Hindawi
Security and Communication Networks
Volume 2018, Article ID 7080673, 11 pages
https://doi.org/10.1155/2018/7080673

Research Article

A Novel AMR-WB Speech Steganography Based on Diameter-Neighbor Codebook Partition

Junhui He,[1] **Junxi Chen**,[1] **Shichang Xiao**,[1] **Xiaoyu Huang**,[2] **and Shaohua Tang**[1]

[1]*School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China*
[2]*School of Economics and Commerce, South China University of Technology, Guangzhou 510006, China*

Correspondence should be addressed to Junhui He; hejh@scut.edu.cn

Received 28 September 2017; Accepted 26 December 2017; Published 13 February 2018

Academic Editor: Rémi Cogranne

Steganography is a means of covert communication without revealing the occurrence and the real purpose of communication. The adaptive multirate wideband (AMR-WB) is a widely adapted format in mobile handsets and is also the recommended speech codec for VoLTE. In this paper, a novel AMR-WB speech steganography is proposed based on diameter-neighbor codebook partition algorithm. Different embedding capacity may be achieved by adjusting the iterative parameters during codebook division. The experimental results prove that the presented AMR-WB steganography may provide higher and flexible embedding capacity without inducing perceptible distortion compared with the state-of-the-art methods. With 48 iterations of cluster merging, twice the embedding capacity of complementary-neighbor-vertices-based embedding method may be obtained with a decrease of only around 2% in speech quality and much the same undetectability. Moreover, both the quality of stego speech and the security regarding statistical steganalysis are better than the recent speech steganography based on neighbor-index-division codebook partition.

1. Introduction

With the rapid development of the Internet and the growing popularity of instant messaging application, people are increasingly using audio-based communication. How to avoid interception and secure communication turns into one of the most important research problems. Encryption is a conventional method of protecting communication; however, the transmission of ciphered content may easily arouse attackers' suspicion. In recent years, steganography has been presented as an effective means of covert communication. Audio steganography can transfer important messages secretly by embedding them into cover audio files with the use of information hiding techniques [1]. Data hiding in audio is especially challenging because the human auditory system operates over a wider dynamic range in comparison with human visual system.

Many works on audio steganography have been already reported. Gruhl et al. [2] proposed an audio steganographic method of echo hiding by the introduction of synthetic resonances in the form of closely spaced echoes. Gopalan [3] presented a method of embedding a covert audio message into a cover utterance by altering one bit in each of the cover utterance samples. Gopalan et al. [4] provided two methods of secret message embedding by modifying the phase or amplitude of perceptually masked or significant regions of a host. And a direct-sequence spread-spectrum watermarking method with strong robustness against common audio editing procedures was proposed in [5]. And many audio steganographic applications including Steghide and Hide4PGP can be freely downloaded from the Internet. But most of these methods are not resilient to AMR-WB speech.

Based on segmental SNR analysis of modification to the encoded bits in a frame, Liu et al. [6] selected the perceptually least important bits to embed secret message in G.729 speech. In [7], a simple and effective steganographic approach, which may be applied to 5.3 Kbps G.723.1 speech, was presented based on analyzing the redundancy of code parameters, and augmented identity matrix was utilized to lower the distortion of cover speech. Similarly, by calculating speech quality sensitivity on each encoded bit out of 244 bits using perceptual evaluation of speech quality (PESQ) criterion, a

data hiding approach to embedding data in enhanced full rate (EFR) compressed speech bitstream is proposed in [8]. In addition, Nishimura [9] proposed three methods of hiding data in the pitch delay data of the AMR speech.

Based on complementary neighbor vertices codebook partition algorithm (CNV), Xiao et al. [10] presented an approach to information hiding in compressed speech with the use of quantization index modulation (QIM) [11]. Huang et al. [12] proposed a steganographic algorithm for embedding data in different speech encoding parameters of the inactive frames, the embedding capacity of which is bounded by the number of inactive frames in the cover speech. In [13], Huang et al. also presented a method for steganography in low bit-rate VoIP streams based on pitch period prediction. It can achieve high quality of stegospeech and prevent statistical steganalysis, but the embedding rate is still low (only about 133.3 bps). And an adaptive suboptimal pulse combination constrained (ASOPCC) method was presented in [14] to embed data into compressed speech signal of AMR-WB codec. However, most of the PESQ scores in different coding modes are not high. In [15], a key-based codebook partition strategy, which dynamically determines the adopted division scheme, was designed to improve the security of the QIM steganography in speech bitstream. Although the stegospeech quality is guaranteed to be good, the embedding capacity is very limited and not adjustable. Liu et al. [16] proposed a neighbor-index-division codebook division algorithm (NID) for G.723.1 speech. Differing from the existing CNV method, NID divides neighbor-indexed codewords into separated subcodebooks according to a suitable stegocoding strategy. The embedding capacity is improved by using multiple division and multi-ary coding strategy.

The adaptive multirate wideband (AMR-WB) is a widely adapted format in mobile handsets and is also the recommended speech codec for VoLTE. AMR-WB speech may be a good candidate for cover medium in audio steganography. Therefore, we will focus on AMR-WB speech steganography in this paper. Firstly, a new diameter-neighbor (DN) codebook partition algorithm toward AMR-WB speech is proposed. Based on DN codebook division, we develop a novel AMR-WB speech steganography capable of providing flexible embedding capacity with different iterative parameter N_i. For example, when $N_i = 48$, twice the embedding capacity of CNV-based method may be obtained with a decrease of only about 2% in speech quality and much the same undetectability. Moreover, both the quality of stego speech and the security of defending against statistical steganalysis [17, 18] are better than the recent NID-based speech steganography.

The remainder of this paper is organized as follows. In Section 2, the related work is briefly introduced. In Section 3, the proposed DN codebook partition algorithm and the novel AMR-WB speech steganography are described in detail. The experimental results and analysis are provided in Section 4. Finally, conclusions are presented.

2. Related Work

In this section, a technical overview of AMR-WB codec is firstly presented. Then two related codebook partition algorithms CNV [10] and NID [16] are also briefly reviewed.

2.1. AMR-WB Codec. The AMR-WB speech codec is standardized by 3GPP (3rd Generation Partnership Project) and adopted as the standard G.722.2 by ITU-T in 2002 [19]. It is a multirate wideband speech codec applied in modern mobile communication systems to remarkably improve the speech quality. The AMR-WB codec operates at a multitude of bit rates ranging from 6.6 kbit/s to 23.85 kbit/s.

