Speech steganalysis based on the delay vector variance method

OSMAN HİLMİ KOÇAL

EMRAH YÜRÜKLÜ

ERDOĞAN DİLAVEROĞLU

Research Article

Speech steganalysis based on the delay vector variance method

Osman Hilmi KOÇAL[1], **Emrah YÜRÜKLÜ**[2,*], **Erdoğan DİLAVEROĞLU**[3]
[1]Department of Computer Engineering, Yalova University, Yalova, Turkey
[2]Department of Electrical and Electronics Engineering, Bursa Orhangazi University, Bursa, Turkey
[3]Department of Electrical and Electronics Engineering, Uludağ University, Bursa, Turkey

Abstract: This study investigates the use of delay vector variance-based features for steganalysis of recorded speech. Because data hidden within a speech signal distort the properties of the original speech signal, we designed a new audio steganalyzer that utilizes delay vector variance (DVV) features based on surrogate data in order to detect the existence of hidden data. The proposed DVV features are evaluated individually and together with other chaotic-type features. The performance of the proposed steganalyzer method is also discussed with a focus on the effect of different hiding capacities. The results of the study show that using the proposed DVV features alone or in cooperation with other features helps in designing a distinctive audio steganalyzer, as cooperation with other chaotic-type features provides higher performances for stego and cover objects.

Key words: Steganography, steganalysis, speech, chaos, false neighbors, Lyapunov exponent, surrogate data, delay vector variance

1. Introduction

Steganography is the science of covert communication that is applied by hiding secret messages in digital signals. To achieve secure and undetectable communication, a stego signal (used to conceal a secret message) should be indistinguishable from a cover signal, which does not contain any secret message. Analyzing the detection of stego and cover signals is called steganalysis. The set of techniques used for steganalysis is called a steganalyzer.

'Embedding' is a term generally used in the steganalysis literature to express hidden information. Embedding, however, has another meaning in chaos theory. To prevent possible misunderstandings, 'embedding' is used in this paper as it is used in chaos theory, not in steganalysis.

Existing methods of audio steganalysis target various data-hiding techniques. The work in [1] focused on least significant bit-based steganography, whereas the work in [2] addressed the steganalysis of the MP3stega algorithm. The steganalysis of the Hide4pgp algorithm was explored in [3]. Spread spectrum watermarking and stochastic modulation steganography were considered in [4]. Watermarking and steganographic data-hiding methods for time and frequency domains were studied in [5]. An essential reference for the basics of steganography can be found in [6]. See [7] for a brief summary of steganalysis approaches.

Chaotic feature-based audio steganalysis was investigated in [8] and [9] and found useful for distinguishing stego from cover signals. Chaotic phenomena in speech are still the subject of various research studies. Though many of the reported approaches assume the linearity and stationarity of audio, theoretical and experimental

---

*Correspondence: emrah.yuruklu@bou.edu.tr

evidence for the existence of chaotic phenomena in speech signals also exists, which cannot be covered by linear modeling [8–10]. The delay vector variance (DVV) method, which is based on surrogate time series [11], is one technique that determines the level of nonlinearity in signals [12]. Assuming that hiding data leads to additive distortion of a speech signal, it is anticipated that this process will change the nonlinear and chaotic structure of the speech signal, and consequently the DVV and chaotic-based features.

A change in chaotic structure also means a change in nonlinear structure, which manifests as a deviation from the numerical values of the DVVs of the cover signal and therefore enables the detection of the possible existence of a hidden signal. Using DVV-based features was proposed for the first time in [9]. Performance results in [9], which was studied only one special test case, showed that DVV-based features are very promising for speech steganalysis even though they were used alone, i.e. not combined with any other chaotic-based features. In this paper, research cases are extended and the performance results of [9] are improved with a combination of DVV-based features when compared to the authors' previously proposed chaotic features [8].

Section 2 of this paper provides an overview of DVV analysis for nonlinear signals. Section 3 shows the application of DVV analysis to steganalysis and DVV feature extraction, and also shows how hiding data changes DVV-based features. Feature selection and the results of the experiments are given in Section 3. Conclusions are drawn in Section 4.

2. DVV method

To assess the presence of nonlinearity in time series, the 'surrogate time series' method is a widely used technique offered by Theiler et al. [12]. A surrogate time series is produced with the same magnitude and a similar Fourier phase to transform the original time series. Although there are different approaches to producing surrogate time series, the most widely used is the iterative amplitude adjusted Fourier transform (iAAFT), which was proposed by Schreiber and Schmitz [13]. In this paper, the iAAFT approach is used to produce surrogate time series. For detailed information, please refer to [13].

