Hindawi
Mobile Information Systems
Volume 2017, Article ID 5418978, 18 pages
https://doi.org/10.1155/2017/5418978

Research Article

Detecting Steganography of Adaptive Multirate Speech with Unknown Embedding Rate

Hui Tian,[1] Jun Sun,[1] Yongfeng Huang,[2] Tian Wang,[1] Yonghong Chen,[1] and Yiqiao Cai[1]

[1]*College of Computer Science and Technology, National Huaqiao University, Xiamen 361021, China*
[2]*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Hui Tian; htian@hqu.edu.cn and Tian Wang; wangtian@hqu.edu.cn

Received 9 December 2016; Accepted 23 April 2017; Published 18 May 2017

Academic Editor: Elio Masciari

Steganalysis of adaptive multirate (AMR) speech is a significant research topic for preventing cybercrimes based on steganography in mobile speech services. Differing from the state-of-the-art works, this paper focuses on steganalysis of AMR speech with unknown embedding rate, where we present three schemes based on support-vector-machine to address the concern. The first two schemes evolve from the existing image steganalysis schemes, which adopt different global classifiers. One is trained on a comprehensive speech sample set including original samples and steganographic samples with various embedding rates, while the other is trained on a particular speech sample set containing original samples and steganographic samples with uniform distributions of embedded information. Further, we present a hybrid steganalysis scheme, which employs Dempster–Shafer theory (DST) to fuse all the evidence from multiple specific classifiers and provide a synthesized detection result. All the steganalysis schemes are evaluated using the well-selected feature set based on statistical characteristics of pulse pairs and compared with the optimal steganalysis that adopts specialized classifiers for corresponding embedding rates. The experimental results demonstrate that all the three steganalysis schemes are feasible and effective for detecting the existing steganographic methods with unknown embedding rates in AMR speech streams, while the DST-based scheme outperforms the others overall.

1. Introduction

Steganography is an ancient but effective technique for covert communications through hiding confidential messages into seemingly innocent carriers with imperceptible distortion. Although its history can date back to 440 BC [1], its candidate carriers have been ceaselessly evolving with the elapsing of years [2]. Over the last years, the steganographic carriers have developed from image [3, 4] to almost all media forms (e.g., video [5, 6], audio [7, 8], text [9, 10], network protocol [11, 12], and Voice over IP [13–16]). However, steganography is a double-edged sword. Illegal usage of this technique would facilitate cybercrime activities and thereby pose a great threat to information security. Thus, its countermeasure, steganalysis, has been also attracting considerable attention [17–25], whose purpose is to detect potential steganographic behaviors effectively.

In today's mobile world, adaptive multirate (AMR) codec has become a well-known and important compression standard for speech coding and been widely employed in not only 3G and 4G speech services [26–28] but also various mobile instant messaging apps (such as WhatsApp, Snapchat, LINE, and WeChat). Moreover, it is also a popular file format for storing AMR-encoded spoken audio supported by almost all mobile communication devices. Due to its increasing popularity and broad influence in mobile communications, AMR speech is spontaneously considered as an ideal carrier by the steganographic research community, and some relevant studies have been successfully performed [29–33].

AMR is a typical codec based on an algebraic code-excited linear prediction algorithm, in which algebraic codebook indices (ACIs), also called fixed codebook indices (FCIs), occupy a large percentage of each speech frame [26–28]. Taking the AMR speech codec at 12.2 kbps mode [28], for example, 140 bits out of 244 frame bits is allocated to FCIs,

suggesting that FCIs account for a large proportion (57.38%) of all frame bits [33]. Therefore, they are popularly regarded as nice candidates for steganographic carriers in the existing studies [29–33]. Geiser and Vary [29] first incorporated information hiding into speech coding of the AMR codec by modifying the fixed-codebook-search algorithm. Specifically, two secret bits can be hidden into a track pulse through limiting the searching range of the second FCI to two of eight candidate values. Their experimental results demonstrate that this method can offer a steganographic bandwidth of 2 kbit/s for the AMR speech codec at 12.2 kbps mode, while guaranteeing an imperceptible impact on speech quality and fairly small computational complexity. Moreover, following the similar idea, Miao et al. [30] proposed an adaptive suboptimal pulse combination constrained method for steganography in the AMR speech stream. Their main advantage over the previous method is enabling regulation of the steganographic capacity by introducing an embedding factor η. For example, for the AMR speech codec at 12.2 kbps mode, η can be typically set as 1, 2, or 4, so the steganographic bandwidths are correspondingly 1, 2, or 3 kbit/s [32, 33]. It has been demonstrated that, by choosing a befitting η, this method can achieve a nice trade-off between the distortion of speech quality and the embedding capacity [30].

To prevent potential cybercrimes based on the above steganographic methods, some steganalysis studies have accordingly been conducted. Miao et al. [31] first presented two steganalysis methods for AMR speech. One is called Markov-based method that adopts Markov transition probabilities to evaluate the relationship between pulse positions in each track, while the other is Entropy-based method that employs the joint entropy and the conditional entropy to measure the uncertainty of pulse positions [31]. However, the above two kinds of statistical features are not accurate enough for characterizing AMR speech, because they ignore the fact that the pulse positions may often be interchanged in the AMR encoding process [33]. Moreover, Ren et al. [32] presented a steganalysis method called Fast-SPP, which employs probabilities of same pulse positions (SPP) as the features to detect the existing steganographic methods [29, 30]. However, the SPP features only reflect the distributions of two track-pulses being in the same position, which are not comprehensive enough to characterize AMR speech [33]. Particularly, if a steganographic method designedly abandons the track-pulses with the same positions and the ones that would be the same after the embedding operation, Fast-SPP could not detect any abnormalities [33]. Therefore, in our previous work [33], we presented more accurate and more complete features for steganalysis of AMR speech. To avoid the impact induced by possible interchange of pulse positions in each track, we employ the statistical features of pulse pairs to characterize AMR speech, including the probability distributions of pulse pairs reflecting the long-term distribution of speech signals, Markov transition probabilities of pulse pairs depicting the short-term invariant characteristic of speech signals, and joint probability matrices of pulse pairs characterizing the track-to-track correlation [33]. Moreover, to optimize the feature set as well as cut down the dimension, a feature selection mechanism using adaptive boosting (AdaBoost) [34–38] is designed. Employing the selected optimal feature set, a support-vector-machine (SVM) based steganalysis of AMR speech was presented. The experimental results show that the proposed method significantly outperforms the previous ones.

However, all the above steganalysis methods assume that the embedding rate (also called the usage rate of the cover, which is the ratio between the practical embedded bits and the total number of cover bits) of steganographic samples in a given test set is exactly known. In other words, they generally train specific classifiers for steganographic samples with predefined embedding rates, and each specialized classifier is expected to detect the steganographic samples with the corresponding embedding rate. Unfortunately, in practice, we usually cannot ascertain whether the steganographic operation has been performed on a given sample, let alone knowing the concrete embedding rate. Thus, it is necessary and significant to develop detection technique for steganography with unknown embedding rate [39–41]. To the best of our knowledge, this work in this paper is the first one dedicated to address the concern in the speech steganalysis field. In the image steganalysis field, however, some pioneer researchers have presented two useful schemes for detecting image steganography with unknown embedding rate. Both the two schemes adopt global classifiers based on a machine-learning algorithm (e.g., SVM) as the detectors, but the components of their training set are different. Specifically, the training set of the first scheme includes original (untouched) samples and steganographic samples with various embedding rates [40, 41], while that of the other one consists of original samples and steganographic samples with uniform distributions of embedded data [40]. In this work, we would like to attempt to first extend the two existing schemes to AMR speech steganalysis with unknown embedding rate employing the state-of-the-art steganalysis features presented in our recent work [33]. Besides, incorporating with Dempster–Shafer theory (DST) [42, 43], we further present a hybrid steganalysis scheme for AMR speech based steganography with unknown embedding rate. DST, also called evidence theory, is a well-established framework for uncertain reasoning, which can fuse available evidence from different sources and achieve a level of belief (confidence; trust) by considering all of them [42–46]. The main idea behind the presented steganalysis scheme is employing an algorithm based on DST to combine all the evidence from a set of classifiers intended for detecting steganographic approaches with specific embedding rates and accordingly providing a synthesized judgement for having or not having hidden information. All the three steganalysis schemes are evaluated with a great number of AMR-encoded speech samples and compared with the optimal steganalysis that uses every specialized classifier to detect the steganography with the corresponding embedding rate. The experimental results show that all these steganalysis schemes are feasible and efficient for detecting the state-of-the-art steganographic methods with unknown embedding rates in AMR speech streams, while the DST-based scheme can achieve better detection performance than the other ones.

The remaining of this paper is organized as follows. To make this paper self-contained, Section 2 first reviews the state-of-the-art steganalysis features based on statistical characteristics of pulse pairs. Section 3 presents the

three steganalysis schemes for detecting AMR speech based steganography with unknown embedding rate. Section 4 evaluates the performance of the three steganalysis schemes by a set of comprehensive experiments, which is followed by concluding remarks given in Section 5.

2. Steganalysis Features Based on Statistical Characteristics of Pulse Pairs

In this work, all the presented steganalysis schemes would adopt the state-of-the-art detection features based on statistical characteristics of pulse pairs for AMR speech, which consists of long- term features, short-term features, and track-to-track features [33].

The probability distributions of the pulse pairs are employed to depict the long-term features of AMR speech. Assume that the given AMR speech sample to be detected has N subframes and each subframe contains T tracks. For the jth ($0 \le j \le T-1$) track in the ith ($0 \le i \le N-1$) subframe, two pulse positions as a pulse pair $(\rho_{i,j}, \rho_{i,j+T})$ can be extracted. For a pulse pair (α, β), its probability (denoted by $P_{(\alpha,\beta)}$) appearing in all subframes can be determined as follows:

$$P_{(\alpha,\beta)} = \frac{\sum_{i=0}^{N-1} \left(\left(\delta\left(\rho_{i,j} = \alpha\right) \& \delta\left(\rho_{i,j+T} = \beta\right) \right) \| \left(\delta\left(\rho_{i,j} = \beta\right) \& \delta\left(\rho_{i,j+T} = \alpha\right) \right) \right)}{N}, \tag{1}$$

where "&" is the binary AND operation, "‖" is the binary OR operation, and $\delta(x = y)$ is a characteristic function defined as follows:

$$\delta\left(x = y\right) = \begin{cases} 1, & x = y \\ 0, & x \neq y. \end{cases} \tag{2}$$

Let the number of candidate positions for every pulse in each track be τ; the number of the possible pulse pairs (denoted by ψ) is

$$\psi = \tau^2 - \frac{\tau\left(\tau - 1\right)}{2} = \frac{\tau^2 + \tau}{2}. \tag{3}$$

Therefore, there are $\psi \times T$ pulse pairs in each subframe. That is to say, the dimension of the long-term feature set (LTFS) for pulse pairs is $\psi \times T$.