The input audio signal is separated into 20 ms long frame using 16 kHz sampling rate. Every frame contains a linear prediction analysis (LPA) and the LP coefficients are converted to immittance spectrum pairs (ISP) coefficients. ISP coefficients are then converted to frequency domain (ISF) for quantization. Except for mode 0 (6.6 kbit/s), the ISF coefficients are quantized using two-stage vector quantization with split-by-2 in first stage and split-by-5 in the second stage. Both the second and the third codebooks in the second stage have 128 codewords, and the ISF indices of the codewords in these codebooks may be employed to embed secret message.

In the decoder, the transmitted indices are first parsed from the received bitstream and then decoded to obtain the code parameters for each transmitted frame, such as the ISP vector, the 4 fractional pitch lags, the 4 LTP filtering parameters, the 4 innovative code vectors, and the 4 sets of vector quantized pitch and innovative gains. For a more detailed description, one should refer to [19]. From the received ISF indices, which may have been modified because of secret message embedding, the receiver can recover the embedded secret message.

2.2. Complementary Neighbor Vertices. CNV is a new type of codebook partition algorithm proposed in [10], in which each codeword in a codebook is viewed as a vertex in the multidimensional space. The relationship between two codewords X and Y is described as an edge connecting the two codewords' vertices. And the weight of an edge is defined as the Euclidean distance $D(XY)$ between two codewords X and Y. Small value of $D(XY)$ indicates that X and Y bear a close resemblance to each other. The vertex nearest to X is referred to as X's neighbor vertex, which is denoted by $N(X)$. The vertex set V together with the edge set E form a graph $G(V, E)$ in a multidimensional space.

The codebook partition is realized by the construction of the graph $G(V, E)$ and vertex labelling. First, each vertex X in $G(V, E)$ is connected with its neighbor vertex $N(X)$ using an edge. Thus, the graph $G(V, E)$ would be divided into several isolated subgraphs, each of which may be proved to be acyclic and 2-colorable. Second, every vertex and its neighbor vertex in a subgraph are labelled oppositely using "0" or "1." Third, all of the vertices with same label are collected into a subcodebook; hence, two subcodebooks will be obtained.

Based on the generated subgraphs and the label assigned to each codeword in them, CNV-based steganography applies QIM concept to embed secret message. More specifically, when the label of the codeword X, which is associated with the cover quantization index I_a, agrees with the secret message, I_a remains unchanged, or else it should be replaced

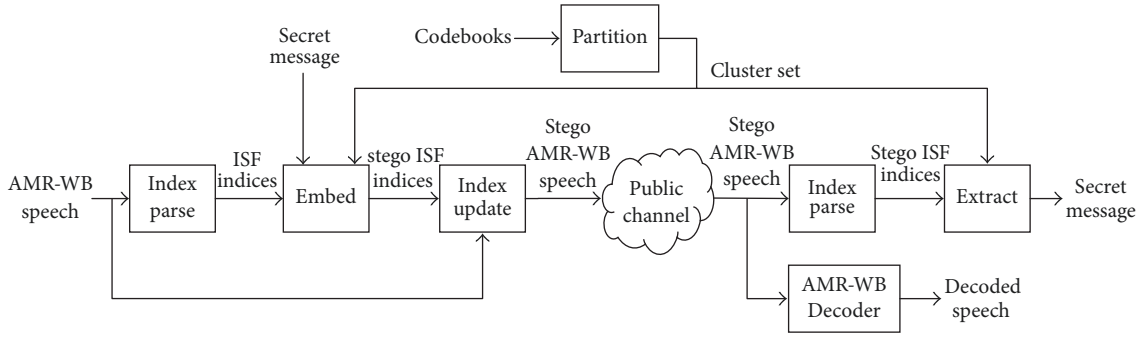


Figure 1: Diagram of the proposed method.

with the quantization index of the neighbor codeword $N(X)$ which belongs to the opposite subcodebook.

The key characteristic of CNV-based steganography is that the distortion is bound even in the worst case. However, the embedding capacity is limited, which is analyzed experimentally in Section 4. Moreover, the number of possible combinations of flipping coefficients which determine whether the labels in a subgraph will be flipped is large. Extra information about the flipping process must be transmitted to the receiver and thus the effective embedding capacity may be decreased further.

2.3. Neighbor Index Division. NID assumes that the codewords of neighbor indices (i.e., neighbor positions) in a codebook would be close together. Hence the codewords in a codebook can be easily separated into subcodebooks according to their indices instead of the Euclidean distance. Specifically, select an appropriate integer k according to the demand for embedding capacity and label the ith codeword with digit $(i-1) \bmod k$, respectively. Then collect all the codewords with same label into a subcodebook and obtain k different subcodebooks.

In order to take full use of the embedding capacity, the binary secret message should be transformed into k-ary digits denoted by m ($m \in \{0, 1, \ldots, k-1\}$). When the codeword related to the cover quantization index belongs to the subcodebook whose label differs from the k-ary digit m to be embedded, this index should be substituted with that of the closest codeword in the corresponding subcodebook$_m$.

NID-based steganography is an information hiding method based on neighbor-index codebook partition, of which the embedding capacity may be controlled by the number of subcodebooks k. However, as illustrated in [16], only about 34% of the pairs of neighbor-index codewords happened to be the pairs of neighbor-vertex codewords. And the mean distance between neighbor-index codewords is apparently larger than that of neighbor-vertex codewords. Therefore, the amount of distortion induced by NID-based steganography may be a little large, which is proved by the experimental results provided in Section 4.

3. Proposed Method

The diagram of the proposed method is shown in Figure 1. Based on DN codebook partition of the codebooks described in Section 2.1, secret message can be embedded into an AMR-WB speech file. After the stego AMR-WB speech file is received, the embedded secret message can be extracted without errors. At the same time, the decoded speech without perceptible distortion will also be obtained. In the following section, the diameter-neighbor codebook partition algorithm (DN) is first introduced. Then the embedding and extraction procedure of our proposed method are described.

3.1. Codebook Partition. A codebook may be viewed as a list of isolated code vectors (i.e., codewords) in the multidimensional space. The codebook partition algorithm used for audio steganography is to divide the codebook into several clusters, in each of which the codewords can be replaced with each other without causing perceptible distortion.

Let B denote the original codebook with N_b codewords, and C denote a cluster with N_c codewords W_t ($t = 1, 2, \ldots, N_c$), and the centroid G of a cluster C is defined as follows:

$$G(i) = \frac{1}{N_c} \sum_{t=1}^{N_c} W_t(i), \tag{1}$$

where $G(i)$ and $W_t(i)$ are the ith components of G and W_t, respectively.