By comparing the characteristics of the original and the surrogate time series, the level of nonlinearity in the time series can be anticipated. Metrics based on surrogate times are defined for performing such a comparison.

2.1. Nonlinearity measurements

Besides the DVV method, there are two other approaches used to predict the nonlinearity of the time series. These are third-order autocovariance and asymmetry due to time reversal [14].

Third-order autocovariance is a higher-order extension of the traditional autocovariance, and can be given by:

$$t^{C3}(\tau) = \langle x_n x_{n-\tau} x_{n-2\tau} \rangle \tag{1}$$

Here, τ is the time lag. Given that the time series is accepted as time-reversible, if the probabilistic properties do not change with regards to time reversal, invariance for the probabilistic features can be measured by using the following equation:

$$t^{REV}(\tau) = \left\langle (x_n - x_{n-\tau})^3 \right\rangle \tag{2}$$

In [13,14], it was shown that with the combination of DVV, these two methods are very helpful for two-tailed tests of nonlinearity.

2.2. DVV method

The DVV method was first proposed in 2003 [14]. It is a generic predictor of nonlinearity that uses phase spaces of time series. For an embedding dimension D_E, a set of delay vectors $\mathbf{x}(n)$ are generated with time lag τ:

$$\mathbf{x}(n) = [x_{n-D_E\tau}, ..., x_{n-\tau}] \tag{3}$$

The definition of proper values and the effects of selecting improper values for the embedding dimension D_E and time delay τ were extensively described in [8,11]. Figure 1 illustrates a sample of the embedding operation of a real audio time series into a 3D phase space.
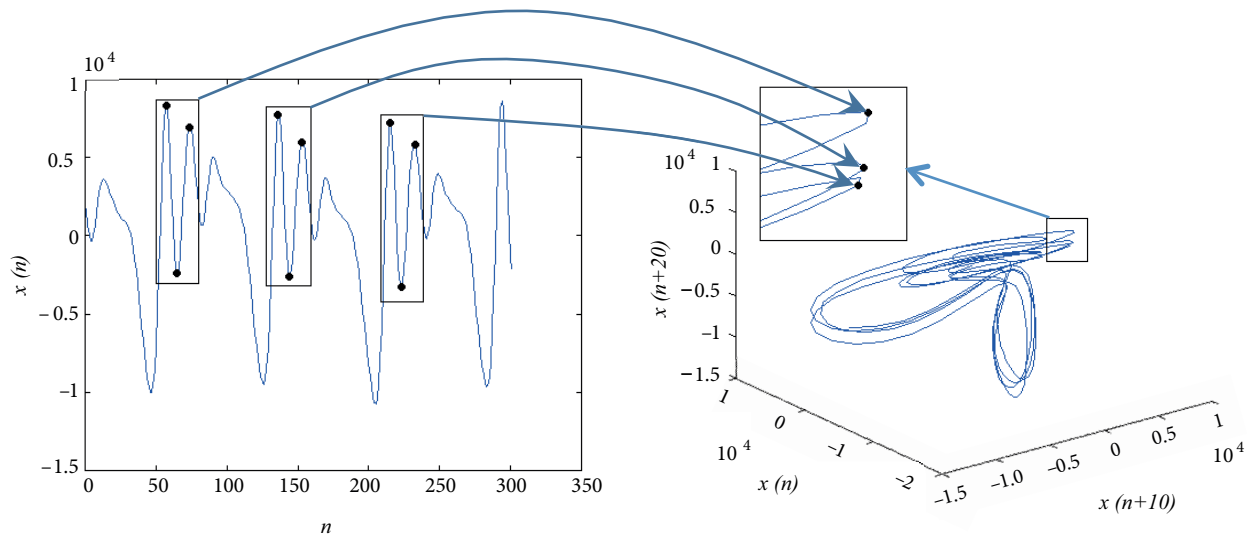

Figure 1. Phase space of a real speech segment for $T = 10$ and $D_E = 3$.

To determine a specific embedding dimension, D_E, the DVV method computes the mean target variance, σ^{*2}, for all sets of Ω_n. Here, every Ω_n is a group that consists of delay vectors that are a certain distance from $\mathbf{x}(n)$. The distance is varied in relation to the distribution of pairwise distances between delay vectors [14].