According to the short-term invariance of speech signals [47], the pulse pair of a track in the current subframe is bound to have a strong correlation with the one of the same track in the prior subframe [33]. In this sense, for the ith ($0 \le i \le T-1$) pulse pairs (i.e., the pulse pairs of the ith tracks) in all subframes, the sequence of pulse-position pairs $S_i = \{s_{i,0}, s_{i,1}, \ldots, s_{i,N-1}\}$ can be considered as a Markov chain. Accordingly, the Markov transition matrix (MTM) can be employed to describe the transitive correlation of pulse-pair states in the given track. Moreover, as a first-order Markov chain, S_i satisfies

$$P\left(s_{i,j} \mid s_{i,0}, s_{i,1}, \ldots, s_{i,j-1}\right) = P\left(s_{i,j} \mid s_{i,j-1}\right). \tag{4}$$

In the ith tracks of all subframes, the probability $P((\alpha_1, \beta_1) \mid (\alpha_2, \beta_2))$ that the pulse pair (α_1, β_1) occurs after the pulse pair (α_2, β_2) is

$$\begin{aligned} &P\left((\alpha_1, \beta_1) \mid (\alpha_2, \beta_2)\right) \\ &\quad = P\left(s_{i,j} = (\alpha_1, \beta_1) \mid s_{i,j-1} = (\alpha_2, \beta_2)\right) \\ &\quad = \frac{P\left(s_{i,j} = (\alpha_1, \beta_1), s_{i,j-1} = (\alpha_2, \beta_2)\right)}{P\left(s_{i,j-1} = (\alpha_2, \beta_2)\right)}. \end{aligned} \tag{5}$$

Further, the MTM for the ith track (denoted by $\mathbf{M}_i$) can be determined as follows:

$$\begin{aligned} &\mathbf{M}_i \\ &= \begin{bmatrix} P\left(v_{i,0} \mid v_{i,0}\right) & P\left(v_{i,0} \mid v_{i,1}\right) & \cdots & P\left(v_{i,0} \mid v_{i,\psi-1}\right) \\ P\left(v_{i,1} \mid v_{i,0}\right) & \ddots & & \\ \vdots & & \ddots & \\ P\left(v_{i,\psi-1} \mid v_{i,0}\right) & \cdots & \cdots & P\left(v_{i,\psi-1} \mid v_{i,\psi-1}\right) \end{bmatrix}, \end{aligned} \tag{6}$$

where ψ is the number of all possible pulse-position pairs for the ith track that can be determined as (3); $v_{i,k} = (u_{i,x}, u_{i,y})$ is the kth ($0 \le k \le \psi - 1$) possible pulse-position pair for the ith track, where $u_{i,x}$ and $u_{i,y}$ are the potential pulse positions for the ith track. Moreover, assume that there are τ candidate positions for each pulse; x, y, and k satisfy the following relation:

$$\begin{aligned} &k \\ &= \begin{cases} y, & x = 0,\ 0 \le y \le \tau - 1 \\ \sum_{i=0}^{x-1} (\tau - i) + y - x, & 1 \le x \le \tau - 1,\ 0 \le y \le \tau - 1. \end{cases} \end{aligned} \tag{7}$$

Since there are ψ possible pulse-position pairs in each track, the size of each MTM is $\psi \times \psi$. Taking the MTMs of all T tracks into account, the dimension of the feature set would be very large. However, the characteristics of all the MTMs are similar. Therefore, we often adopt the average Markov transition probabilities (MTPs) as the steganalysis features instead. Apparently, the average MTM (denoted by $\mathbf{M}$) is determined as

$$\mathbf{M} = \frac{\sum_{i=0}^{T-1} \mathbf{M}_i}{T}. \tag{8}$$

Accordingly, the dimension of the short-term feature set (STFS) for pulse pairs is $\psi \times \psi$.

Furthermore, the joint probability matrices of the pulse pairs in different tracks are employed to characterize the

Table 1: The composition of the reduced feature set [33].

Feature sets	Original dimensions	Reduced dimensions	Proportion in the reduced feature set
LTFS	180	58	11.65%
STFS	1296	250	50.20%
TTFS	1296	190	38.15%
Total	2772	498	100%

track-to-track features. To be specific, for the pulse pair of the ith track and the one of the jth track ($0 \le i, j \le T-1$), the joint probability matrix (JPM) $\mathbf{J}_{i,j}$ is

$$\mathbf{J}_{i,j} = \begin{bmatrix} P\left(v_{i,0}, v_{j,0}\right) & P\left(v_{i,0}, v_{j,1}\right) & \cdots & P\left(v_{i,\psi-1}, v_{j,\psi-1}\right) \\ P\left(v_{i,1}, v_{j,0}\right) & \ddots & & \\ \vdots & & \ddots & \\ P\left(v_{i,\psi-1}, v_{j,0}\right) & \cdots & \cdots & P\left(v_{i,\psi-1}, v_{j,\psi-1}\right) \end{bmatrix} \quad (9)$$

where ψ is the number of all possible pulse-position pairs for the ith track that can be determined by (3); $v_{i,k}$ ($v_{j,k}$) is the kth ($0 \le k \le \psi - 1$) possible pulse-position pair for the ith (jth) track; and $P(v_{i,k}, v_{j,h})$ is the joint probability of $v_{i,k}$ and $v_{j,h}$ ($0 \le k, h \le \psi - 1$). Specifically, the joint probability of the pulse-position pair (α_i, β_i) in the ith track and the pulse-position pair (α_j, β_j) in the jth track can be determined as follows:

$$P\left(\left(\alpha_i, \beta_i\right), \left(\alpha_j, \beta_j\right)\right) = \frac{\sum_{k=0}^{N-1} \left(\delta\left(u_{k,i} = \left(\alpha_i, \beta_i\right)\right) \& \delta\left(u_{k,j} = \left(\alpha_j, \beta_j\right)\right)\right)}{N}, \quad (10)$$

where N is the number of the subframes, $u_{k,i}$ ($u_{k,j}$) is the pulse pair in the ith (jth) track of the kth subframe ($0 \le k \le N-1$), $\delta(x = y)$ is a characteristic function defined as (2), and "&" is the binary AND operation.

Like STFS above, we adopt the average JPM as the track-to-track feature set (TTFS) instead of all JPMs to reduce the computational complexity. Specifically, the average JPM (denoted by $\mathbf{J}$) is

$$\mathbf{J} = \frac{2\sum_{i=0}^{T-1}\sum_{j=i+1}^{T-1} \mathbf{J}_{i,j}}{T(T-1)}. \quad (11)$$

Apparently, the dimension of the TTFS is $\psi \times \psi$. Accordingly, the total dimension of all the three feature sets is $\psi \times T + 2\psi \times \psi$. Taking the AMR speech codec at 12.2 kbps mode as an example, there are five tracks in each subframe (i.e., $T = 5$), where two pulses share eight candidate positions, that is, $\tau = 8$. Thus, there are $\psi = 36$ pulse pairs in each track, and the total dimension of all feature sets is 2772. These features are still too large to be directly adopted in the machine-learning based steganalysis scheme, since very-high-dimensional features would not only cause huge computational costs in the detection phase but also be more likely to induce overfitting in the training phase [33]. Thus, a feature selection mechanism based on AdaBoost [34–38] is employed to optimize the feature set as well as reduce the dimension. In the previous work [33], by this mechanism a reduced feature set with the 498 most effective features is obtained for the AMR speech codec at 12.2 kbps mode, of which the composition is shown in Table 1. Given that the excellent effectiveness of the selected feature set for steganalysis of AMR speech has been verified, we directly employ it in this paper.

3. Steganalysis Schemes for Detecting AMR Speech Steganography with Unknown Embedding Rate

In this section, we present three steganalysis schemes for detecting AMR speech based steganography with unknown embedding rate employing SVM, which is a well-known machine-learning tool with excellent performance on classification [48–53] and popularly employed in the steganalysis field [17–20, 24, 25, 33]. The first two schemes are extended from the existing image steganalysis schemes [40–42], which both employ global classifiers to detect the steganography but adopt different training sets. As depicted in Figures 1 and 2, the first scheme trains the global classifier using a comprehensive speech sample set, including original samples and steganographic samples with various embedding rates, while the second one adopts a particular speech sample set, consisting of original samples and steganographic samples with uniform distributions of embedded data, to train the global classifier. For ease of description, we denote the first scheme as GC-M, meaning that it trains the global classifier on mixed samples with various embedding rates, and the second scheme as GC-U, meaning that it trains the global classifier on particular samples with uniform distributions of embedded data. In this work, for each AMR speech based steganographic method, the training set of GC-M involves the steganographic samples with the embedding rates from 10% to 100%. Moreover, to obtain the steganographic AMR speech samples with uniform distributions of embedded data for GC-U, we choose the tracks for hiding information in each subframe in a uniform random manner during the steganographic processes.

In addition, we further present a steganalysis scheme based on Dempster–Shafer theory (DST) for AMR speech based steganography with unknown embedding rate, as shown in Figure 3. To make the paper self-contained, we first review DST briefly. DST is a well-established mathematical theory of evidence first presented by Dempster [42] and Shafer [43], which can combine the evidence from different

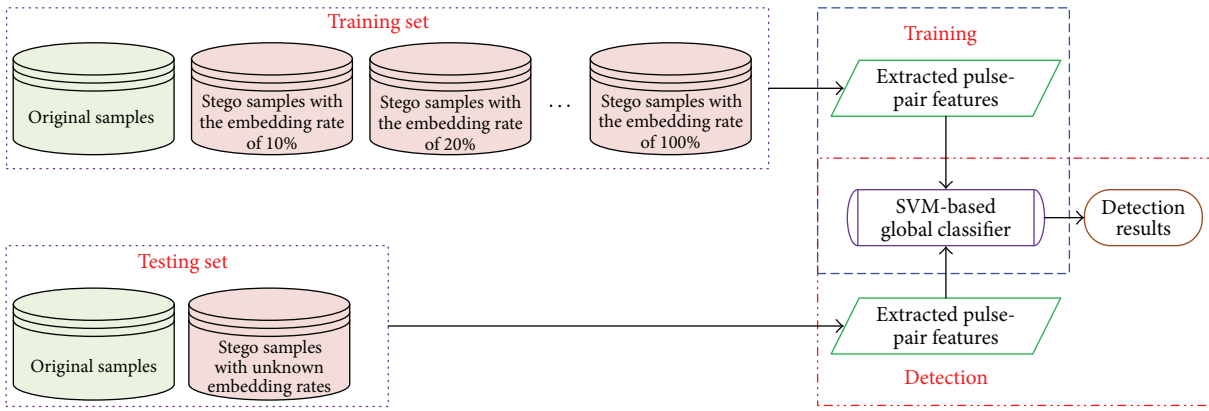


Figure 1: The first steganalysis scheme (GC-M) that trains the global classifier on mixed samples with various embedding rates.