The centroid G (average code vector) is used to represent the corresponding cluster C; hence, the cluster C may also be considered as a vector in the multidimensional codebook space. In order to describe the similarity between two clusters C_1 and C_2, the Euclidean distance between them is defined as follows:

$$D(C_1, C_2) = \sqrt{\sum_{i=1}^{n} \left(G_1(i) - G_2(i)\right)^2}, \tag{2}$$

where G_1 and G_2 are the corresponding geometric center points of the two clusters C_1 and C_2. And n is the dimension of a codeword; $G_1(i)$ and $G_2(i)$ are the ith components of G_1 and G_2, respectively.

Let S denote a cluster set. The diameter of S is defined as the maximal Euclidean distance D_m of all cluster pairs in the cluster set S, that is,

$$D\left(C_p, C_q\right) \le D_m \quad \forall p, q = 1, 2, \ldots, |S|, \tag{3}$$

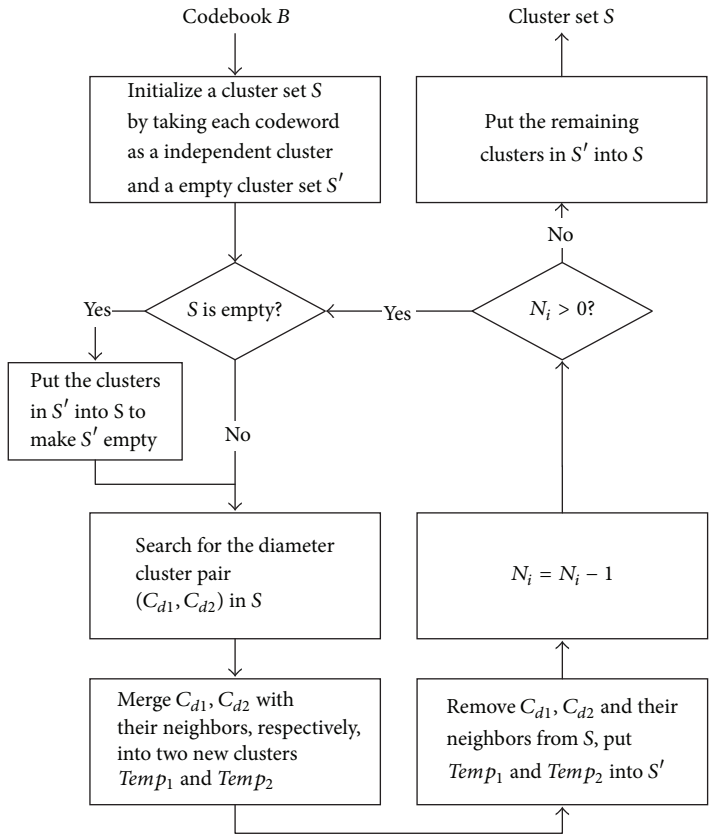


Figure 2: Diagram of our proposed codebook partition.

where $|S|$ is the number of clusters within the cluster set S. The cluster pair with maximal Euclidean distance D_m, called diameter cluster pair, is denoted by (C_{d1}, C_{d2}). And the neighbor of a cluster C in S is represented by $N(C, S)$; then we have

$$D(C, N(C, S)) \leq D\left(C, C_p\right) \quad \forall p = 1, 2, \ldots, |S|. \tag{4}$$

Figure 2 illustrates the diagram of the proposed DN codebook partition algorithm. And its detailed procedure is given in Algorithm 1. The original codebook B will be divided into $|S|$ clusters by iteratively merging the diameter cluster pair with their respective neighbors. An iteration parameter N_i is applied to obtain flexible embedding capacity through controlling the merging procedure. The relationship between N_i and the embedding capacity will be discussed in Section 4.3.

Figure 3 is provided as an example to illustrate the proposed codebook partition algorithm. The white circle "○" denotes a codeword. And the oval "⬭" with shadow denotes a codeword and its neighbor in S being processed, while the oval "⬭" without shadow represents a cluster in S' that has been formed. The "0," "1," "00," "01," "10," or "11" in a circle "○" is the label of a codeword in the cluster. The cross "×" means the centroid of the cluster it belongs to, and a line "–" represents the diameter of a cluster set. The first to third merging iterations are shown in Figures 3(a)–3(c), respectively. The fourth merging iteration is comprised of Figures 3(d) and 3(e), and Figure 3(f) demonstrates the labelling of the codewords.

3.2. Embedding Procedure. In our proposed method, the ISF indices corresponding to the codewords in the codebook are first obtained by parsing the host AMR-WB speech. Then the ISF indices are employed to embed secret message based on codebook partition. Generally, the codewords in the same cluster as the codeword referred by I_a lies in are considered to be replaceable with each other. According to the secret message to be embedded, I_a may be substituted by one of the other codewords' indices within the same cluster. The number of secret message bits that can be embedded depends on the size of the specific cluster. The embedding procedures are given in the following.

Step 1. Search cluster set S for the cluster C which contains the codeword referred by the ISF index I_a.

Step 2. If there are N codewords in C, the number of secret bits that can be embedded into I_a is calculated as $n = \lfloor \log_2 N \rfloor$.

Step 3. Read n not-yet-embedded bits, denoted by m, from the secret message. I_a is replaced with I_b which indexes the codeword with the same label as m.

Step 4. Repeat Steps 1–3 until all the secret bits are embedded.

