

# A Novel Audio Steganalysis Based on High-Order Statistics of a Distortion Measure with Hausdorff Distance

Yali Liu<sup>1</sup>, Ken Chiang<sup>2</sup>, Cherita Corbett<sup>2</sup>, Rennie Archibald<sup>3</sup>, Biswanath Mukherjee<sup>3</sup>, and Dipak Ghosal<sup>3</sup>

<sup>1</sup> Electrical & Computer Engineering, University of California, Davis,  
Davis, CA 95616 USA  
yliu@ece.ucdavis.edu

<sup>2</sup> Sandia\* National Laboratories, Livermore, CA 94551, USA  
{clcorbe, kchiang}@sandia.gov

<sup>3</sup> Department of Computer Science, University of California, Davis,  
Davis, CA 95616 USA  
{rvarchibald, mukherjee, ghosal}@cs.ucdavis.edu

**Abstract.** Steganography can be used to hide information in audio media both for the purposes of digital watermarking and establishing covert communication channels. Digital audio provides a suitable cover for high-throughput steganography as a result of its transient and unpredictable characteristics. Distortion measure plays an important role in audio steganalysis - the analysis and classification method of determining if an audio medium is carrying hidden information. In this paper, we propose a novel distortion metric based on Hausdorff distance. Given an audio object  $x$  which could potentially be a stego-audio object, we consider its de-noised version  $x'$  as an estimate of the cover-object. We then use Hausdorff distance to measure the distortion from  $x$  to  $x'$ . The distortion measurement is obtained at various wavelet decomposition levels from which we derive high-order statistics as features for a classifier to determine the presence of hidden information in an audio signal. Extensive experimental results for the Least Significant Bit (LSB) substitution based steganography tool show that the proposed algorithm has a strong discriminatory ability and the performance is significantly superior to existing methods. The proposed approach can be easily applied to other steganography tools and algorithms.

## 1 Introduction

Steganography is the art and science of hiding a secret message within an innocuous and open carrier medium, such as digital audio, image, and video. To achieve covert communications without raising suspicion, media containing some

---

\* Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-04AL85000.

hidden information (stego-objects) should be indistinguishable from media without any hidden information (cover-objects). The rapid proliferation of Voice over Internet Protocol (VoIP) and other Peer-to-Peer (P2P) audio services provide vast opportunities for covert communications [1]. By slightly altering the binary sequence of the audio samples with existing steganography tools, covert communication channels may be relatively easy to establish. Moreover, the inherent redundancy in the audio signal and its transient and unpredictable characteristics imply a high hidden capacity. This is further aided by the fact that the human ear is insensitive to small distortions in the audio signal. All these make audio a good candidate for use as a “cover” for covert communications to hide secret messages.

The countermeasure to steganography is steganalysis. In particular, steganalysis seeks to identify suspected information streams; determine whether or not they have hidden messages encoded into them; and, if possible, recover the hidden information. Steganalysis can also serve as an effective way to judge the security performance of steganography techniques, leading to new steganography methods followed by new and improved steganalysis techniques for their detection. As covert communications greatly increase the possibility of unknown malicious activities from a security standpoint, there is significant demand in steganalysis technique to detect hidden information in open digital media content. In this paper, similar to most of the steganalysis work, we focus on the detection of the presence of hidden content, rather than message recovery or other functions.

In recent years, there has been significant effort in the steganalysis of digital images [2][3][4]. Typically, natural images tend to be contiguous and smooth, which leads to high spatial coherence among adjacent pixels. Since the hidden message is usually independent of the cover-image, embedding the hidden message into the cover-image may decrease or even destroy the inherent natural correlation. As a result, most of the image steganalysis algorithms attempt to determine some particular statistical features that can capture this change. Although most image steganalysis algorithms claim that they can be easily extended to other types of media files (e.g., audio), many of the models capture statistical regularities inherent to the spatial composition of images, which are not present in other types of media files such as audio [5]. As a result, deeper research must be conducted on the feature extraction when identifying stego-audio files.

Compared to image steganalysis, audio steganalysis is relatively unexplored. Johnson et al. proposed a universal steganalysis algorithm that exploits the statistical regularities of recorded speech [5]. In [6], audio quality metrics were adopted to capture the distortion introduced by the hidden information. Later in [7], Avcibas proposed an audio steganalysis algorithm using content-independent distortion measures. However, all these audio distortion measurements are seeking ways in which the existing quality metrics can reflect the sensitivity to the presence of hidden messages.

In this paper, we propose an audio steganalysis scheme that measures audio distortion using Hausdorff Distance [8]. Among various distance measures, Hausdorff distance was chosen because of its successful applications in matching given templates in arbitrary target images [9]. Its strong discriminatory power makes it very helpful in the distortion measurement process. High order statistics derived from this distortion measure can then be used to generate features for a classifier. Unlike previous work in audio steganalysis that used the traditional audio quality metrics, such as signal-to-noise ratio (SNR), Perceptual Audio Quality Measure (PAQM) [6], and other such metrics, the proposed distortion measure is designed specifically to detect modifications to pure audio content as follows.