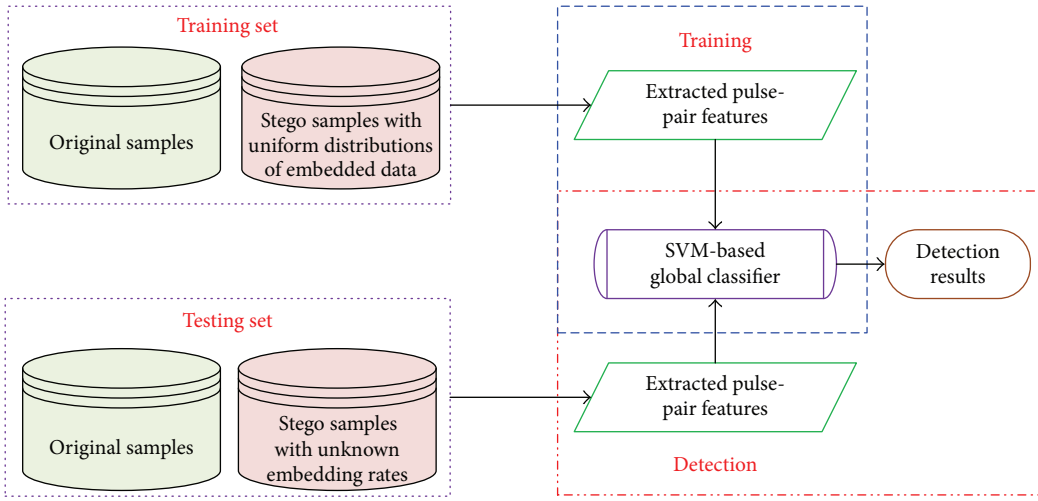


Figure 2: The second steganalysis scheme (GC-U) that trains the global classifier on particular samples with uniform distributions of embedded data.

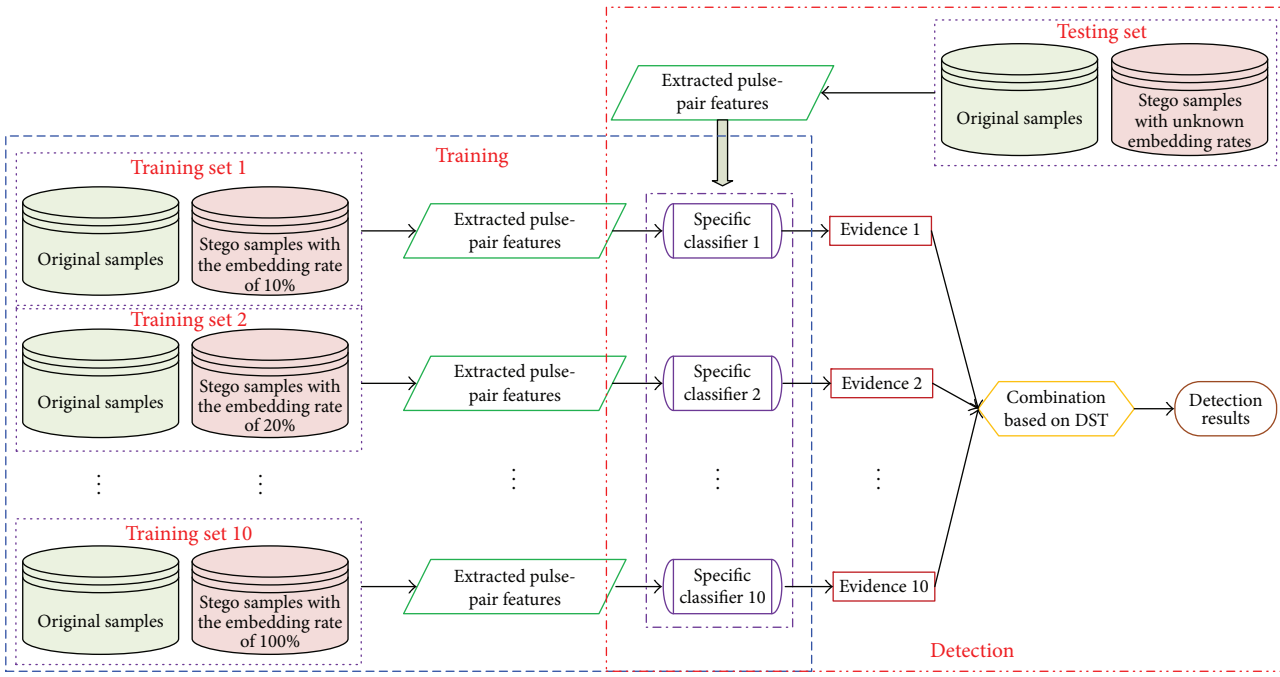


Figure 3: The steganalysis scheme based on Dempster–Shafer theory.

sources to obtain the probability of a certain event [43]. Owing to its powerful reasoning function based on evidence combination, DST has been popularly employed in many fields, such as information fusion [44], classification [45], and intrusion detection [46].

Generally, DST is constructed on a finite set of n possible elements (denoted by $\Theta = \{e_1, e_2, \ldots, e_n\}$) under consideration, called a frame of discernment. Note that Θ is exhaustive, and all elements in Θ are mutually exclusive. Let 2^Θ be the set including all possible subsets of Θ. A mass function for assigning a probability mass to each element, also called basic probability assignment, is defined as follows:

$$\begin{aligned} & m(X) : 2^\Theta \longrightarrow [0,1], \\ \text{s.t.} \quad & m(\emptyset) = 0, \\ & \sum_{X \in 2^\Theta} m(X) = 1, \end{aligned} \tag{12}$$

where $\emptyset$ is the empty set. Each nonempty subset X of Θ is called a focal element, and its mass function $m(X)$ represents the exact belief for the proposition described by X. Further, the belief function for a subset X of Θ, denoted by $\mathrm{Bel}(X)$, is the sum of the mass values of all its subsets; namely,

$$\mathrm{Bel}(X) = \sum_{Y \subseteq X} m(Y), \quad \left(Y \in 2^\Theta\right). \tag{13}$$

The plausibility function for a subset X of Θ, denoted by $\mathrm{Pl}(X)$, is the sum of the mass values of all the subsets of Θ that intersect X; namely,

$$\mathrm{Pl}(X) = \sum_{Y \cap X \neq \emptyset} m(Y) = 1 = \mathrm{Bel}\left(\overline{X}\right), \quad \left(Y \in 2^\Theta\right). \tag{14}$$

Moreover, DST provides a combination rule to obtain a synthesized belief value for an element by fusing the evidence from different sources. Formally, assume that $m_1, m_2, \ldots, m_n$ are mass functions for a subset X of Θ from different evidence, the combination rule can be stated as follows:

$$m(X) = \begin{cases} 0, & X = \emptyset \\ \dfrac{\sum_{\bigcap_{i=1}^n X_i = A} \prod_{i=1}^n m_i(X_i)}{1-k}, & X \neq \emptyset, \end{cases} \tag{15}$$

where k is a conflict factor that measures the degree of conflict for all the evidence and can be determined as follows:

$$k = \sum_{\bigcap_{i=1}^n X_i = \emptyset} \prod_{i=1}^n m_i(X_i). \tag{16}$$

Note that if $k = 1$, all the available evidence is highly contradictory and thereby cannot be directly combined.

In our work, the frame of discernment Θ for detecting AMR speech based steganography with unknown embedding rate is defined as $\Theta = \{C, S\}$, where C and S represent the cover (original) and steganographic samples, respectively, and accordingly, $2^\Theta = \{\emptyset, \{C\}, \{S\}, \{C, S\}\}$. As shown in Figure 3, we adopt the specific SVM-based classifiers for the embedding rates from 10% to 100% as ten independent evidence sources. That is to say, there are ten mass functions from the specific SVM-based classifiers for various embedding rates. Specifically, the ith mass function from the classifier for the embedding rate of $10\% \times i$ is defined as follows:

$$\begin{aligned} m_i(\{C\}) &= P_C(SC_i), \\ m_i(\{S\}) &= P_S(SC_i), \\ m_i(\Theta) &= 0, \end{aligned} \tag{17}$$

where $P_C(SC_i)$ ($P_S(SC_i)$) is the confidence probability for the test sample belonging to the cover (steganographic) classification, offered by the SVM-based classifier for the embedding rate of $10\% \times i$.

According to (15), we can get

$$\begin{aligned} m(\{C\}) &= \frac{\prod_{i=1}^{10} P_C(SC_i)}{\prod_{i=1}^{10} P_C(SC_i) + \prod_{i=1}^{10} P_S(SC_i)}, \\ m(\{S\}) &= \frac{\prod_{i=1}^{10} P_S(SC_i)}{\prod_{i=1}^{10} P_C(SC_i) + \prod_{i=1}^{10} P_S(SC_i)}, \\ m(\{C, S\}) &= 0. \end{aligned} \tag{18}$$

Incorporating (13) and (14), we can further obtain

$$\begin{aligned} \mathrm{Bel}(\{C\}) &= \mathrm{Pl}(\{C\}) = m(\{C\}), \\ \mathrm{Bel}(\{S\}) &= \mathrm{Pl}(\{S\}) = m(\{S\}). \end{aligned} \tag{19}$$

Thus, we can finally make a decision by comparing $m(\{C\})$ and $m(\{S\})$. That is, for a test sample, its classification (denoted by Λ) can be determined as follows:

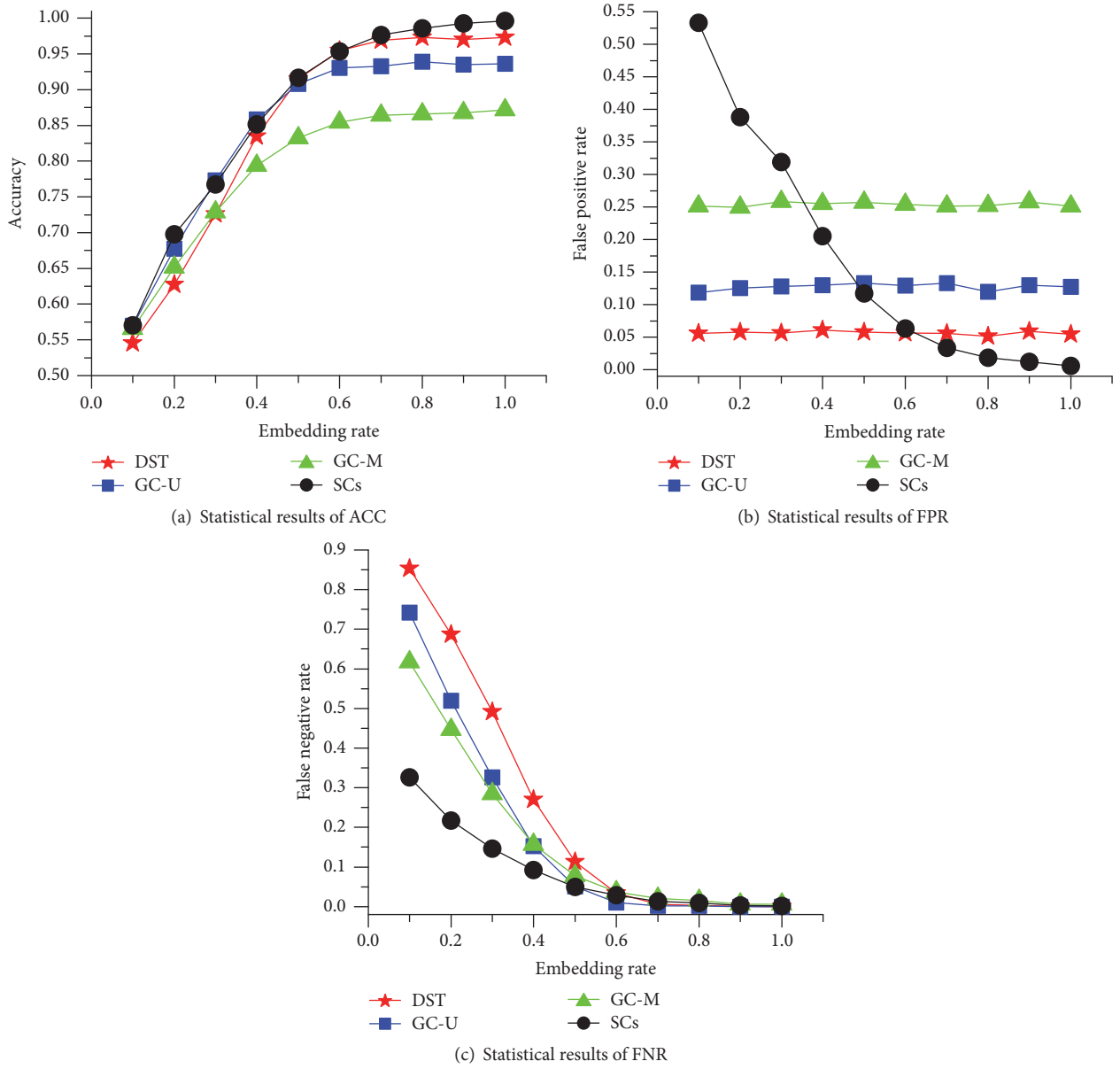
$$\Lambda = \begin{cases} C, & \text{If } m(\{C\}) \geq m(\{S\}), \\ S, & \text{Otherwise.} \end{cases} \tag{20}$$