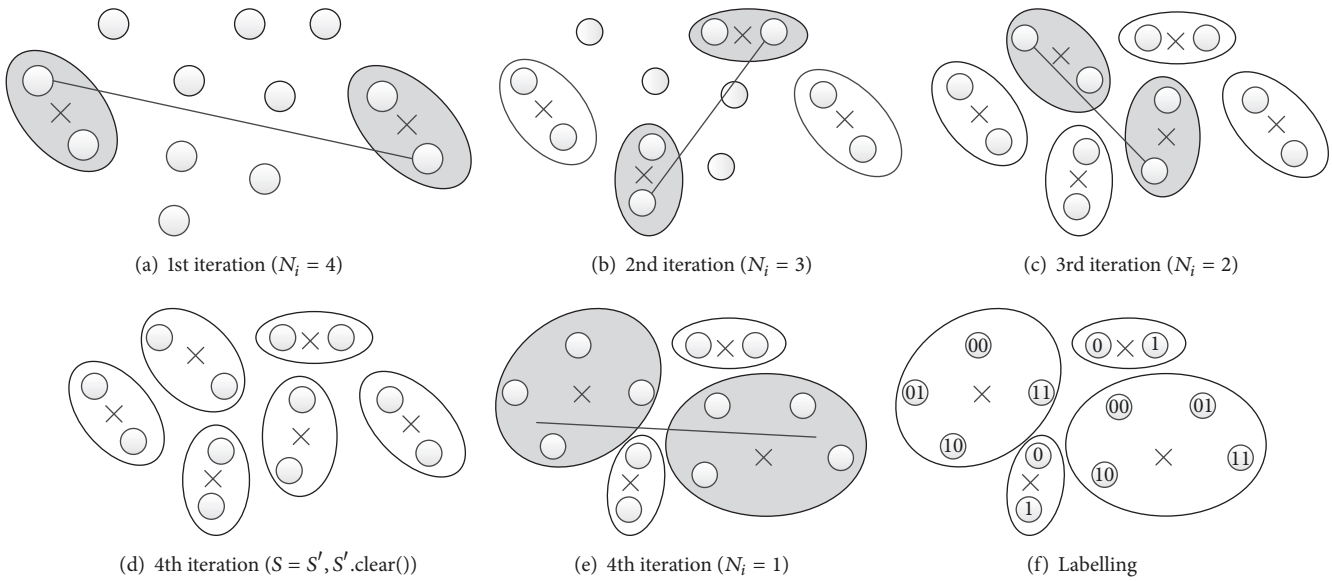


Figure 3: An example of our proposed codebook partition.

Input: Codebook B, iterative parameter N_i
Output: Cluster set S
/* S' is a helper cluster set */
S'.clear();
S.clear();
/* Each codeword is taken as a initial cluster */
for $i = 0$; $i < N_c$; $++i$ **do**
  S.push (C_i);
end
/* Iterative merging */
while $N_i > 0$ **do**
  if S *is empty* **then**
    $S = S'$;
    S'.clear();
  end
  $(C_{d1}, C_{d1}) = \arg\max_{i,j\in\{1,2,\ldots,|S|\}} D(C_i, C_j)$;
  $Temp_1 = C_{d1} \cup N(C_{d1}, S)$;
  $Temp_2 = C_{d2} \cup N(C_{d2}, S)$;
  S'.push ($Temp_1$);
  S'.push ($Temp_2$);
  S.remove (C_{d1});
  S.remove (C_{d2});
  S.remove ($N(C_{d1}, S)$);
  S.remove ($N(C_{d2}, S)$);
  $N_i = N_i - 1$;
end
/* Put the remaining clusters in S' into S */
for iter = S'.begin(); iter < S'.end(); ++iter
 do
  S.push (*iter);
end
return S;

Algorithm 1: DN-based codebook partition algorithm.

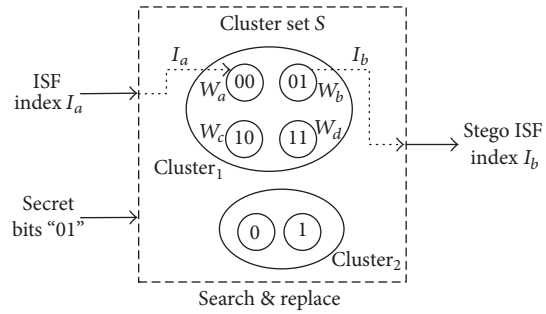


Figure 4: Embedding two bits into one cover ISF index.

Figure 4 is an example of embedding two secret bits into one cover ISF index. Let us assume the cluster set S contains two clusters and the corresponding codeword indexed by I_x is W_x; for example, I_b indexes the codeword W_b. Hence, the ISF index I_a shown in Figure 4 will be replaced with I_b which indexes the codeword W_b with the same label as the secret bits "01."

3.3. Extracting Procedure. When the stego AMR-WB speech is transferred to the intended receiver, the stego indices may be obtained by parsing AMR-WB speech stream and used to extract the embedded secret message. The message extraction procedures from the stegoindex I_b are given below.

Step 1. Search cluster set S, which is the same as that employed in the embedding procedure, for the cluster C which contains the codeword W_b referred by the ISF index I_b.

Step 2. If there are totally N codewords in C, the number of secret bits carried by I_b is computed by $n = \lfloor \log_2 N \rfloor$.

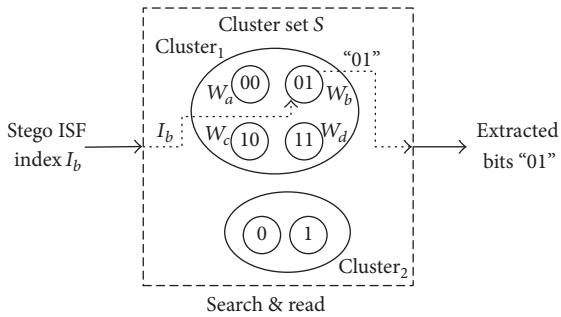


Figure 5: Extracting two bits from one stego-ISF index.

Step 3. Read the label of W_b as the extracted n bits, which are appended to the secret message bit sequence.

Step 4. Repeat Steps 1–3 until all the secret bits are recovered.

Figure 5 is the corresponding example of extracting two secret bits from the stegoindex I_b generated by the previous embedding instance shown in Figure 4. It can be easily seen that the extracted secret bits are identical to the embedded secret bits.

4. Experimental Results and Analysis

In order to demonstrate the performance of the proposed method, the perceptual quality of the stego AMR-WB speech with secret message embedded using our method is computed and compared to that of the stego AMR-WB speech generated with CNV and NID steganography. Moreover, the flexibility of embedding capacity and the security regarding statistical detection are analyzed in detail.

4.1. Audio Database. TIMIT acoustic-phonetic continuous speech corpus (https://catalog.ldc.upenn.edu/ldc93s1) is an audio database which contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences, and all audio sentences are sampled at 16 kHz. In our experiments, 1000 audio sentences are randomly chosen from TIMIT database. The average, maximum, and minimum length of the chosen audio sentences are 3.47 s, 3.96 s, and 3.12 s. All audio files are converted into AMR-WB format using standard codec.

4.2. Speech Quality Evaluation. The perceptual evaluation of speech quality (PESQ) described in the ITU-T P.862 Recommendation [20] may be employed to evaluate speech quality. Moreover, according to ITU-T P.862.2 [21], the raw PESQ score can be converted to mean opinion score-listening quality objective (MOS-LQO), which is more suitable for evaluating wideband speech. Hence, MOS-LQO is applied in our experiments. The normal range of MOS-LQO score is 1.017 to 4.549. The higher the score, the better the quality.