The DVV method can be summarized as follows:

$$\sigma^{*2}(r_d) = \frac{\frac{1}{N}\sum_{n=1}^{N} \sigma_n^2(r_d)}{\sigma_x^2} \tag{4}$$

- For the given embedding dimension D_E, the mean, μ_d, and the standard deviation, σ_d, are computed over all pairwise Euclidean distances between delay vectors.

- For the given embedding dimension, D_E, and $\Omega_n(r_d)$ sets are generated according to the following formula: $\Omega_n(r_d) = \{\mathbf{x}(i)|\,\|\mathbf{x}(n) - \mathbf{x}(i)\| \leq r_d\}$; $i = 1, 2,.., N$. Here, N is the sample count in the time series of $\mathbf{x}$, and r_d is the particular distance that must be chosen from the interval $[\mu_d - -n_d\sigma_d; \mu_d + n_d\sigma_d]$, where n_d is a parameter that controls the span over which to compute the DVV plot.

- For the given embedding dimension D_E, the variance of every set of $\Omega_n(r_d)$, $\sigma_n^2(r_d)$ is computed. The average value of all sets, $\Omega_n(r_d)$, is normalized by the variance of the time series σ_x^2. At the end, the measure of unpredictability is computed by:

Unless one set, $\Omega_n(r_d)$, contains more than 30 delay vectors, it is not taken into account when performing the computations [14].

As a result of standardizing the distance axis, the DVV plots are easily interpreted. Figure 2 shows the DVV plots of four benchmark signals to further explain DVV. According to the plots, the Henon Map signal (A) is the most predictable signal, while colored noise (C) is the most unpredictable. Note that all the x-axes are standardized as distances to make the comparison easier [15].
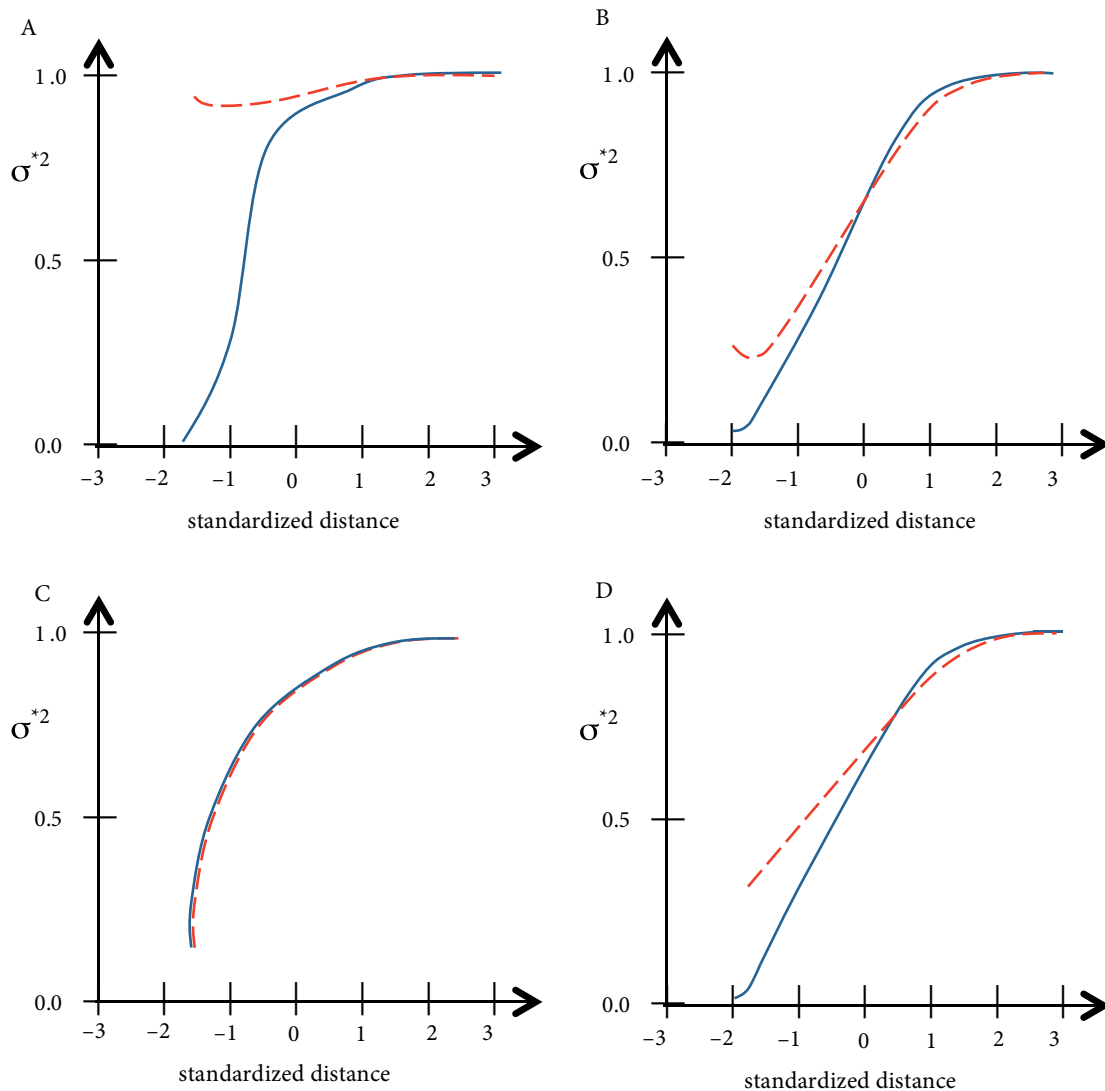
Figure 2. Solid curves represent the DVV plots for (a) the Henon Map, (b) Mackey–Glass, (c) colored noise, and (d) the laser time series. Average DVV plots computed over 99 surrogates are shown as dashed curves [15].