Table 2: The components of adopted ten-second speech samples.

Category	Chinese		English		Total
	Male	Female	Male	Female	
Number	743	739	905	979	3366

4. Performance Evaluation and Analysis

In this paper, all the SVM-based classifiers are implemented employing LibSVM [49], a popular open-source software library for SVM. Specifically, the classifiers are constructed on the linear SVM (C-style) with RBF kernel, in which the default parameters are employed, that is, $c = 1$ and $g = 1/1064$. Moreover, we collect a total of 3366 ten-second speech samples from audio materials for language learning, of which the components are shown in Table 2. Without loss of generality, we typically choose the AMR codec at 12.2 kbps mode as the cover codec. In the experiments, all steganalysis schemes are evaluated on through detecting



(a) Statistical results of ACC

(b) Statistical results of FPR

(c) Statistical results of FNR

Figure 4: Experimental results for detecting Geiser's method.

the state-of-the-art steganographic methods, namely, Geiser's method [29] and Miao's methods at the modes of $\eta = 1$, 2, and 4 [30]. Prior to the steganographic experiments, we randomly select a half (1683) of the total speech samples as the cover sample set for training (CSST) and take the remaining samples as the cover sample set for detection (CSSD). In the steganographic experiments, the embedded messages are all randomly produced. For the three steganalysis schemes, we define their training sets as follows:

(i) The training set of the first scheme (GC-M): for each steganographic method, the training set includes 1400 speech samples randomly selected from CSST and 1400 mixed steganographic speech samples at the embedding rates from 10% to 100%, where there are 140 speech samples at each embedding rate.

(ii) The training set of the second scheme (GC-U): for each steganographic method, the training set includes 1400 speech samples randomly selected from CSST and 1400 steganographic speech samples with uniform distributions of embedded messages.

(iii) The training sets of the third scheme (DST-based scheme): for each steganographic method, it is necessary to train the specific classifiers for different embedding rates. Accordingly, for each embedding rate, a training set needs to be created, which includes 1400 speech samples randomly selected from CSST and 1400 samples generated by performing the given steganographic method at the corresponding embedding rate.

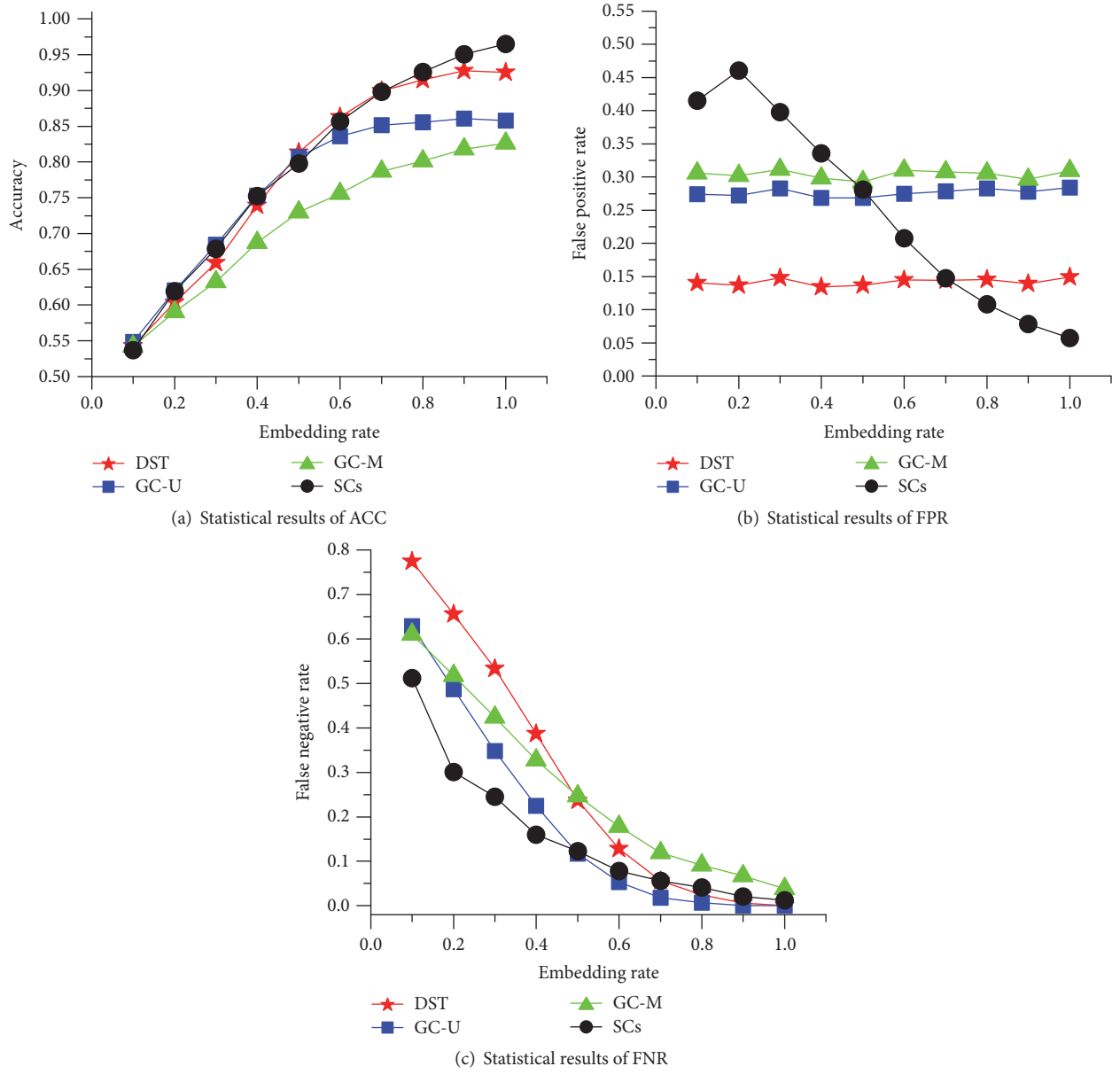


Figure 5: Experimental results for detecting Miao's method ($\eta = 1$).

In addition, to evaluate the steganalysis performance at the various embedding rates from 10% to 100%, we create ten detection sample sets for each steganographic method. Specifically, for each embedding rate, the detection sample set consists of 1400 speech samples randomly chosen from CSSD and 1400 speech samples generated by performing the given steganographic method at the corresponding embedding rate. Further, we evaluate the performance of the three steganalysis schemes by comparing them with the steganalysis based on specific classifiers (SCs) [33]. In all steganalysis experiments, we make the statistical analyses on accuracy (ACC, the proportion of true detection results), false positive rate (FPR, the proportion of false positives out of all negatives), and false negative rate (FNR, the proportion of false negatives out of all positives).

Figures 4, 5, 6, and 7, respectively, show the experimental results of detecting all the four steganographic methods for the ten-second speech samples at the embedding rates from 10% to 100%, from which we can learn that all the three steganalysis schemes in this paper are feasible and effective, while there are some differences in their detection performance. To be specific, the DST-based scheme outperforms GC-U and GC-M on the whole as also shown in Tables 3–6, since the detection accuracies of the DST-based scheme are better than the others in most cases and closer to those of the scheme based on SCs overall. Moreover, the FPRs of the DST-based scheme are smaller than the others in any case. By the way, for a given steganographic method, the FNRs of each steganalysis scheme presented in this paper are almost the same at any embedding rate, since each scheme adopts the