Given an audio object  $x$  which could potentially be a stego-audio object, we consider its de-noised version  $x'$  as an estimate of the cover-object. After appropriate segmentation, we apply wavelet decomposition to both  $x$  and  $x'$  to generate wavelet coefficients [10] at different levels of resolution. Next, Hausdorff distances are used to test the similarities between the wavelet coefficients of the audio signals and their de-noised versions. The statistical moments of these Hausdorff distances are used as the features to train a classifier on the difference between known cover-audio objects and stego-audio objects with different hidden content loadings. Simulations with numerous audio sequences show that our algorithm provides significantly higher classification rates than existing schemes that use standard audio quality metrics or statistical moments without considering audio quality. Moreover, as the proposed scheme makes no assumptions about the embedding technique used, it should be easily applicable to other steganography tools and algorithms.

The paper is organized as follows. Section 2 provides the related work. In Section 3, we briefly introduce the general idea of steganalysis based on audio distortion before presenting the novel steganalysis distortion metric based on Hausdorff distance and the corresponding high-order statistics used as a feature vector. In Section 4, we implement our technique and present our experimental results and performance comparisons. Section 5 concludes the paper.

## 2 Related Work

In recent years, there has been significant research effort in steganalysis with primary focus on digital images. Since there are similarities between images and audios, in this section, we review some image steganalysis algorithms which may be helpful for the audio steganalysis. It should be noted that many steganalysis techniques are specific to some particular data hiding methods [2][11][12][13][14]. However, since the data-embedding method is typically unknown prior to detection, we focus on the design of a unified steganalysis algorithm to detect the presence of steganography independent of the steganography algorithms used. Moreover, we focus on passive detection as opposed to active warden steganalysis [6] which aim to detect and modify the hidden content.

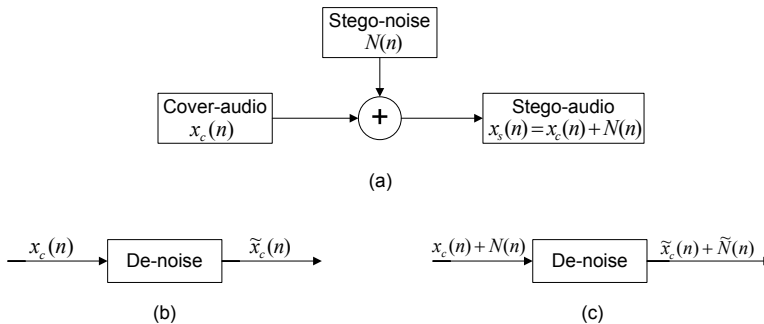
## 2.1 High-Order Statistics and Steganalysis

A number of prior studies have shown that high-order statistics are very effective in differentiating stego-images from cover-images. In [15], Farid proposed a general steganalysis algorithm based on image high-order statistics. In this method, a statistical model based on the first (mean) and higher-order (variance, skewness, and kurtosis) magnitude statistics, extracted from wavelet decomposition, is used for image steganography detection. In [16], a steganalysis method based on the moments of the histogram characteristic function was proposed. It has been proved that, after a message is embedded into an image, the mass center (the first moment) of histogram characteristic function will decrease. In [10], Holotyak et al. used higher-order moments of the probability density function (PDF) of the estimated stego-object in the finest wavelet level to construct the feature vectors. Due to the limited number of features used in the steganalysis technique proposed in [16], Shi et al. proposed the use of statistical moments of the characteristic functions of the wavelet sub-bands [17]. Because the  $n^{th}$  statistical moment of a wavelet characteristic function is related to the  $n^{th}$  derivative of the corresponding wavelet histogram, the constructed 39-dimensional feature vector has proved to be sensitive to embedded data.

Usually, the steganalysis algorithms based on the high-order statistics can achieve satisfactory performance on image files, regardless of the underlying embedding algorithm. However, these statistical models may not be appropriate for audio files because these models capture statistical regularities inherent to the spatial composition of images which is not present in audio [5].

## 2.2 Distortion Measures and Steganalysis

The concept of using distortion measures to classify cover-objects and stego-objects was introduced by Avcibas et al. in 2003 [18]. Since the presence of steganography communication in a signal can be modeled as additive noise in the time or frequency domains [16], the de-noised versions of the image signals can be used to represent close approximations of the cover-images. It has been shown that the distortion (measured by the distance in the feature space) of the cover-image to its de-noised version is different than the distortion between a stego-image and its de-noised version. Specifically, some image quality metrics, e.g., Minkowsky [18], correlation, and human visual system (HVS) based measures [19][20], are selected as the feature set to distinguish between cover-images and stego-images. This concept was extended to audio steganalysis in [6]. Similar to [18], the potential of distortion audio metrics is used to build a steganalyzer to discriminate between cover-audio objects and stego-audio objects. Particularly, the traditional audio quality metrics, such as SNR, PAQM, and other such metrics are tested for their sensitivity to the presence of steganographic content. In [7], Avcibas proposed an audio steganalysis algorithm using content-independent distortion measures. By removing content dependency during the distortion measurement, the paper shows that the discriminatory power is enhanced.



**Fig. 1.** Schematic descriptions of (a) additive noise steganography model, (b) de-noising a cover-audio object, and (c) de-noising a stego-audio object.