Figure 6 shows the MOS-LQO scores of the 1000 cover AMR-WB speeches in 23.85 kbit/s mode and the corresponding stego AMR-WB speeches using three different codebook partition algorithms. Three progressive embedding rates, that

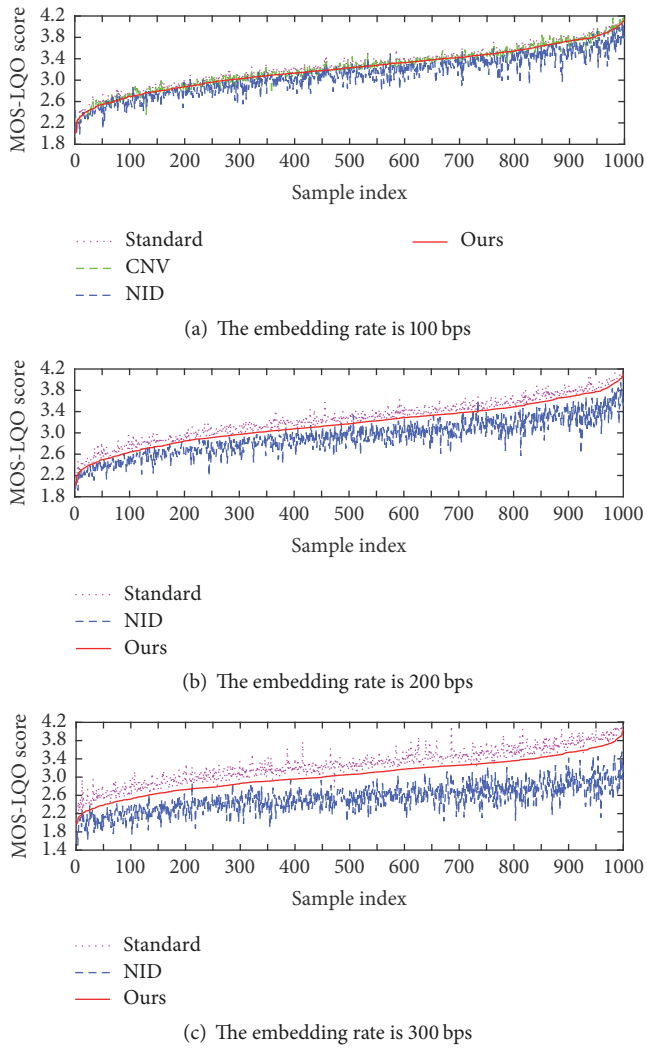


Figure 6: Comparisons of MOS-LQO values for 1000 samples between the standard AMR-WB codec, CNV-based steganography, NID-based steganography, and the proposed DN-based steganography.

is, 100 bps, 200 bps, and 300 bps, are employed in our experiments. The indices of speech samples are sorted according to the MOS-LQO scores of our proposed method. It can be seen from Figure 6 that the overall scores of the stego AMR-WB speeches generated with our method are higher than those of the NID-based stego AMR-WB speeches, especially when the embedding rates are 200 bps and 300 bps. And the MOS-LQO scores of the CNV-based stego AMR-WB speeches are slightly higher than ours when the embedding rate is 100 bps, which means there are no obvious discrepancies in speech quality between them. Besides, when the high embedding rate, that is, 200 bps or 300 bps, is used, the decrease in MOS-LQO scores of our stego AMR-WB speeches is significantly smaller than that of NID-based steganography.

Moreover, the average MOS-LQO scores of the cover AMR-WB speeches and the stego AMR-WB speeches with three different codebook partition algorithms, that is, CNV,

Table 1: MOS-LQO scores of the standard codec, CNV-based, NID-based, and our proposed steganography in four different rate modes and three embedding rates.

Embedding rate	Method	Rate mode (kbit/s)			
		12.65	15.85	19.85	23.85
	Standard	2.929	3.073	3.199	3.269
100 bps	CNV	2.871 (−2.0%)	3.021 (−1.7%)	3.153 (−1.4%)	3.225 (−1.3%)
	NID	2.750 (−6.1%)	2.895 (−5.8%)	3.020 (−5.6%)	3.091 (−5.4%)
	Ours	2.864 (−2.2%)	3.010 (−2.0%)	3.139 (−1.9%)	3.216 (−1.6%)
200 bps	CNV	/	/	/	/
	NID	2.601 (−11.2%)	2.736 (−11.0%)	2.875 (−10.7%)	2.921 (−10.6%)
	Ours	2.807 (−4.2%)	2.955 (−3.8%)	3.084 (−3.6%)	3.164 (−3.2%)
300 bps	CNV	/	/	/	/
	NID	2.284 (−22.0%)	2.386 (−22.3%)	2.475 (−22.6%)	2.533 (−22.5%)
	Ours	2.699 (−7.9%)	2.841 (−7.5%)	2.971 (−7.1%)	3.046 (−6.8%)

NID, and DN, including four rate modes (12.65 kbit/s, 15.85 kbit/s, 19.85 kbit/s, and 23.85 kbit/s) together with three kinds of embedding rate (100 bps, 200 bps, and 300 bps), are given in Table 1. Only the MOS-LQO scores of NID-based and DN-based steganographic methods with embedding rates 200 bps and 300 bps are given in Table 1 because the embedding capacity of CNV-based steganography may not be larger than 100 bps.

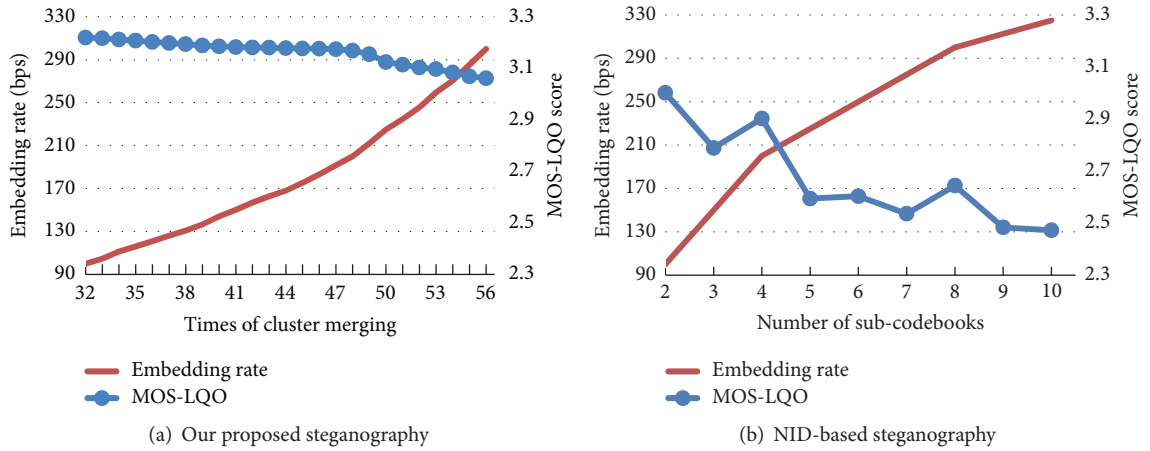
When the embedding rate is 100 bps, which is almost the limit of CNV steganography, we can see from Table 1 that the mean MOS-LQO scores of our proposed method are only about 0.3% worse than CNV-based steganography. The slight decrease may be almost imperceptible by human auditory system (HAS). And there are significant increases of approximately 3.8% in the mean MOS-LQO scores when our presented method is compared to NID-based steganography. And it can be observed that when the embedding rates are 200 bps and 300 bps, the scores of our approach are improved by about 7% and 15% correspondingly in contrast to those of NID-based steganography.