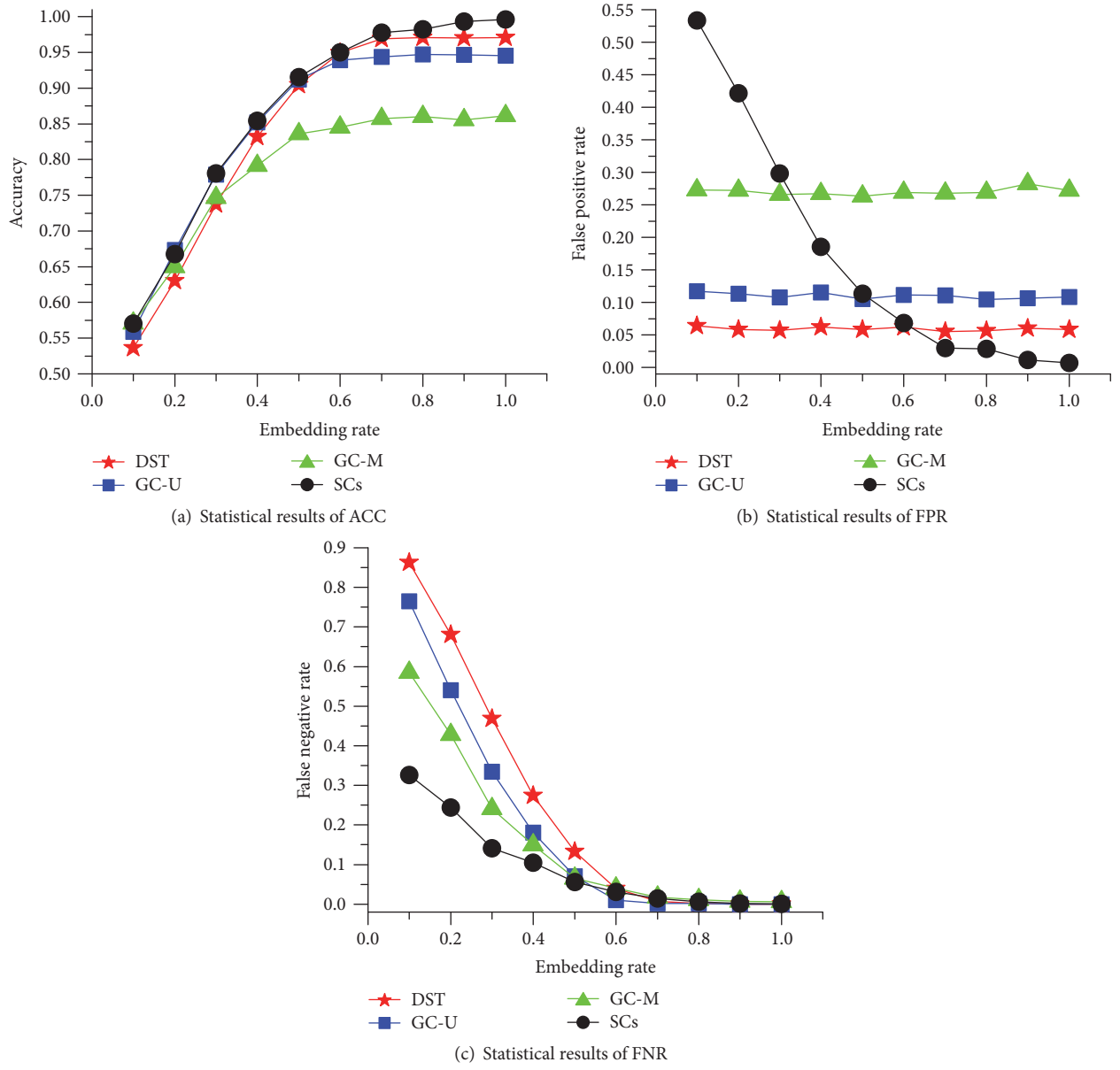


Figure 6: Experimental results for detecting Miao's method ($\eta = 2$).

Table 3: Statistical results of accuracies for detecting Geiser's method.

Embedding rate	GC-M	GC-U	DST	SCs	Difference between DST and other schemes		
					GC-M	GC-U	SCs
10%	56.57%	56.96%	54.57%	57.04%	−2.00%	−2.39%	−2.46%
20%	65.18%	67.75%	62.75%	69.75%	−2.43%	−5.00%	−7.00%
30%	72.86%	77.32%	72.61%	76.71%	−0.25%	−4.71%	−4.11%
40%	79.39%	85.86%	83.46%	85.14%	4.07%	−2.39%	−1.68%
50%	83.25%	90.82%	91.46%	91.68%	8.21%	0.64%	−0.21%
60%	85.46%	93.04%	95.46%	95.36%	10.00%	2.43%	0.11%
70%	86.39%	93.29%	96.89%	97.68%	10.50%	3.61%	−0.79%
80%	86.61%	93.93%	97.29%	98.61%	10.68%	3.36%	−1.32%
90%	86.79%	93.50%	97.00%	99.25%	10.21%	3.50%	−2.25%
100%	87.14%	93.64%	97.29%	99.64%	10.14%	3.64%	−2.36%

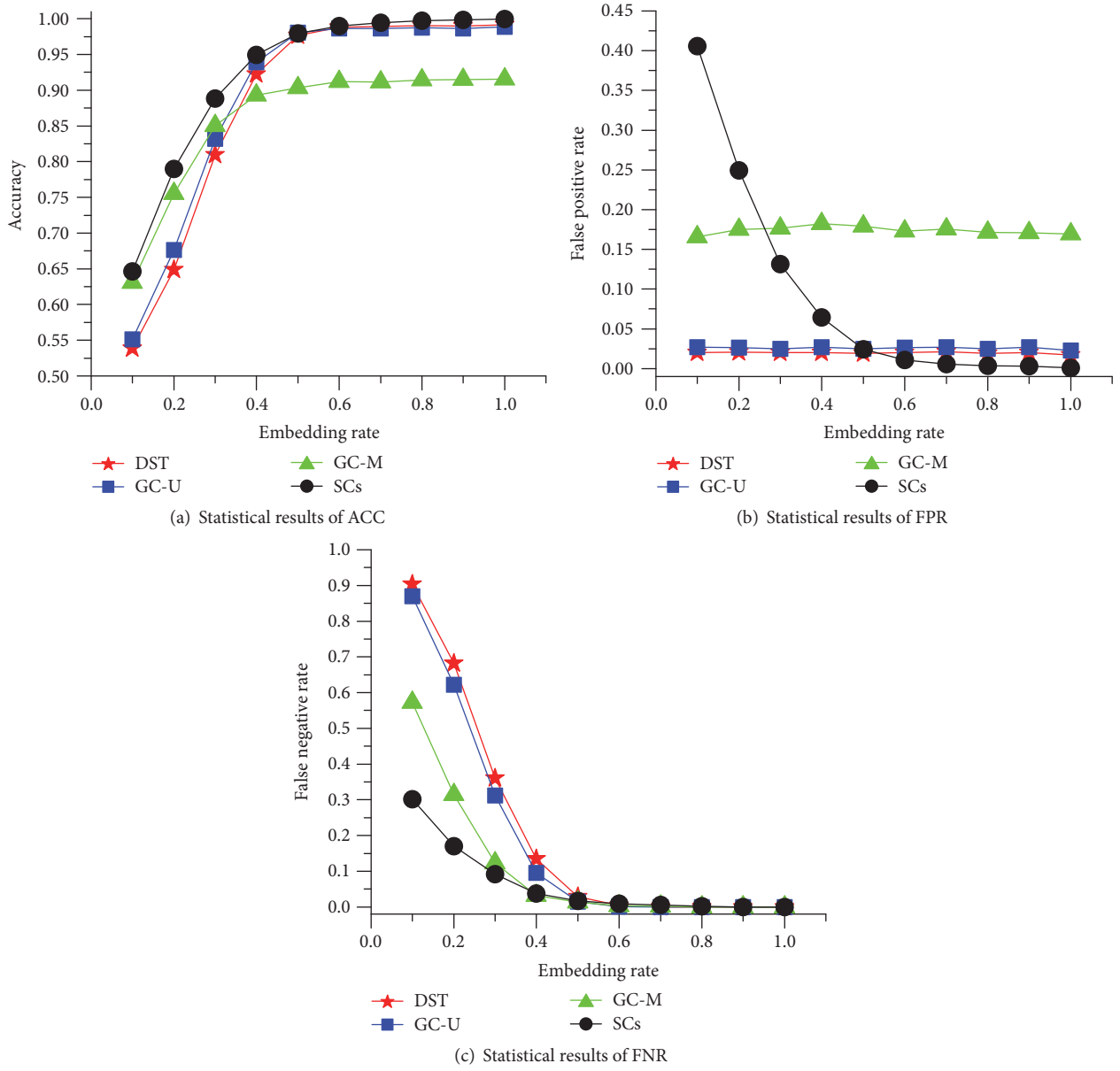


Figure 7: Experimental results for detecting Miao's method ($\eta = 4$).

Table 4: Statistical results of accuracies for detecting Miao's method ($\eta = 1$).

Embedding rate	GC-M	GC-U	DST	SCs	Difference between DST and other schemes		
					GC-M	GC-U	SCs
10%	54.21%	54.86%	54.25%	53.68%	0.04%	−0.61%	0.57%
20%	59.04%	62.07%	60.39%	61.93%	1.36%	−1.68%	−1.54%
30%	63.25%	68.46%	65.93%	67.86%	2.68%	−2.54%	−1.93%
40%	68.71%	75.32%	73.93%	75.21%	5.21%	−1.39%	−1.29%
50%	73.00%	80.71%	81.36%	79.82%	8.36%	0.64%	1.54%
60%	75.57%	83.57%	86.32%	85.68%	10.75%	2.75%	0.64%
70%	78.68%	85.18%	89.96%	89.82%	11.29%	4.79%	0.14%
80%	80.14%	85.54%	91.50%	92.57%	11.36%	5.96%	−1.07%
90%	81.82%	86.07%	92.75%	95.04%	10.93%	6.68%	−2.29%
100%	82.61%	85.79%	92.50%	96.50%	9.89%	6.71%	−4.00%