Note that all these algorithms attempt to find good features from the standard quality metrics which are designed to evaluate the perceptual and objective quality performance of images or audio. As the primary motivation for developing these quality metrics was for purposes other than steganalysis, the capability of distinguishing changes in quality due to embedding content using steganography may be limited. Consequently, we argue that it is better to define a distortion metric that is designed specifically to detect modifications to audio content. Furthermore, since the high-order moments have been helpful in image steganalysis, we believe they can also contribute in audio steganalysis if used properly.

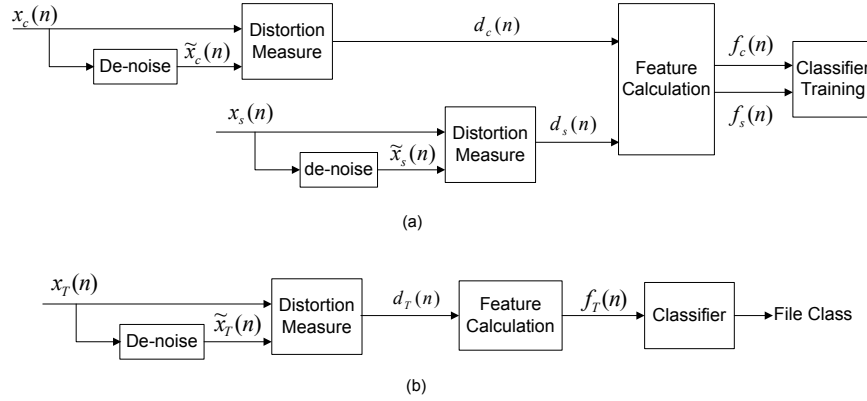
### 3 Methodology

In this section, we first review steganography message embedding techniques and set up a steganalyser based on audio distortion. Then based on the approximate additive noise model, an audio steganalysis using high-order statistics of a distortion measure with Hausdorff distance is proposed.

#### 3.1 Steganalysis Based on Audio Distortion

Due to the natural noise in the media transmission process, e.g., quantization, sensor, and channel, a number of steganography hiding schemes try to disguise the hidden message as a naturally present noise. As such, a generalized additive noise scheme has been developed in [20] that is capable of embedding data with any given distribution. Moreover, the work in [15] shows that most of the steganography algorithms, e.g., Least Significant Bit (LSB) steganography, spread spectrum image steganography, or even more robust and stealthy steganography schemes such as Discrete Cosine Transform (DCT) steganography, can be approximated as an additive noise scheme.

The same additive noise model can also be applied to audio files. The steganography message embedding process is shown in Figure 1. Let  $x_c(n)$  denote a cover-audio object and  $x_s(n)$  be its stego-version. Let  $N(n)$  be an independent and



**Fig. 2.** Schematic description of (a) training and (b) testing for the steganalysis.

identically distributed (i.i.d) Gaussian noise; then the stego-audio object can be expressed as  $x_s(n) = x_c(n) + N(n)$  with the additive noise model. A good feature should enlarge the distance between  $x_c(n)$  and  $x_s(n)$ . However, it is important to note that, in a real communication environment, a reference audio file needs to be used since we cannot get specific information about the original cover-audio object. The de-noised version of an audio file has already been shown to be a good estimation of the cover signal [6]. Note that the de-noised version of stego-audio is still the de-noised cover-audio with some i.i.d. Gaussian noise.

The training and testing procedures for the steganalysis are shown in Figure 2. Let  $\tilde{x}_c(n)$  and  $\tilde{x}_s(n)$  be the de-noised versions of a cover-audio object and a stego-audio object, respectively. The defined distortion metric, in fact, is simply trained to differentiate between the distances, denoted as  $d_c(n)$  and  $d_s(n)$ , of the cover-audio object and stego-audio object to their de-noised versions. Instead of using  $d_c(n)$  and  $d_s(n)$  as audio features, further feature calculation procedures are performed before going to the classifier training process. The test audio file  $x_T(n)$  will go through the same procedures of distortion measure and feature selection until the resulting feature vector  $f_T(n)$  is achieved, and then used to judge the test file type with the training model.

In addition to feature calculation, the classifier plays an important role in the steganalysis process. It affects the classification performance in terms of success classification rate as well as the computational complexity. In our work, we use the freely available package Library for Support Vector Machines (LIBSVM) [21], which is powerful software for data classification and is widely used in steganalysis.

### 3.2 Feature Calculation

**Wavelet de-noising.** The goal of the de-noising process is to recover the characteristics of the original cover-audio object while also removing as much noise

as possible. Considering the non-stationary characteristics of audio signals [5], a smoothing filter may not be very suitable for estimating the cover-object. Among many other techniques, Wiener filtering is a powerful tool for additive noise reduction. In its basic form, Wiener theory assumes that signals are stationary processes. However, this assumption is not realistic for audio signals, whose characteristics change in time and therefore are considered non-stationary signals. As a result, we consider adopting the wavelet de-noising technique. Using the thresholding technique [22], wavelet approximation allows an adaptive representation of signal discontinuities. Wavelets also provide unconditional basis for a variety of function spaces and thus provide better approximation power than Fourier basis to help recover the characteristics of the cover-audio signal more effectively.