Furthermore, we can also see that the experimental results of four rate modes are analogous. The decrease of speech quality caused by NID-based steganography is more than twice that caused by DN-based steganography. And the proposed method can obtain twice the embedding capacity of CNV-based steganography by sacrificing less than 2% speech quality in average. In addition, only a slight decline in speech quality is observed when 300 bps embedding rate is used in the proposed DN-based method while 200 bps is employed in NID-based method.

4.3. Flexible Embedding Capacity. Compared to CNV-based steganography, flexible embedding capacity may be obtained to satisfy different practical demand with our proposed method. The steganographic capacity can be adjusted by changing the iteration parameter N_i. For different values of N_i, for example, $N_i = 32, 33, \ldots, 54$, the average embedding capacity and the MOS-LQO scores are given in Figure 7(a), and the corresponding results of NID-based steganography are provided in Figure 7(b) for comparison. Without loss of generality, only 23.85 kbit/s mode is used.

From Figure 7, we can observe that the embedding rate significantly increases with N_i while the MOS-LQO score slightly goes down. However, as NID-based steganography is concerned, the MOS-LQO score rapidly declines with the increase of the embedding rate. Therefore, the proposed DN-based steganography can achieve higher embedding capacity with slight decrease in speech quality. For example, when $N_i = 48$, the size of each cluster in S is equal to 4 and we can embed 4 bits per frame; that is, the embedding rate is 200 bps, but, at the same time, the CNV algorithm can embed at most 2 bits per frame (100 bps).

4.4. Resistibility of Statistical Steganalysis. Speech steganography aims to hide secret message into cover speech without arousing suspicion. It is very important for a steganographic method to resist statistical steganalysis, which is the technique of detecting the presence of hidden message. Two state-of-the-art steganalytic methods [17, 18] are used to evaluate the performance of statistical undetectability of our proposed method. In [17], mel-cepstrum coefficients and Markov transition features from the second-order derivative of the audio signal are extracted to capture the statistical distortions caused by audio steganography, while, in [18], the correlation characteristics of split vector quantization codewords of linear predictive coding filter coefficients are



(a) Our proposed steganography (b) NID-based steganography

Figure 7: Relationship between the embedding rates and the MOS-LQO scores for our proposed steganography and NID-based steganography.

Table 2: Steganalysis results of different steganographic methods in 23.85 kbits/s mode.

Training rate	0.4				0.5				0.6			
Method	Markov	MFCC	SS-QCCN	RS-QCCN	Markov	MFCC	SS-QCCN	RS-QCCN	Markov	MFCC	SS-QCCN	RS-QCCN
100 bps												
CNV	49.8%	49.8%	43.7%	49.0%	50.1%	50.2%	44.0%	49.2%	50.0%	50.5%	41.9%	50.0%
NID	51.0%	60.1%	42.2%	50.0%	50.1%	60.9%	42.9%	48.7%	52.1%	59.8%	41.8%	49.4%
Ours	50.0%	50.0%	44.0%	49.4%	50.3%	49.3%	40.3%	49.4%	49.1%	48.6%	41.8%	43.3%
200 bps												
CNV	/	/	/	/	/	/	/	/	/	/	/	/
NID	53.5%	74.5%	46.9%	50.0%	53.3%	76.2%	47.6%	50.0%	53.6%	75.8%	44.4%	50.1%
Ours	51.0%	48.3%	45.2%	50.0%	49.8%	48.7%	42.2%	50.0%	50.5%	48.6%	45.0%	50.0%
300 bps												
CNV	/	/	/	/	/	/	/	/	/	/	/	/
NID	54.8%	74.6%	49.3%	50.0%	56.3%	77.2%	50.0%	50.0%	55.4%	78.3%	50.5%	50.6%
Ours	52.4%	49.7%	47.9%	50.0%	52.8%	60.9%	48.2%	50.0%	53.8%	50.1%	46.6%	50.0%

utilized to steganalyze QIM-based steganography in low-bit-rate speech (such as G.723.1 and G.729). Both steganalytic methods use a support vector machine to predict the existence of hidden message in given audios.

In our experiments, the sentences chosen from "TIMIT" databases as stated in Section 4.1 are first encoded using the standard AMR-WB codec. These AMR-WB recordings constitute the cover speech set. Then secret message is embedded into each cover AMR-WB speech with different embedding rates, that is, 100 bps, 200 bps, and 300 bps, by CNV-based, NID-based, and DN-based steganography. Of course, 200 bps and 300 bps may be omitted for CNV-based steganography because of its limited embedding capacity. And seven stegospeech sets are generated, among which one set is related to CNV-based steganographic method, and each of three sets is associated with NID-based and DN-based steganography, respectively. Moreover, only 23.85 kbit/s mode is used without loss of generality.

In each experiment, a pair of cover and stego speech sets is randomly divided into training and testing sets according to three kinds of training rates, that is, 0.4, 0.5, and 0.6. For example, if the training rate is 0.4, the training set contains 40% speech samples randomly chosen from each of the cover and stegospeech sets, and the remaining 60% samples go into the testing set. As described in [17, 18], LIBSVM [22] is used as a classifier, and radial basis function (RBF) kernel and grid-search technique are employed to obtain better classification performance. For Li et al.'s steganalytic method, the principal component analysis (PCA) is first used, as suggested in [18], to reduce the dimension of feature vectors to 300. Let the samples in cover speech set denote negatives and those in stego speech set stand for positives. Hence, the accuracy may be defined as follows:

$$\text{Accuracy} = \frac{1}{2} \times \left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{FP} + \text{TN}} \right), \tag{5}$$

where TP are true positives, TN are true negatives, FN are false negatives, and FP are false positives.

