QIM method, which means the MEL method has smaller influence on host.

Table 12 The comparison of the two methods with the maximum data embedding rate

Parameter	CCR		HR		EE	
k	MEL (%)	LPC-QIM (%)	MEL (%)	LPC-QIM (%)	MEL	LPC-QIM
3	25	50	66.67	100	2	1
7	12.5	50	42.86	100	3	1
15	6.25	50	26.67	100	4	1
31	3.13	50	16.13	100	5	1
63	1.56	50	9.52	100	6	1
127	0.78	50	5.51	100	7	1

Table 13 The comparison of the two methods with the same data embedding rate

Parameter		CCR		HR		EE	
k	n	MEL (%)	LPC-QIM (%)	MEL (%)	LPC-QIM (%)	MEL	LPC-QIM
2	3	25	33.33	66.67	66.67	2	1
3	7	12.5	21.43	42.86	42.86	3	1
4	15	6.25	13.33	26.67	26.67	4	1
5	31	3.13	8.06	16.13	16.13	5	1
6	63	1.56	4.76	9.52	9.52	6	1
7	127	0.78	2.76	5.51	5.51	7	1

2. MEL method is more efficient than LPC-QIM method under the same data embedding rate. Since MEL method can embed bits by vector replacement at a time, the embedding efficiency of our method is times as that of LPC-QIM method.

4 Conclusion

The main contribution of this paper is to propose an information hiding method for low bit-rate VoIP speech codecs based on matrix embedding. It can be used in all the LPC-based speech codecs. We embed secret information during the VQ process of linear predictive coefficients by codeword replacements. And the replacements are conducted based on the criterion of minimum distance of LP-Coefficient-Vectors regained from the codewords before and after embedding. The embedding efficiency of our method is k ($k > 1$) times as that of LPC-QIM method under the same data embedding rate.

Two criteria were adopted to evaluate the performance of the proposed algorithm: influence of steganography on speech quality analysis and security analysis. In the influence of steganography on speech quality analysis, we compared the two steganographic methods based on subjective speech assessment and objective speech assessment. The experimental results show that our steganographic method introduces less distortion. In the security analysis, we employed the effective steganalysis method proposed in [18] to detect both our steganographic method and the LPC-QIM steganographic method. The experimental results prove that our method has lower probability of being detected. Algorithm proposed in this paper can be applied in independent covert communication system based on VoIP, or modify current VoIP applications via hooking techniques, without modifying corresponding clients, to construct integrated covert communication system, such as compressed speech payload in G-talk.

The future work is to further improve the embedding efficiency so as to minimize the distortion introduced by embedding. The steganalysis methods have also been developing rapidly in recent years, so we need to further enhance the security of the proposed method. The application of our steganographic method is promising. One can directly use our method to hide information in the real-time speech streams in VoIP.