**Distortion Measure.** Once we get the de-noised version of an audio signal, a distance measurement will be applied to measure the distortion or degradation of the original audio signal. Such a measurement should respond to the presence of a hidden message in an accurate, consistent, and monotonic (with respect to the size of the hidden message) way. It should be noted that, instead of gathering information directly from audio files, signatures of the audio files are generated based on their wavelet coefficients at different levels of resolution and will be used to test the distance of the audio file and its de-noised version. The wavelet transform is chosen for its well-known capability of multi-resolution decomposition [23], which can help to enlarge the influence of the additive noise present as a result of embedding. Since the hidden information may only modify a small portion of the cover-objects, the distortion is calculated at their pre-defined small segments separately. At this point, the time-frequency localization [23] characteristic of the wavelet transform may also provide some information about the discontinuities that occur.

Since the distortion metric should be sensitive to the presence of a hidden message and its reaction should be proportional to the embedding strength, Hausdorff distance [8] is used as a dissimilarity measure. Among dissimilarity measures over binary images, the Hausdorff distance has often been used in the content-based retrieval domain and is known to have successful applications in object matching [18]. On the other hand, Hausdorff distance is very sensitive to noise [24][9]. A small distortion can result in a significant distance between two objects. However, in steganalysis, the main issue under consideration is not the content of an audio file but the minor distortions introduced during the data-hiding process. As a result, this characteristic of Hausdorff distance makes it very helpful in the steganalysis.

The Hausdorff distance is basically a max-min distance. Suppose the length of each segment of the audio file is  $M$ . After de-noising and wavelet decomposition at level  $p$ , for  $m^{th}$  segment, the wavelet coefficients of the audio file and its de-noised version are  $C_m^p = \{c_m^1, c_m^2, \dots, c_m^q\}$  and  $\tilde{C}_m^p = \{\tilde{c}_m^1, \tilde{c}_m^2, \dots, \tilde{c}_m^q\}$ , where  $q = M/2^p$ . Then, its distortion measure with Hausdorff distance is:

$$H_m^p = \max\{h(C_m^p, \tilde{C}_m^p), h(\tilde{C}_m^p, C_m^p)\} \quad (1)$$

where

$$h(C_m^p, \tilde{C}_m^p) = \max_{i=1,2,\dots,q} \{ \min_{j=1,2,\dots,q} \|c_m^i - \tilde{c}_m^j\| \} \quad (2)$$

is the directed Hausdorff distance from  $C_m^p$  to  $\tilde{C}_m^p$  and  $\|c_m^i - \tilde{c}_m^j\|$  is some underlying norm on the point of  $c_m^i$  and  $\tilde{c}_m^j$ . Here, we use the absolute difference.

**Feature Calculation.** As in [6], to get good local distortion estimation, the segment size  $M$  that the audio file is split into should not be very large (in our experiment, we set  $M$  as 1024 audio samples). As a result, the number of Hausdorff distances after the distortion measurement is still very large. It is unrealistic to use these distances directly for steganalysis. A feasible approach is to extract a certain amount of data (features) from these distances and use them to represent the distortion measurement for steganalysis. Because the task of the segmentation is to test the distortion regularity for the audio files, high-order statistics based on the moments will be used as the final feature.

Suppose the entire audio length is  $L$  samples, then the total number of segments is  $\lceil \frac{L}{M} \rceil$ . For wavelet decomposition level  $p$ , where  $p = 0, 1, \dots, P$ , the overall distortion measured using Hausdorff distance is:

$$D^p = \langle H_1^p, H_2^p, \dots, H_{\lceil \frac{L}{M} \rceil}^p \rangle \quad (3)$$

and the feature vector  $V^p = \langle v_1^p, v_2^p, \dots, v_K^p \rangle$  can be extracted according to the following equation.

$$v_i^p = \frac{\sum_{j=1}^n (f_j^p)^i \cdot d_j^p}{\sum_{j=1}^n d_j^p}, i = 1, 2, \dots, K \quad (4)$$

where  $d_j^p$  is the amplitude of  $j^{th}$  frequency component  $f_j^p$  to the distortion distances  $D^p$  and  $K$  is the total number of moments.

In this way, for each wavelet decomposition level, we have  $K$  features. For a total  $P$ -level wavelet decomposition plus the level 0 which is the signal itself, we have  $(P + 1) \times K$  features which form a high-dimensional feature vector:

$$V = \langle V^0, V^1, \dots, V^P \rangle \quad (5)$$

for steganalysis.

### 3.3 Algorithm Summary

In summary, the proposed feature calculation algorithm proceeds along the following steps:

**Step 1.** For a given audio file  $x(n)$ , apply wavelet de-noising to get its de-noised version  $\tilde{x}(n)$ .

**Step 2.** Partition the signal  $x(n)$  and  $\tilde{x}(n)$  with pre-defined segment length  $M$ . Calculate the wavelet coefficients  $C_m^p$  and  $\tilde{C}_m^p$  at different levels  $p$  for segment  $m$ .

**Step 3.** For each wavelet decomposition level  $p$ , calculate the distortion measure  $H_m^p$  with Hausdorff distance in Equation (1) for all the segments.