The steganalytic results are given in Table 2, It can be seen that when the embedding rate is 100 bps, the accuracy of detecting both CNV-based and DN-based methods is almost the same, say, 50% or so, while that of detecting

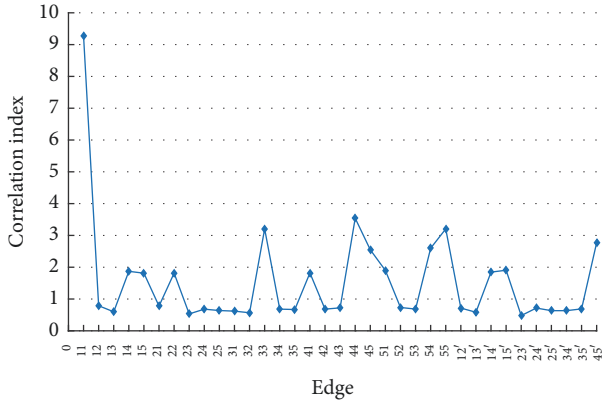


Figure 8: The correlation index of 1000 AMR-WB speeches, where the interframe edge ii connects two vertices $V_i[k]$ and $V_i[k+1]$ in two neighboring frames, and the intraframe edge ij' connects two vertices $V_i[k]$ and $V_j[k]$ in the same frame.

NID-based steganography increases to 60% when MFCC-based steganalytic method is applied. Moreover, there is an apparent increase in the accuracy of detecting NID-based hiding method with the embedding rate increases to 200 bps or 300 bps when Liu et al.'s methods (i.e., Markov and MFCC-based steganalytic methods) are applied. But the accuracy of steganalyzing our proposed method, DN-based steganography, stays at the same level of 50%. Therefore, the proposed method may defend against Liu et al.'s statistical steganalysis [17] even with higher embedding rates.

According to the definition of the correlation index given in [18], the experimental results of the correlation indices of 1000 AMR-WB speeches, which are randomly selected from "TIMIT," are shown in Figure 8. Based on these results, two strong quantization codeword correlation network (QCCN) models, say, SS-QCCN and RS-QCCN, can be constructed as illustrated in Figure 9. These two models are then used to steganalyze our proposed steganography. The steganalytic results are also presented in Table 2. It can be seen from Table 2 that the accuracy of both SS-QCCN and RS-QCCN is less than 50% for all of the AMR-WB stegospeeches. The possible reasons may be that only the second and third codebooks in the second stage are employed in the AMR-WB speech steganography, which means merely the vertices $V_2[k]$ and $V_3[k]$ in the kth frame may be changed during steganography while none of them are utilized in Li et al.'s steganalytic method except for the edge "33" in RS-QCCN model. Besides, we also used an adapted QCCN model (i.e., utilize edges "22," "33," and "$23'$") targeted at AMR-WB speech, but the accuracy is still less than 50%. It may be because the correlation of those edges is not strong enough for steganalysis according to Figure 8. Therefore, it is reasonable to conclude that the AMR-WB speech steganography can defend against the steganalytic method proposed in [18].

In order to visualize the detection performance, we give some receiver operating characteristic (ROC) curves of steganalyzing CNV-based steganography with 100 bps embedding rate and NID-based and DN-based steganography with 100 bps, 200 bps, and 300 bps embedding rates are

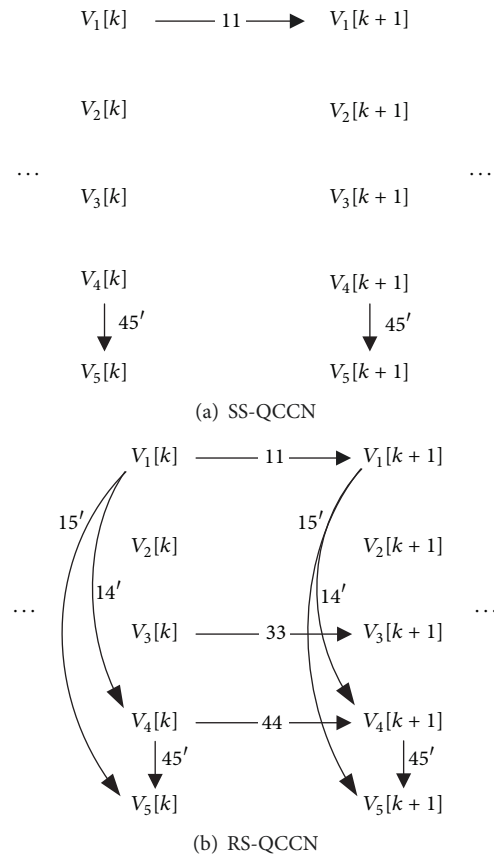


Figure 9: Two AMR-WB strong correlation network models.

provided in Figure 10 (ROC curves for SS-QCCN and RS-QCCN are omitted for these two methods fail to steganalyze AMR-WB steganography in spite of embedding capacity). It shows that all of the three steganographic methods can resist statistical steganalysis when the embedding rate is 100 bps. While the statistical steganalytic methods, especially MFCC-based steganalysis, may detect the existence of hidden message embedded with NID-based steganography when the embedding rate is above 100 bps, the proposed DN-based steganography may still have good security against both Markov-based and MFCC-based steganalysis.

5. Conclusion

The adaptive multirate wideband (AMR-WB) is a widely adapted format in mobile handsets and is also the recommended speech codec for VoLTE. AMR-WB speech may be a good candidate for cover medium in speech steganography. In this paper, a novel AMR-WB speech steganographic method is proposed. The experimental results demonstrated the effectiveness of our proposed method. The main contributions of this paper are as follows:

(1) A novel AMR-WB speech steganography is proposed based on diameter-neighbor codebook partition algorithm. It can provide higher capacity without noticeable decrease in speech quality and better