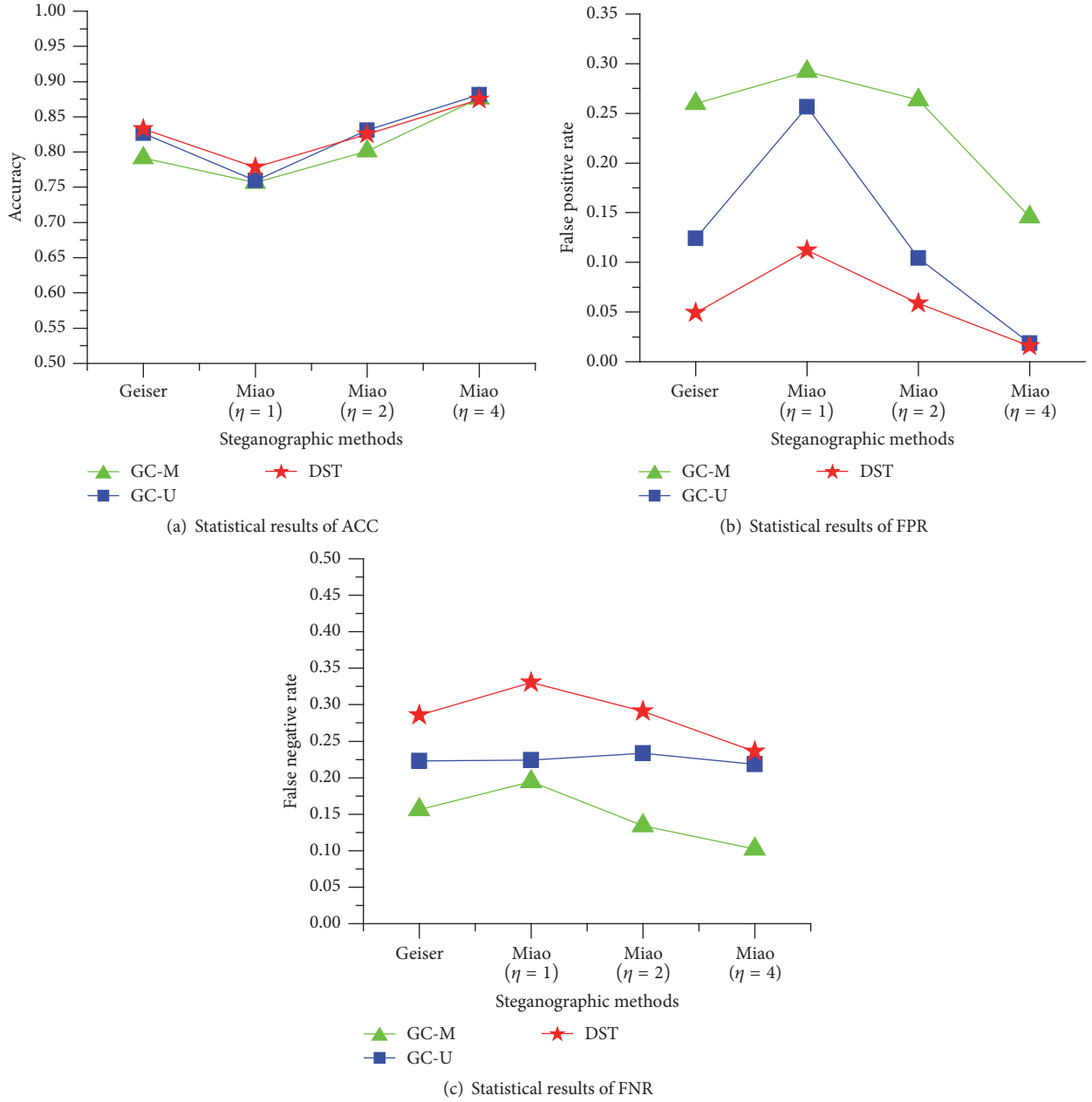
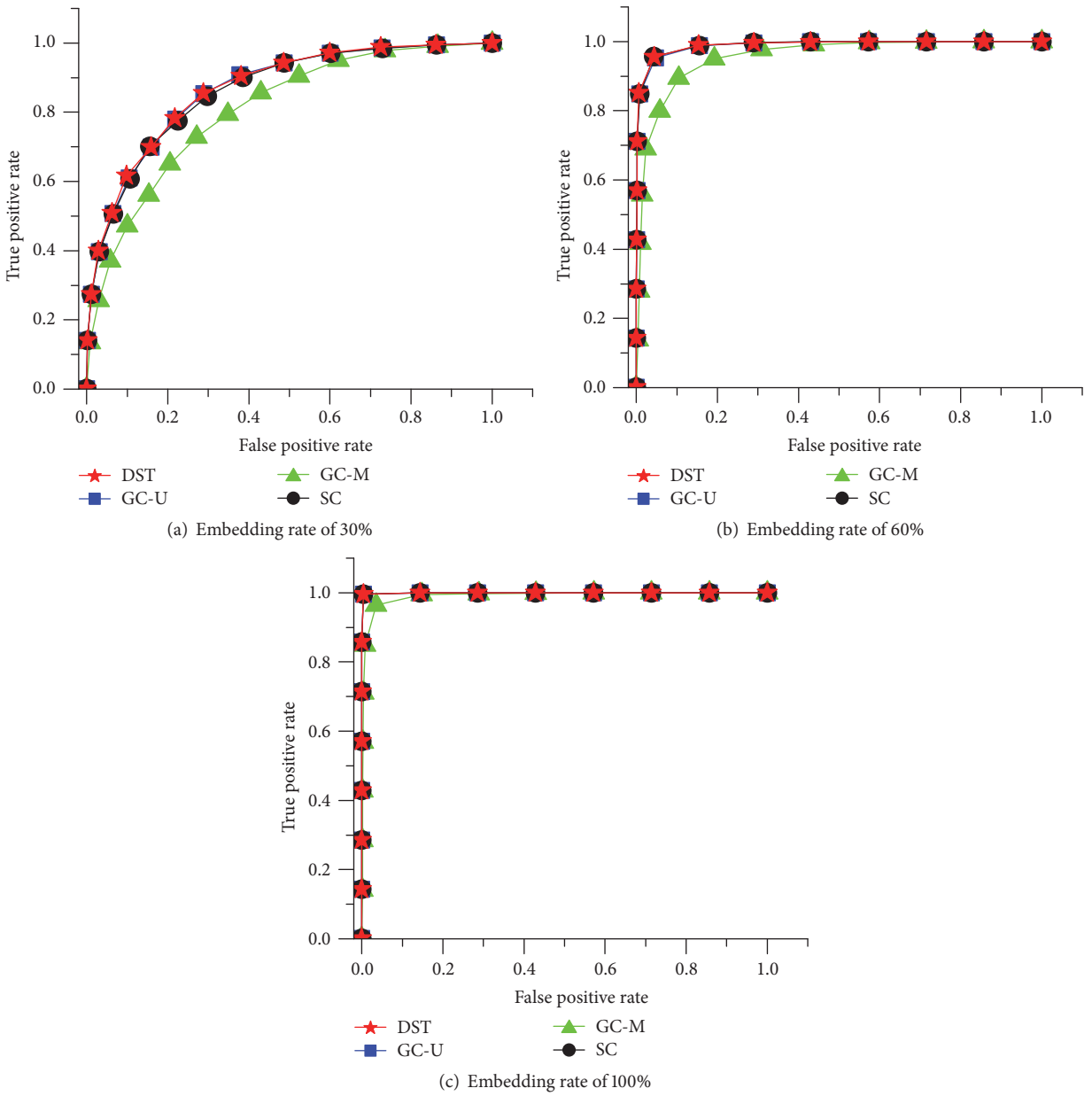


Figure 8: Experimental results for detecting steganographic methods at variable embedding rates.

identical classifier to detect the cover samples. In the cases of the embedding rates smaller than 40%, some detection accuracies of the DST-based scheme are very slightly lower than GC-U or GC-M. The main reason behind this phenomenon is that the detection accuracies of the specific classifiers are relatively low and thereby more likely make the evidence from them highly contradictory. Overall, since the embedding capacities of ten-second speech samples under the embedding rates lower than 40% are very small, the detection performance of all the steganalysis schemes is not so good (particularly, the accuracies are lower than 80% for Geiser's method and Miao's methods at the modes of $\eta = 1$ and 2). In this sense, how to further improve the steganalysis performance for relatively low embedding rates is still a question worthy of study.

In addition, to comprehensively evaluate the performance of the presented schemes for detecting steganographic methods at variable embedding rates, we prepare a mixed detection sample set for each steganographic method, which consists of 1400 speech samples randomly chosen from CSSD and 140 steganographic samples generated by performing the given steganographic method at each embedding rate from 10% to 100%. Figure 8 shows the statistical results of the steganalysis experiments. From these charts, we can learn that all the presented three schemes can achieve relatively good accuracies for detecting the existing steganographic methods.



(a) Embedding rate of 30%

(b) Embedding rate of 60%

(c) Embedding rate of 100%

Figure 9: The ROC curves for detecting Geiser's method.

Specifically, for Geiser's method, the accuracies are more than 79%; for Miao's method ($\eta = 1$), the accuracies more than 75%; for Miao's method ($\eta = 2$), the accuracies more than 80%; and for Miao's method ($\eta = 4$), the accuracies more than 87%. In a word, the presented three schemes are effective for detecting the existing steganographic methods with any given embedding rates.

To further assess the performance of the presented three steganalysis schemes and compare them with the steganalysis based on SCs, we draw receiver-operating-characteristic (ROC) curves for detecting all the state-of-the-art steganographic methods at the typical embedding rates of 30%, 60%, and 100%, as shown in Figures 9, 10, 11, and 12, and calculate their areas under the curves (AUC), as shown in Table 7. The experimental results demonstrate again that the presented three steganalysis schemes are really feasible and effective for detecting the state-of-the-art steganographic methods, while the DST-based scheme can offer better detection performance than GC-U and GC-M overall.

5. Conclusions

Due to its increasing popularity and broad influence in mobile communications, AMR speech is spontaneously considered as an ideal carrier by the steganographic research community, and some relevant steganographic techniques have been successfully developed. However, AMR speech based steganography is a double-edged sword. Illegal usage