**Step 4.** Set up the feature vector  $V^p$  by calculating the moments of  $D^p$  using Equation (3) for each wavelet decomposition level  $p$ .

**Step 5.** Set up the high-dimensional feature  $V$  using Equation (5).

## 4 Experimental Results

To evaluate the performance of the proposed steganalysis algorithm, we randomly picked 994 wav files from the wav surfer database [25]. All these wav files are parts of movies or television programs and have different audio characteristics. These audio files (compressed to MP3 formats) are transformed into standard PCM wav format using Nero Wave Edit before processing. The sampling rate is 44.1 kHz with 16 bits per sample. The audio file lengths vary from one second to 298 seconds. As for the steganography tool, we have used Steghide [26] due to its robustness against a number of different steganalysis tools. For the results presented in this paper, we have set  $P$  (the number of wavelet decomposition levels) to 4 and  $K$  (the number of high order moments) to 5. This implies that a 25-D feature vector is generated for each audio file.

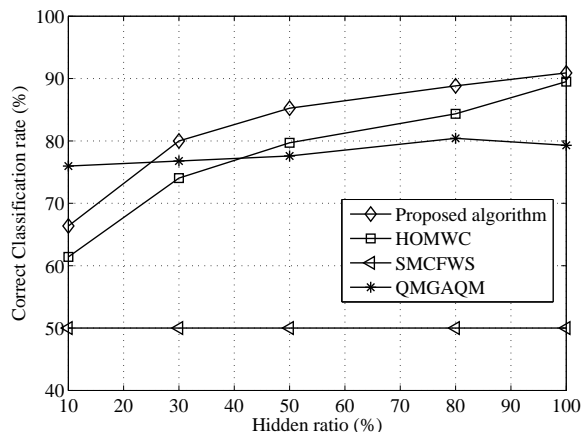
### 4.1 Performance Comparison with Other Audio Steganalysis Algorithms

In this section, we compare the performance of the proposed algorithm with three known algorithms referred to in Section 2. Each experiment randomly selects 895 of the 994 original audio as cover-audio objects. For each audio object, we create a corresponding stego-audio object with a specific amount of hidden content (measured with hidden ratio). As a result, 895 stego-audio files and their original 895 cover-audio files are used for training the classifier. Here, the hidden ratio is the percentage of the size of the hidden message to the hidden capacity (the maximum size of the information that can be hidden) which is determined by Steghide. From the remaining 99 audio files, we obtain 99 pairs, each pair consisting of the original audio and the corresponding stego-audio with a specific hidden ratio. These 99 pairs of audio are used for testing. Note that in this section, we only focus on the feature effectiveness and assume that the hidden ratio information is known before testing, i.e., the hidden ratios are the same in the training and testing processes. The performance metric used is the correct classification rate<sup>4</sup> with the average computed from 10 independent experiments.

Of the three different reference algorithms considered for comparison, the first two were selected to test the importance of high-order statistics in audio steganalysis. Particularly, one algorithm (HOMWC) [15] is directly based on the

---

<sup>4</sup> The correct classification rate is the average detection rate to all the original audio objects and stego-audio objects.



**Fig. 3.** Comparison of correct classification rate with other audio steganalysis algorithms.

high-order moments of the wavelet coefficients of the audio signal. The second algorithm (SMCFWS) [17] is based on statistical moments of the characteristic functions of wavelet sub-bands. In order to make a fair comparison of the performance of the algorithms, we have considered the first 5 moments for the first 4 wavelet decomposition levels as in our algorithm as well. The third algorithm (QMGAQM) [6] is based on the quality measurement with general audio quality metrics. Similarly to the study in [6], a 5-D feature vector based on SNR and PAQM and other such metrics is used to train the classifier.

Figure 3 plots the correct classification rate as a function of the hidden ratio. It is clearly observed that the correct classification rate achieved by our proposed algorithm is more than 90% with 100% hidden ratio and 85% for 50% hidden ratio. Even with only 10% hidden ratio, our approach can still achieve more than 66% successful detection. More importantly, the proposed algorithm shows strong monotonic characteristics with different hidden ratios. On the other hand, it is observed that the SMCFS algorithm does not perform well. Although the algorithm using moments of wavelet coefficients (HOMWC) is fairly good, our algorithm can still get more than 10% improvement in the correct classification rate. This does not come as a surprise since these algorithms work very well in the stego-image identification due to their ability to capture the statistical regularities inherent in the spatial composition of images which are not present in audio. Note that the performance of the algorithm with audio quality metrics (QMGAQM) is fairly good at low hidden ratio. However, the classification rate does not show strong monotonic characteristics with respect to different hidden ratios. This confirms our aforementioned doubt that the standard audio quality metrics may or may not reflect modifications to pure audio content.