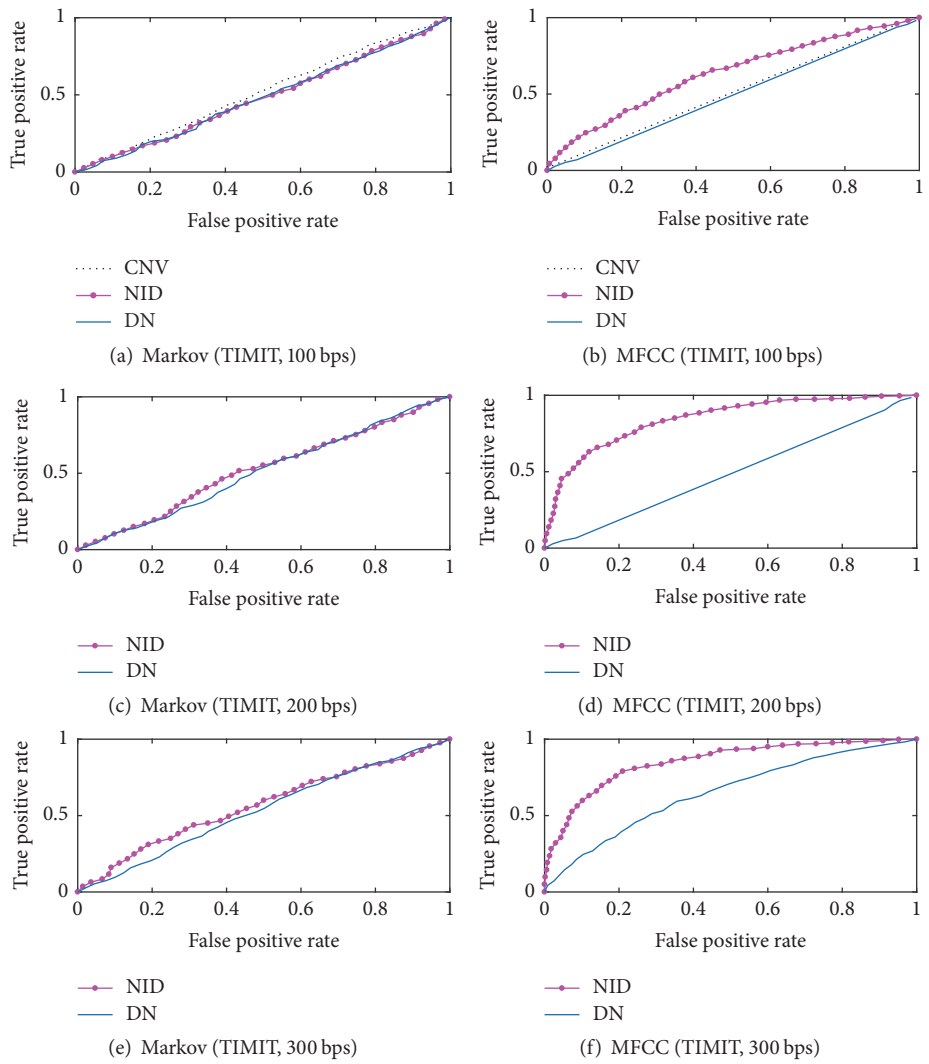


Figure 10: ROC curves for steganalysis of CNV-based, NID-based, and our proposed steganography (50% training rate).

performance against statistical steganalysis than NID-based method.

(2) Flexible embedding capacity may be easily achieved with different iterations of cluster merging. Twice the embedding capacity of CNV-based embedding method may be obtained with $N_i = 48$.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grant no. 61632013.

References

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3-4, pp. 313–335, 1996.

[2] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *Information Hiding*, R. Anderson, Ed., vol. 1174 of *Lecture Notes in Computer Science*, pp. 295–315, Springer Berlin Heidelberg, Berlin, Germany, 1996.

[3] K. Gopalan, "Audio steganography using bit modification," in *Proceedings of the 2003 International Conference on Multimedia and Expo, ICME 2003*, pp. I629–I632, USA, July 2003.

[4] K. Gopalan, S. Wenndt, S. Adams, and D. Haddad, "Audio steganography by amplitude or phase modification," in *Proceedings of the Security and Watermarking of Multimedia Contents V*, pp. 67–76, USA, January 2003.

[5] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1020–1033, 2003.

[6] L. Liu, M. Li, Q. Li, and Y. Liang, "Perceptually transparent information hiding in G.729 bitstream," in *Proceedings of the 2008 4th International Conference on Intelligent Information Hiding and Multiedia Signal Processing, IIH-MSP 2008*, pp. 406–409, China, August 2008.

[7] T. Xu and Z. Yang, "Simple and effective speech steganography in G.723.1 low-rate codes," in *Proceedings of the 2009*

International Conference on Wireless Communications and Signal Processing, WCSP 2009, China, November 2009.

[8] A. Shahbazi, A. H. Rezaie, and R. Shahbazi, "MELPe coded speech hiding on enhanced full rate compressed domain," in *Proceedings of the Asia Modelling Symposium 2010: 4th International Conference on Mathematical Modelling and Computer Simulation, AMS2010*, pp. 267–270, Malaysia, May 2010.

[9] A. Nishimura, "Data hiding in pitch delay data of the adaptive multi-rate narrow-band speech codec," in *Proceedings of the IIH-MSP 2009-2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 483–486, Japan, September 2009.

[10] B. Xiao, Y. Huang, and S. Tang, "An approach to information hiding in low bit-rate speech stream," in *Proceedings of the 2008 IEEE Global Telecommunications Conference, GLOBECOM 2008*, pp. 1940–1944, USA, December 2008.

[11] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[12] Y. F. Huang, S. Tang, and J. Yuan, "Steganography in inactive frames of VoIP streams encoded by source codec," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 296–306, 2011.

[13] Y. Huang, C. Liu, S. Tang, and S. Bai, "Steganography integration into a low-bit rate speech codec," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1865–1875, 2012.

[14] H. Miao, L. Huang, Z. Chen, W. Yang, and A. Al-Hawbani, "A new scheme for covert communication via 3G encoded speech," *Computers and Electrical Engineering*, vol. 38, no. 6, pp. 1490–1501, 2012.

[15] H. Tian, J. Liu, and S. Li, "Improving security of quantization-index-modulation steganography in low bit-rate speech streams," *Multimedia Systems*, vol. 20, no. 2, pp. 143–154, 2014.

[16] J. Liu, H. Tian, J. Lu, and Y. Chen, "Neighbor-index-division steganography based on QIM method for G.723.1 speech streams," *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 1, pp. 139–147, 2016.

[17] Q. Liu, A. H. Sung, and M. Qiao, "Derivative-based audio steganalysis," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 7, no. 3, article no. 18, 2011.

[18] S. Li, Y. Jia, and C.-C. J. Kuo, "Steganalysis of QIM Steganography in Low-Bit-Rate Speech Signals," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 5, pp. 1011–1022, 2017.

[19] ITU-T, Wideband Coding of Speech at around 16 Kbps Using Adaptive Multi-rate Wideband (AMR-WB), International Telecommunication Union Std. G.722.2, 2002.

[20] Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Net-works and Speech Codecs, International Telecommunication Union Std. P.862, 2001.

[21] Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs, International Telecommunication Union Std. P.862.2, 2007.

[22] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.