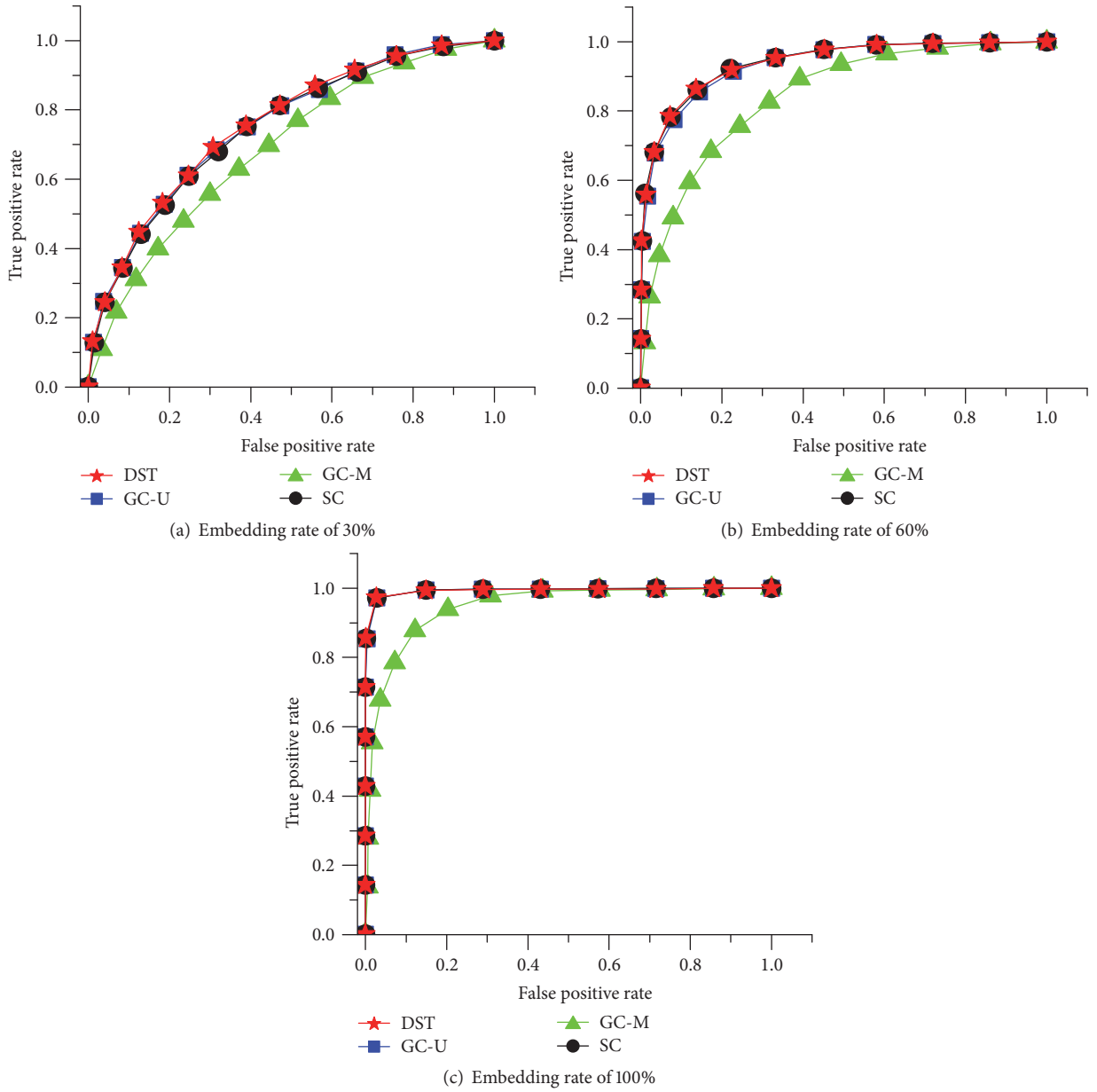
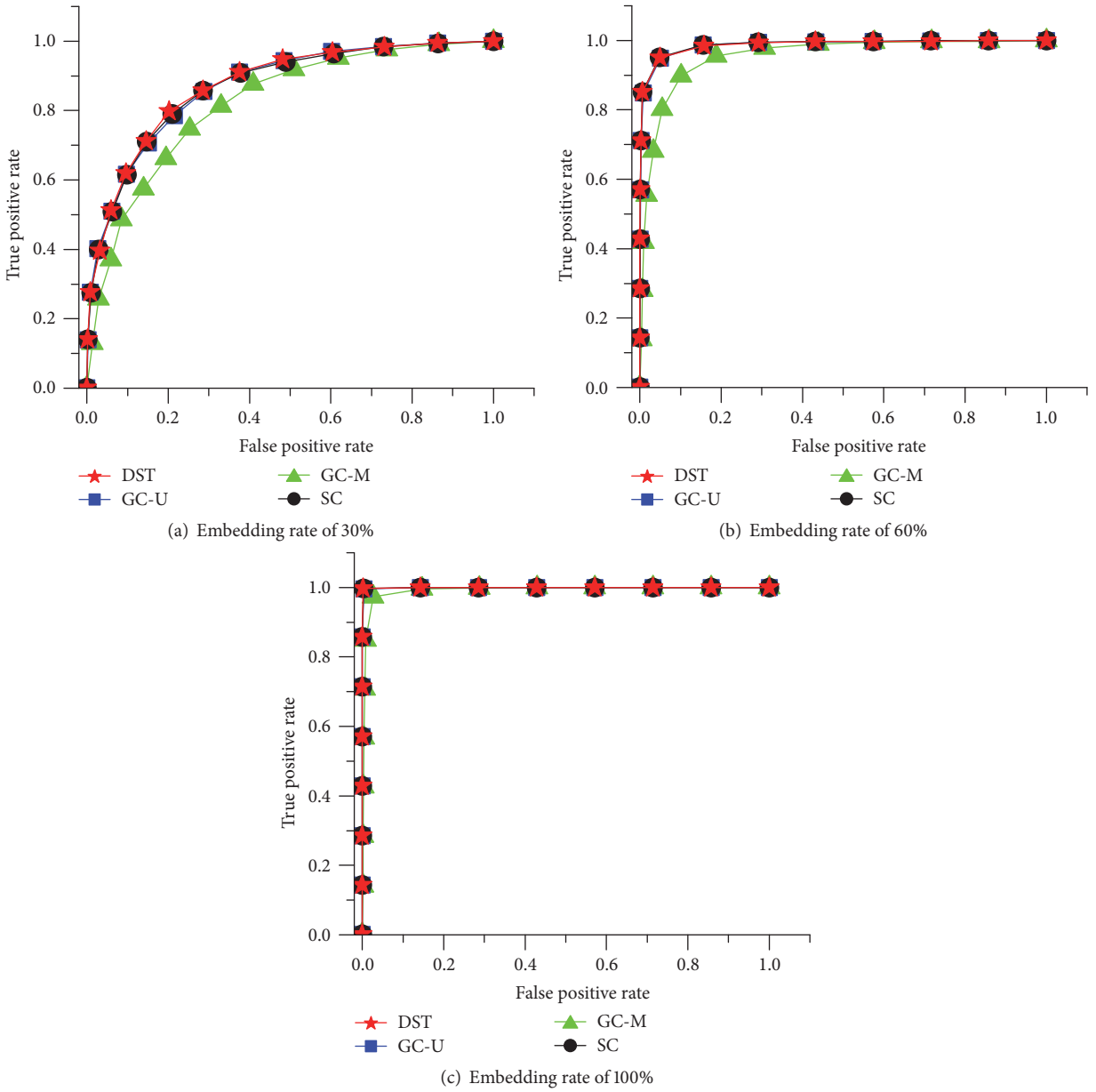


Figure 10: The ROC curves for detecting Miao's method ($\eta = 1$).

Table 5: Statistical results of accuracies for detecting Miao's method ($\eta = 2$).

Embedding rate	GC-M	GC-U	DST	SCs	Difference between DST and other schemes		
					GC-M	GC-U	SCs
10%	57.11%	55.89%	53.64%	57.04%	−3.46%	−2.25%	−3.39%
20%	65.00%	67.32%	63.04%	66.75%	−1.96%	−4.29%	−3.71%
30%	74.68%	77.89%	73.71%	78.04%	−0.96%	−4.18%	−4.32%
40%	79.18%	85.21%	83.18%	85.46%	4.00%	−2.04%	−2.29%
50%	83.57%	91.21%	90.43%	91.54%	6.86%	−0.79%	−1.11%
60%	84.50%	93.93%	94.93%	95.00%	10.43%	1.00%	−0.07%
70%	85.75%	94.39%	96.89%	97.79%	11.14%	2.50%	−0.89%
80%	86.00%	94.71%	97.11%	98.25%	11.11%	2.39%	−1.14%
90%	85.54%	94.68%	97.00%	99.36%	11.46%	2.32%	−2.36%
100%	86.11%	94.57%	97.07%	99.61%	10.96%	2.50%	−2.54%



(a) Embedding rate of 30%

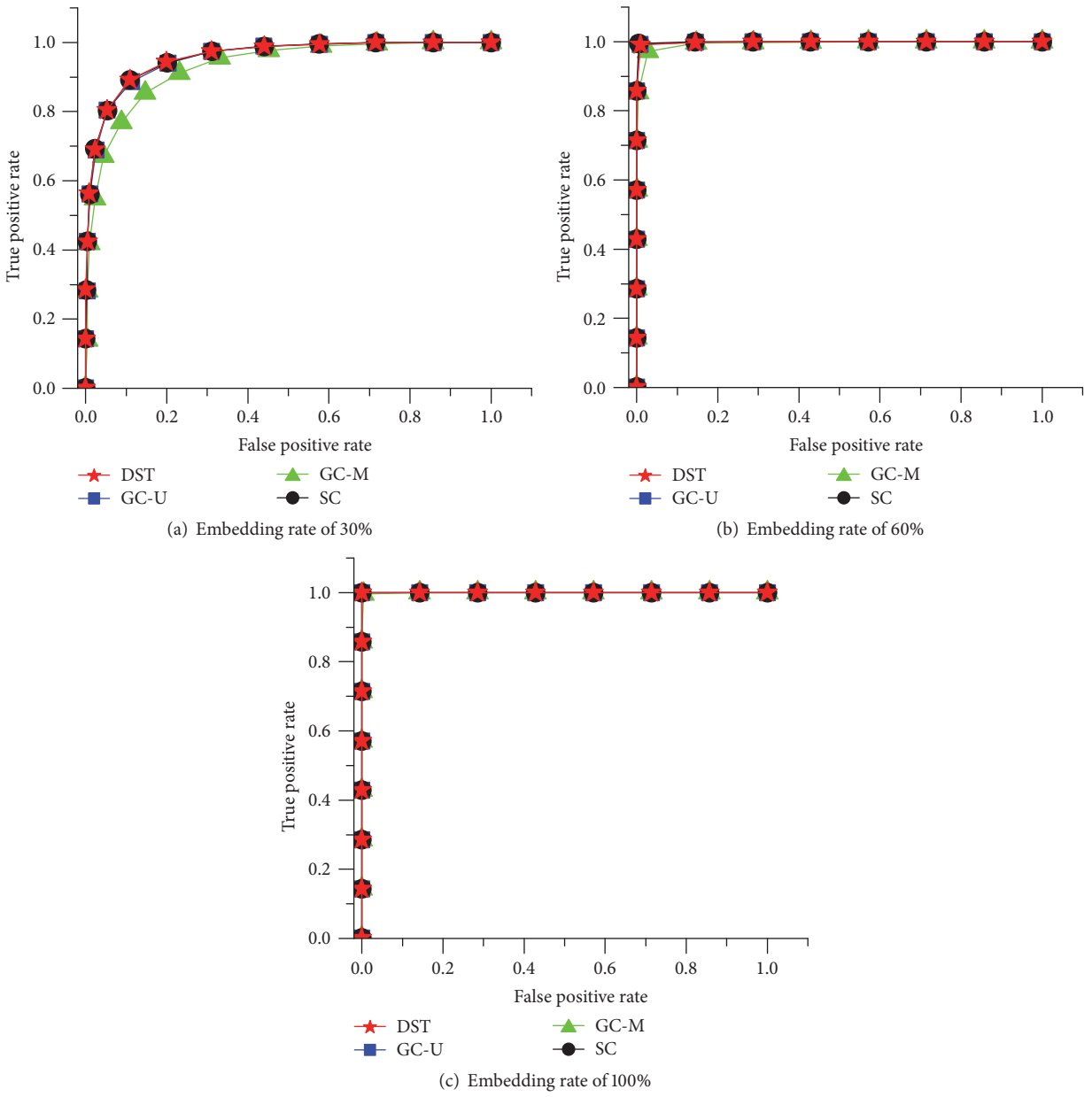
(b) Embedding rate of 60%

(c) Embedding rate of 100%

Figure 11: The ROC curves for detecting Miao's method ($\eta = 2$).

Table 6: Statistical results of accuracies for detecting Miao's method ($\eta = 4$).

Embedding rate	GC-M	GC-U	DST	SCs	Difference between DST and other schemes		
					GC-M	GC-U	SCs
10%	63.07%	55.14%	53.86%	64.64%	−9.21%	−1.29%	−10.79%
20%	75.54%	67.61%	64.86%	79.00%	−10.68%	−2.75%	−14.14%
30%	85.04%	83.18%	80.96%	88.82%	−4.07%	−2.21%	−7.86%
40%	89.29%	93.89%	92.25%	94.93%	2.96%	−1.64%	−2.68%
50%	90.36%	98.04%	97.64%	97.96%	7.29%	−0.39%	−0.32%
60%	91.18%	98.64%	98.79%	99.00%	7.61%	0.14%	−0.21%
70%	91.14%	98.64%	98.93%	99.46%	7.79%	0.29%	−0.54%
80%	91.43%	98.75%	99.04%	99.75%	7.61%	0.29%	−0.71%
90%	91.46%	98.64%	99.00%	99.86%	7.54%	0.36%	−0.86%
100%	91.54%	98.86%	99.14%	99.96%	7.61%	0.29%	−0.82%



(a) Embedding rate of 30%

(b) Embedding rate of 60%

(c) Embedding rate of 100%

Figure 12: The ROC curves for detecting Miao's method ($\eta = 4$).

of this technique would facilitate cybercrime activities and thereby pose a great threat to information security. Thus, its countermeasure, steganalysis of AMR speech, has been also a significant problem worthy of study. Although some fruitful steganalysis studies for AMR speech have been conducted, all the state-of-the-art methods deal with the problem under the assumption that the embedding rate of steganographic samples to be tested is exactly known, which is actually unpractical. Therefore, we are motivated to study steganalysis of AMR speech with unknown embedding rate in this paper. To address this problem, we came up with three different schemes based on SVM. The first two schemes are extended from the existing image steganalysis schemes, which both use global classifiers to detect the steganography but adopt different training sets. Specifically, the first scheme trains the global classifier on a comprehensive speech sample set including original samples and steganographic samples with various embedding rates, while the second one trains the global classifier on a particular speech sample set consisting of original samples and steganographic samples with uniform distributions of embedded data. Besides, we further presented the third hybrid steganalysis scheme based DST, which adopts DST to combine all the evidence from a set of specific classifiers and accordingly provide a synthesized decision for having or not having hidden information. All the three steganalysis schemes are evaluated employing the optimized feature set based on statistical characteristics of pulse pairs and compared with the optimal steganalysis that uses each

Table 7: The areas under the ROC curves for detecting the existing steganographic methods.