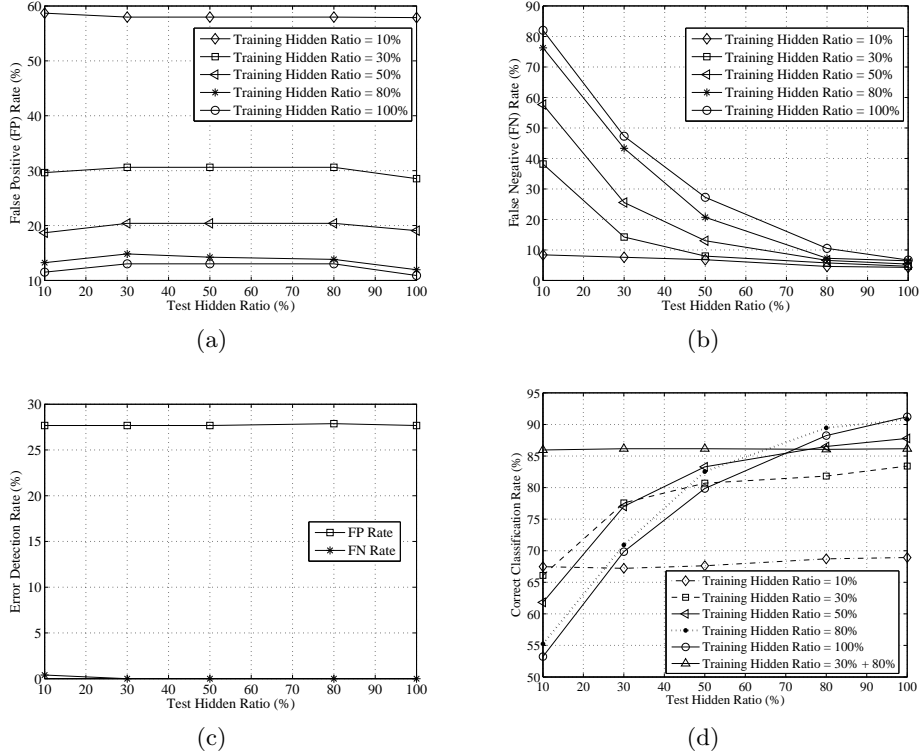
## 4.2 Performance with Respect to Different Hidden Ratios

In the previous section, we compared the feature effectiveness with different audio steganalysis algorithms using the same hidden ratio at the training process and testing. However, in a real system, the hidden ratio information will be unknown before the test. In this section, we evaluate the performance of the proposed algorithm by considering different hidden ratios during the training process. Similar to the previous section, we randomly select 895 original audio files as cover-audio objects. Their corresponding stego-audio files are generated based on five different hidden ratios: 10%, 30%, 50%, 80%, and 100%. For each hidden ratio, we created 895 stego-audio objects and used them in conjunction with the original 895 cover-audio objects for training the classifier. This leaves 99 pairs of audio files (original audio objects and their stego-versions with the same hidden ratio) to be used for testing. The performance metrics used are the false positive (FP) and false negative (FN) rates reported as an average of ten independent experiments.

Figure 4(a) and 4(b) plot the detection performance at different hidden ratios during the training and testing processes. They show that both FP and FN of our algorithm are influenced by the hidden ratio in the training process. Specifically, the higher the hidden ratios used in the training process, the lower the FP. This is consistent with the fact that distortion is higher with higher hidden ratios. Thus, at high hidden ratios, the test cover-audio object is less likely to be misjudged as a stego-audio object. However, the large distortion introduced by the high hidden ratios in the training process will make the system more likely to misclassify stego-audio objects with lower hidden ratio. Consequently, there is a trade-off between FP and FN. Concerning this trade-off, we find it is reasonable to train our SVM models with audio files embedded with multiple hidden ratios. Therefore, during the training process, for each cover-audio object, multiple versions of the stego-audio objects with the selected set of the hidden ratios are used. Considering the unknown properties of the test audio and computation cost for the training process, only limited combinations of the hidden ratios may be selected. In our study, we find 30% and 80% hidden ratios can help to train the system with a good representation of the cover-audio objects and stego-audio objects simultaneously.

Figure 4(c) plots the simulation results for the test audio objects containing different hidden ratios with a training set that contains stego-audio objects with both 30% and 80% hidden ratios (denoted as 30% + 80%). The low FP rate and FN rate indicate that, in most cases, the system can distinguish the cover-audio objects and stego-audio objects successfully. More importantly, these results show that multiple training hidden ratios greatly improve the robustness of our algorithm. Our improved robustness can be observed as both FP rate and FN rate are almost unchanged with different test hidden ratios, which is very helpful since we usually do not know the hidden ratio of a stego-object in advance.

Figure 4(d) plots the correct classification rates by the influence of different hidden ratios in the training process. Note that the smaller the difference between