References

1. Chang YT, Huang CT, Huang CL, Wang SJ (2014) Data hiding of high compression ratio in VQ indices with neighboring correlations. Multimed Tools Appl 74(5):1645–1666
2. Chen B, Wornell GW (2001) Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE T Inform Theory 47(4):1423–1443
3. Chiang Y-K, Tsai P, Huang F-L (2008) Codebook partition based steganography without member restriction. Fundam Inform 82(1-2):15–27
4. Crandall R (1998) Some notes on steganography. Available: http://os.inf.tu-dresden.de/~westfeld/crandall.pdf. Accessed 1 May 2015
5. Girl (2008) Available: http://en.wikipedia.org/wiki/Girl. Accessed 1 May 2015
6. Huang YF, Liu CH, Tang SY, Bai S (2012) Steganography integration into a low-bit rate speech codec. IEEE T Inf Foren Sec 7(6):1865–1875
7. Huang YF, Tang S, Yuan J (2011) Steganography in inactive frames of VoIP streams encoded by source codec. IEEE T Inf Foren Sec 6(2):296–306
8. Hui T, Ke Z, Hong J, Jin L, Yongfeng H, Dan F (2009) An M-Sequence Based Steganography Model for Voice over IP. In: IEEE International Conference on Communications, 2009. ICC '09. IEEE, pp 1–5
9. Ito A, Abe S, Suzuki Y (2009) Information hiding for G.711 speech based on substitution of least significant bits and estimation of tolerable distortion. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. IEEE, pp 1409–1412
10. ITU-T G.729 (1996) Coding of speech at 8 kbits/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). International Telecommunications Union, Geneva, Switzerland
11. Jin L, Ke Z, Hui T (2012) Least-significant-digit steganography in low bitrate speech. In: IEEE International Conference on Communications (ICC), 2012. IEEE, pp 1133–1137
12. Lee JD, Chiou YH, Guo JM (2013) Lossless data hiding for VQ indices based on neighboring correlation. Inform Sci 221(2):419–438
13. Li SB, Tao HZ, Huang YF (2012) Detection of quantization index modulation steganography in G.723.1 bit stream based on quantization index sequence analysis. J Zhejiang Univ-SC C 13(8):624–634
14. Lin C, Pan JS, Liao BY (2006) Robust VQ-based digital image watermarking for mobile wireless channel. In: Proceedings of the 2006 IEEE International Conference on Systems, Man and Cybernetics. IEEE, pp 2380–2384
15. Liu L, Li M, Li Q, Liang Y (2008) Perceptually transparent information hiding in G. 729 bitstream. In: 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, pp 406–409
16. Lu ZM, Yan B, Sun SH (2005) Watermarking combined with CELP speech coding for authentication. Ieice T Inf Syst E88-D (2):330–334
17. Qin C, Chang CC, Chen YC (2013) Efficient reversible data hiding for VQ-compressed images based on index mapping mechanism. Signal Process 93(9):2687–2695
18. Qingzhong L, Sung AH, Mengyu Q (2009) Temporal derivative-based spectrum and Mel-Cepstrum audio Steganalysis. IEEE T Inf Foren Sec 4(3):359–368
19. Rahmani P, Dastghaibyfard G (2014) A low distortion reversible data hiding scheme for search order coding of VQ indices. Multimed Tools Appl:1–22
20. Recommendation P.563 Erratum 1 (2007) http://www.itu.int/rec/T-REC-P.563-200710-I!Err1/en. Accessed 15 August 2015
21. Recommendation P.800 (1996) http://www.itu.int/rec/T-REC-P.800-199608-I/en. Accessed 15 August 2015
22. Recommendation P.862 (2001) http://www.itu.int/rec/T-REC-P.862-200102-I/en. Accessed 15 August 2015
23. Roselinkiruba R, Balakirshnan R (2013) Secure steganography in audio using inactive frames of VoIP streams. In: IEEE Conference on Information & Communication Technologies (ICT), 2013. IEEE, pp 491–495
24. Simmons G (1984) The prisoners problem and the subliminal channel. In: Chaum D (ed) Advances in Cryptology. Springer US, pp 51–67
25. Simmons G (1985) The subliminal channel and digital signatures. In: Beth T, Cot N, Ingemarsson I (eds) Advances in Cryptology, vol 209. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp 364–378
26. Simmons GJ (1994) Subliminal channels - past and present. Eur T Telecommun 5(4):459–473
27. TeleGeography Report & Database (2014) Available: https://www.telegeography.com/research-services/telegeography-report-database/index.html Accessed 15 August 2015

28. Tingting X, Zhen Y (2009) Simple and effective speech steganography in G.723.1 low-rate codes. In: WCSP '09 International Conference on Wireless Communications & Signal Processing, 2009. IEEE, pp 1–4
29. Tutte WT (2001) Graph theory. Cambridge University Press, pp 233–237
30. Wang FH, Jain LC, Pan JS (2007) VQ-based watermarking scheme with genetic codebook partition. J Netw Comput Appl 30(1):4–23
31. Westfeld A (2001) F5A Steganographic Algorithm. In: Moskowitz I (ed) Information Hiding, vol 2137. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp 289–302
32. Xiao B, Huang Y, Tang S (2008) An approach to information hiding in low bit-rate speech stream. In: Global Telecommunications Conference. IEEE, pp 1–5
33. Yan S, Tang G, Sun Y, Gao Z, Shen L (2014) A triple-layer steganography scheme for low bit-rate speech streams. Multimed Tools Appl:1–20

Peng Liu is a Ph.D. candidate in the Haikou Laboratory, Institute of Acoustics, Chinese Academy of Science, Haikou, China. He received his BS degree in from Hainan University in 2011, majoring in communication engineering. His research interests include multimedia signal processing and data hiding.

Songbin Li received the Ph.D. degree from the Institute of Acoustics, Chinese Academy of Science, in 2010. And then, he did his postdoctoral research at the Department of Electronic Engineering, Tsinghua University. From July of 2012, he is an Associate Professor in the Haikou Laboratory, Institute of Acoustics, Chinese Academy of Science, Haikou, China. His research interests focus on multimedia signal processing and information hiding.

Haiqiang Wang is currently pursuing Ph.D. degree at the University of Southern California (USC). He received his B.S. and M.S. degrees in Signal and Information Processing from Northwestern Polytechnical University (NWPU), Xi'an, China. His research interests include multimedia signal processing and computer vision.