Steganographic method	Embedding rate	AUC			
		GC-M	GC-U	DST	SC
Geiser's method	30%	0.807738	0.867176	0.870095	0.864861
	60%	0.956495	0.991041	0.991186	0.990635
	100%	0.991904	0.999748	0.999778	0.999790
Miao's method ($\eta = 1$)	30%	0.688343	0.754481	0.759237	0.752387
	60%	0.843580	0.937296	0.939769	0.941201
	100%	0.948847	0.995716	0.995116	0.995560
Miao's method ($\eta = 2$)	30%	0.823086	0.871535	0.876069	0.871058
	60%	0.957053	0.989038	0.989208	0.989506
	100%	0.993139	0.999892	0.999890	0.999966
Miao's method ($\eta = 4$)	30%	0.932302	0.957791	0.959466	0.958470
	60%	0.995550	0.999243	0.999135	0.999204
	100%	0.999875	1.000000	1.000000	1.000000

specialized classifier to detect the steganography with the corresponding embedding rate. The experimental results demonstrate that all the presented steganalysis schemes are feasible and effective for detecting the existing steganographic methods with unknown embedding rates in AMR speech streams, while the DST-based scheme can provide better performance than the others in most cases.

Conflicts of Interest

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

Acknowledgments

This work was supported in part by Natural Science Foundation of China under Grant nos. U1536115, 61302094, and U1405254, Program of China Scholarships Council under Grant no. 201507540001, Natural Science Foundation of Fujian Province of China under Grant no. 2014J01238, Program for New Century Excellent Talents in Fujian Province University under Grant no. MJK2016-23, Program for Outstanding Youth Scientific and Technological Talents in Fujian Province University under Grant no. MJK2015-54, Promotion Program for Young and Middle-Aged Teacher in Science & Technology Research of Huaqiao University under Grant no. ZQN-PY115, and Program for Science & Technology Innovation Teams and Leading Talents of Huaqiao University under Grant no. 2014KJTD13.

References

[1] N. Provos and P. Honeyman, "Hide and seek: an introduction to steganography," *IEEE Security and Privacy*, vol. 99, no. 3, pp. 32–44, 2003.

[2] E. Zielińska, W. Mazurczyk, and K. Szczypiorski, "Trends in steganography," *Communications of the ACM*, vol. 57, no. 3, pp. 86–95, 2014.

[3] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727–752, 2010.

[4] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1905–1917, 2015.

[5] M. M. Sadek, A. S. Khalifa, and M. G. M. Mostafa, "Video steganography: a comprehensive review," *Multimedia Tools and Applications*, vol. 74, no. 17, pp. 7063–7094, 2015.

[6] M. Ramalingam and N. A. M. Isa, "A data-hiding technique using scene-change detection for video steganography," *Computers & Electrical Engineering*, vol. 54, pp. 423–434, 2016.

[7] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative study of digital audio steganography techniques," *Eurasip Journal on Audio, Speech, and Music Processing*, vol. 2012, no. 1, article 25, 2012.

[8] G. Hua, J. Huang, Y. Q. Shi, J. Goh, and V. L. L. Thing, "Twenty years of digital audio watermarking - A comprehensive review," *Signal Processing*, vol. 128, pp. 222–242, 2016.

[9] E. Satir and H. Isik, "A compression-based text steganography method," *Journal of Systems and Software*, vol. 85, no. 10, pp. 2385–2394, 2012.

[10] C.-Y. Chang and S. Clark, "Practical linguistic steganography using contextual synonym substitution and a novel vertex coding method," *Computational Linguistics*, vol. 40, no. 2, pp. 403–448, 2014.

[11] J. Lubacz, W. Mazurczyk, and K. Szczypiorski, "Principles and overview of network steganography," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 225–229, 2014.

[12] W. Mazurczyk, S. Wendzel, S. Zander, A. Houmansadr, and K. Szczypiorski, *Information Hiding in Communication Networks: Fundamentals, Mechanisms, Applications, and Countermeasures*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2016.

[13] W. Mazurczyk, "VoIP steganography and its detection-a survey," *ACM Computing Surveys*, vol. 46, no. 2, article 20, 2013.

[14] H. Tian, J. Qin, S. Guo et al., "Improved adaptive partial-matching steganography for Voice over IP," *Computer Communications*, vol. 70, pp. 95–108, 2015.

[15] H. Tian, J. Qin, Y. Huang et al., "Optimal matrix embedding for Voice-over-IP steganography," *Signal Processing*, vol. 117, pp. 33–43, 2015.

[16] Y. Jiang, S. Tang, L. Zhang, M. Xiong, and Y. J. Yip, "Covert voice over internet protocol communications with packet loss based on fractal interpolation," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 4, article 54, pp. 1–20, 2016.

[17] A. Janicki, W. Mazurczyk, and K. Szczypiorski, "Steganalysis of transcoding steganography," *Annals of Telecommunications/Annales des Télécommunications*, vol. 69, no. 7-8, pp. 449–460, 2014.

[18] Z. Xia, X. Wang, X. Sun, and B. Wang, "Steganalysis of least significant bit matching using multi-order differences," *Security and Communication Networks*, vol. 7, no. 8, pp. 1283–1291, 2014.

[19] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

[20] Z. Xia, X. Wang, X. Sun, Q. Liu, and N. Xiong, "Steganalysis of LSB matching using differences between nonadjacent pixels," *Multimedia Tools and Applications*, vol. 75, no. 4, pp. 1947–1962, 2016.

[21] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis based on embedding probabilities of pixels," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 734–744, 2016.

[22] J. Yu, F. Li, H. Cheng, and X. Zhang, "Spatial steganalysis using contrast of residuals," *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 989–992, 2016.

[23] T. Denemark, M. Boroumand, and J. Fridrich, "Steganalysis features for content-adaptive JPEG steganography," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1736–1746, 2016.

[24] H. Tian, Y. Wu, Y. Cai et al., "Distributed steganalysis of compressed speech," *Soft Computing*, vol. 21, no. 3, pp. 795–804, 2017.

[25] H. Tian, Y. Wu, C. C. Chang et al., "Steganalysis of analysis-by-synthesis speech exploiting pulse-position distribution characteristics," *Security and Communication Networks*, vol. 9, no. 15, pp. 2934–2944, 2016.

[26] 3GPP/ETSI, "AMR speech codec: general description, version 10.0.0," Technical Report TS 26 071, Sophia Antipolis Cedex, France, April 2011.

[27] 3GPP/ETSI., "Performance characterization of the adaptive multi-rate (AMR) speech codec," Technical Report TR 126 975, Sophia Antipolis Cedex, France, January 2009.

[28] 3GPP/ETSI, "Digital cellular telecommunications system (phase 2+); Universal mobile telecommunications system (UMTS); LTE: mandatory speech codec speech processing functions; Adaptive multi-rate (AMR) speech codec; Transcoding functions (3GPP TS 26.090 version 13.0.0 Release 13)," Technical Report TR 126 090, Sophia Antipolis Cedex, France, January 2016.

[29] B. Geiser and P. Vary, "High rate data hiding in ACELP speech codecs," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 4005–4008, Las Vegas, Nev, USA, April 2008.

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

[31] H. Miao, L. Huang, Y. Shen, X. Lu, and Z. Chen, "Steganalysis of compressed speech based on Markov and entropy," in *Proceedings of the 12th International Workshop on Digital-Forensics and Watermarking (IWDW)*, pp. 63–76, Auckland, New Zealand, Oct. 2013.

[32] Y. Ren, T. Cai, M. Tang, and L. Wang, "AMR steganalysis based on the probability of same pulse position," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1801–1811, 2015.

[33] H. Tian, Y. Wu, Y. Huang et al., "Steganalysis of adaptive multi-Rate speech using statistical characteristics of pulse pairs," *Signal Processing*, vol. 134, pp. 9–22, 2017.

[34] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, pp. 771–780, 1999.

[35] X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Information Sciences*, vol. 295, pp. 395–406, 2015.

[36] D. D. Le and S. Satoh, "Feature selection by adaboost for SVM-based face detection," *Information Technology Letters*, vol. 3, pp. 183–186, 2004.

[37] Y.-J. Yeh and C.-T. Hsu, "Online selection of tracking features using AdaBoost," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 442–446, 2009.

[38] L. Guo, P.-S. Ge, M.-H. Zhang, L.-H. Li, and Y.-B. Zhao, "Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4274–4286, 2012.

[39] A. D. Ker, P. Bas, R. Böhme et al., "Moving steganography and steganalysis from the laboratory into the real world," in *Proceedings of the 1st ACM Workshop on Information Hiding and Multimedia Security, IH and MMSec 2013*, pp. 45–58, France, June 2013.

[40] T. Pevny, "Detecting messages of unknown length," in *Proceedings of the Media Watermarking, Security, and Forensics III*, vol. 7880, pp. 1–12, San Francisco Airport, California, USA, 2011.

[41] L. Marvel, B. Henz, and C. Boncelet, "A performance study of ±1 steganalysis employing a realistic operating scenario," in *Proceedings of the 2007 IEEE Military Communications Conference*, pp. 1–7, USA, October 2007.

[42] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[43] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.

[44] R. R. Murphy, "Dempster-Shafer theory for sensor fusion in autonomous mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 2, pp. 197–206, 1998.

[45] N. R. Pal and S. Ghosh, "Some classification algorithms integrating Dempster-Shafer theory of evidence with the rank nearest neighbor rules," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, vol. 31, no. 1, pp. 59–66, 2001.

[46] T. M. Chen and V. Venkataramanan, "Dempster-Shafer theory for intrusion detection in ad hoc networks," *IEEE Internet Computing*, vol. 9, no. 6, pp. 35–41, 2005.

[47] J. S. Perkell and D. H. Klatt, *Invariance and Variability in Speech Processes*, Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, 1986.

[48] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[49] 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.

[50] B. Gu, V. S. Sheng, Z. Wang, D. Ho, S. Osman, and S. Li, "Incremental learning for ν-support vector regression," *Neural Networks*, vol. 67, pp. 140–150, 2015.

[51] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1403–1416, 2015.

[52] B. Gu, V. S. Sheng, and S. Li, "Bi-parameter space partition for cost-sensitive SVM," in *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pp. 3532–3539, Buenos Aires, Argentina, July 2015.

[53] B. Gu and V. S. Sheng, "A robust regularization path algorithm for ν-support vector classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1241–1248, 2017.