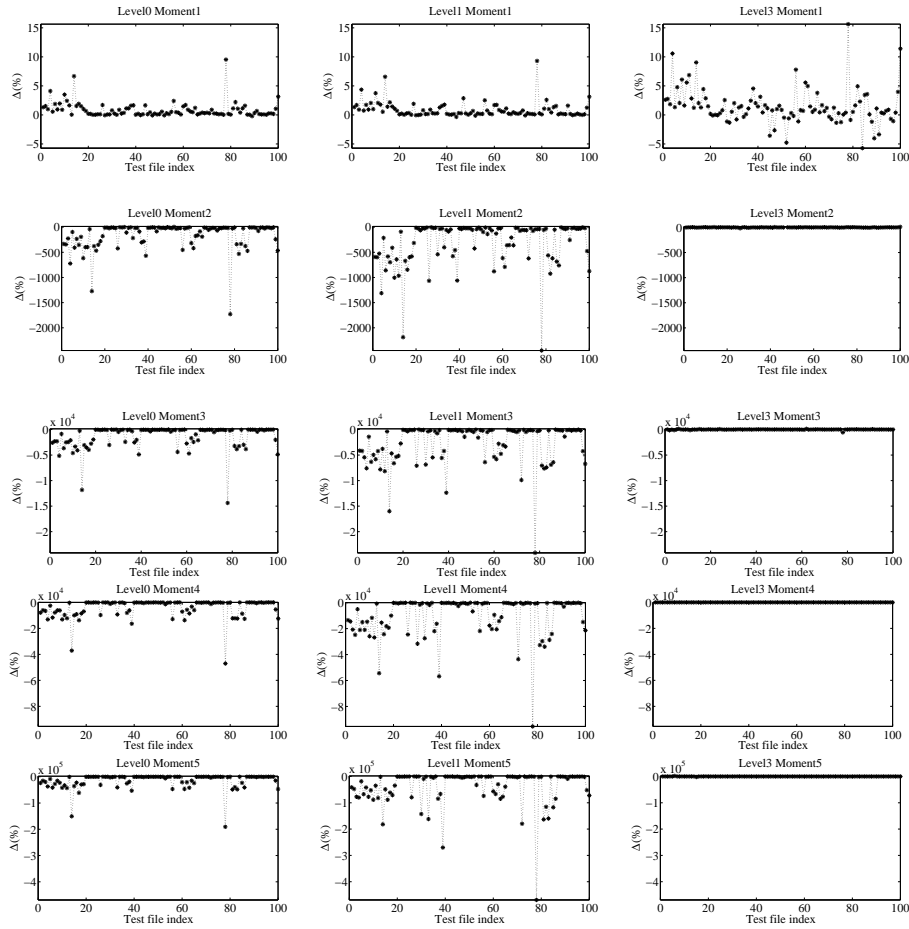
**Fig. 4.** Detection performance at different hidden ratios. (a) FP rate with different training hidden ratios; (b) FN rate with different training hidden ratios; (c) FP rate and FN rate with the 30% + 80% training hidden ratio; (d) correct classification rate with different training hidden ratios.

the test hidden ratio and the training hidden ratio, the better classification performance we achieve. Moreover, with the help of multiple training hidden ratios, the correct classification rate shows strong robust characteristics and for most cases the correct classification rate is much higher than the best one we can get with only one training hidden ratio. Although there is still some small gap between the performance of multiple hidden ratios at higher hidden ratios, e.g., 80% and 100%, the improvement in classification can be achieved by increasing the number of training hidden ratios.

### 4.3 Analysis of Feature Contributions

To measure the effectiveness of each feature in the 25-D feature vector, we define the relative feature distance as:

$$\Delta = \frac{v_s - v_c}{v_s} \times 100\% \quad (6)$$



**Fig. 5.** Feature effectiveness with respect to different wavelet decomposition levels and statistical orders.

where  $v_c$  and  $v_s$  are the feature vectors obtained from cover-audio objects and stego-audio objects, respectively. Figure 5 plots the relative feature distance for 100 audio files randomly selected from our audio database. The hidden ratio in this section is set to be 100%. The results show that the differences between the features of the cover-object and stego-object are less noticeable with the higher wavelet decomposition level. This is because the embedded information corresponds to high frequency noise. In the wavelet decomposition process, the lower levels correspond to the higher frequency bands and higher levels lead to decreasing frequency bands. As a result, the feature at the lower wavelet levels will better detect changes resulting from noise. Finally, Table 1 shows the contribution of each dimension of the 25-D feature vector to the classification performance by

separately applying each dimension as a one-dimensional (1-D) feature for steganalysis. The performance is measured with the correct classification rate for the randomly-selected 895 cover-audio objects and their stego-versions in the training process. The results show that the correct classification rates are different for different 1-D features. Particularly, the correct classification rate is much higher at the lower wavelet decomposition levels compared to the higher levels, and the original signal gets a median correct classification rate within the different wavelet decomposition levels. Also, within the same level, the correct classification rate increases with the increasing moment order. These results confirm our previous analysis for the different feature effectiveness. In addition, it can be observed that the correct classification rates using any 1-D feature vector or any 5-D feature vector are significantly lower than using the combined 25-D feature vector. In other words, the 25 features collectively perform much better, thus these features are complementary in steganalysis.

**Table 1.** Correct classification rate of 1-D feature for the training data.

Moment Order	Level 0	Level 1	Level 2	Level 3	Level 4
1	55.30	57.42	55.02	51.79	51.06
2	62.94	65.56	55.91	53.01	51.78
3	65.90	77.23	56.19	53.57	52.40
4	66.35	79.80	58.25	55.80	52.57
5	67.18	82.75	59.04	56.25	52.68
5-D feature vector	68.97	85.77	69.87	60.71	54.30
25-D feature vector			89.45		

## 5 Conclusion

In this paper, we presented an audio steganalysis method that is based on audio distortion measurement and high-order statistics in the feature selection. A distortion metric based on Hausdorff distance was designed specifically to detect modifications and additions to audio media. We considered the de-noised version of the audio object as an estimate of the cover-object. We then used the Hausdorff distance to measure the distortion. The distortion measurement was obtained at various wavelet decomposition levels from which we derived high-order statistics as features for a classifier to determine the presence of hidden information in an audio signal. Results from simulations with numerous audio sequences showed that our algorithm provides significantly higher detection rates than existing schemes that use standard audio quality metrics or statistical moments without considering audio quality.

## References

1. Dittmann, J., Hesse, D., Hillert, R.: Steganography and steganalysis in voice-over ip scenarios: operational aspects and first experiences with a new steganalysis tool set. In: Security, Steganography, and Watermarking of Multimedia Contents VII, San Jose, CA, USA, SPIE (January 2005) 607–618
2. Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB steganography in color and grayscale images. In: Proceedings of the 2001 workshop on multimedia and security: new challenges, Ottawa, Ontario, Canada, ACM (October 2001) 27–30
3. Johnson, N.F., Jajodia, S.: Steganalysis of images created using current steganography software. In: Proceedings of the Second International Workshop on Information Hiding, Portland, OR, USA, Springer-Verlag (April 1998) 273–289
4. Westfeld, A., Pfitzmann, A.: Attacks on steganographic systems. In: Proceedings of the Third International Workshop on Information Hiding, Dresden, Germany, Springer-Verlag (September 1999) 61–76
5. Johnson, M.K., Lyu, S., Farid, H.: Steganalysis of recorded speech. In Delp III, E.J., Wong, P.W., eds.: Security, Steganography, and Watermarking of Multimedia Contents VII. Volume 5681., SPIE (May 2005) 664–672
6. Ozer, H., Avcibas, I., Sankur, B., Memon, N.D.: Steganalysis of audio based on audio quality metrics. In Delp III, E.J., Wong, P.W., eds.: Security and Watermarking of Multimedia Contents V. Volume 5020., SPIE (January 2003) 55–66
7. Avcibas, I.: Audio steganalysis with content-independent distortion measures. Signal Processing Letters, IEEE **13**(2) (2006) 92–95
8. Huttenlocher, D.P., Klanderma, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on **15**(9) (1993) 850–863
9. Veltkamp, R.C.: Shape matching: similarity measures and algorithms. In: Shape Modeling and Applications, SMI 2001 International Conference on., Genova, Italy (May 2001) 188–197
10. Holtyak, T., Fridrich, J., Voloshynovskiy, S.: Blind statistical steganalysis of additive steganography using wavelet higher order statistics. In: Lecture Notes in Computer Science, Springer Berlin / Heidelberg (2005) 273–274
11. Chandramouli, R., Memon, N.: Analysis of LSB based image steganography techniques. In: Image Processing, 2001. International Conference on, Thessaloniki, Greece (October 2001) 1019–1022
12. Dabeer, O., Sullivan, K., Madhow, U., Chandrasekaran, S., Manjunath, B.S.: Detection of hiding in the least significant bit. Signal Processing, IEEE Transactions on **52**(10) (2004) 3046–3058
13. Dumitrescu, S., Wu, X.: Steganalysis of LSB embedding in multimedia signals. In: Multimedia and Expo, 2002. ICME'02. IEEE International Conference on, Lusanne, Switzerland (August 2002) 581–584
14. Dumitrescu, S., Wu, X., Wang, Z.: Detection of LSB steganography via sample pair analysis. In: Proceedings of the Fifth International Workshop on Information Hiding, Noordwijkerhout, The Netherlands (October 2002) 355–372
15. Farid, H.: Detecting hidden messages using higher-order statistical models. In: Image Processing. 2002. International Conference on, Rochester, NY, USA (September 2002) 905–908
16. Harmse, J.J.: Steganalysis of additive noise modelable information hiding. Master's thesis, Rensselaer Polytechnic Institute, Troy, New York, USA (2003)

17. Shi, Y.Q., Xuan, G., Yang, C., Gao, J., Zhang, Z., Chai, P., Zou, D., Chen, C., Chen, W.: Effective steganalysis based on statistical moments of wavelet characteristic function. In: IEEE International Conference on Information Technology: Coding and Computing, ITCC'05, IEEE Computer Society (April 2005) 768–773
18. Avcibas, I., Memon, N., Sankur, B.: Steganalysis using image quality metrics. Image Processing, IEEE Transactions on **12**(2) (2003) 221–229
19. Watson, A.B., ed.: Digital images and human vision. MIT Press, Cambridge, MA, USA (1993)
20. Nill, N.: A visual model weighted cosine transform for image compression and quality assessment. Communications, IEEE Transactions on **33**(6) (1985) 551–557
21. A Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
22. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association **90**(432) (1995) 1200–1224
23. Burrus, C.S., Gopinath, R.A., Guo, H.: Introduction to wavelets and wavelets transforms, a primer. Prentice Hall, Upper Saddle River, NJ, USA (1998)
24. Veltkamp, R., Hagedoorn, M.: State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands (1999)
25. Wave files. <http://www.wavsurfer.com/>
26. Steghide. <http://steghide.sourceforge.net/>