3. A new steganalyzer feature set using the DVV method

By evaluating Figure 2, the nonlinear dynamics of the signal become easier to interpret. However, this approach is not enough, given that objective decisions are not used when creating evaluations. For this reason, Schreiber

and Schmitz created the method described below, which produces numerical results [13]:

$$n\_surr \geq \frac{2}{\alpha} - 1 \tag{5}$$

1. The number of surrogate data, n_surr, is decided according to the defined limit of significance, α (generally 0.05):

2. The n_surr surrogate time series is produced from the original time series $\mathbf{s}_1$, $\mathbf{s}_2$,...$\mathbf{s}_{n\_surr}$.

3. DVV curves of the original and surrogate time series are constructed.

4. Statistically analyze DVV curves by using one-sided or two-sided tests to find out their coherence with the original data by using various methods.

For the fourth step, we decided to use *RMSE* values for measuring the difference between the DVV curves of the original and the surrogate signal. For n_surr surrogate time series, the *RMSE* vector is constituted as $RMSE = \{RMSE_1, RMSE_2,..., RMSE_{n\_surr}\}$.

For using DVV-based features for speech steganalysis, we have designed a feature set based on calculated *RMSE* values of surrogate time series. The proposed feature vector, which must be constituted for every original time series (i.e. audio record), consists of four elements—mean, variance, skewness, and kurtosis—which are values that are calculated over *RMSE* values [9]:

$$\mathbf{F}_{surr} = \{mean\,(RMSE)\,var\,(RMSE)\,ske\,(RMSE)\,kur\,(RMSE) \tag{6}$$

In the Ω_n set, the square value of the neighborhood distances between all delayed vector couples is:

$$d(x(n), x(m))^2 = \|x(m) - x(n)\|^2 = \sum_{k=1}^{D_E} (x(n+k) - x(m+k))^2 \tag{7}$$

The mean, μ_d, and variance values, σ_d^2, in the Ω_n set are:

$$\mu_d \underset{x(m)\in\Omega_n}{=} E[d] \; \sigma_d^2 \underset{x(m)\in\Omega_n}{=} E\left[(d-\mu_d)^2\right] = E\left[d^2\right] - \mu_d^2 \tag{8}$$

Assuming that hiding data in cover signals means adding zero-mean and σ_ε^2 varianced white noise with a magnitude of $\varepsilon(n)$, if we reorganize Eq. (7) for both cover and stego signal, then:

$$d_C(x(n), x(m))^2 = d_C^2 = \sum_{k=1}^{D_E} (x(n+k) - x(m+k))^2 \tag{9}$$

$$d_S(x(n), x(m))^2 = d_S^2 = \sum_{k=1}^{D_E} ([x(n+k) + \varepsilon(n+k)] - [x(m+k) + \varepsilon(m+k)])^2 \tag{10}$$

If we expand the term of d_S^2, which is used for stego signals, then:

$$
\begin{aligned}
d_S^2 \;=\; & \sum_{k=1}^{D_E} \{x^2(n+k) + 2x(n+k)\varepsilon(n+k) + \varepsilon^2(n+k) + x^2(m+k) + 2x(m+k)\varepsilon(m+k) \\
& + \varepsilon^2(m+k) - 2\left[x(n+k)x(m+k) + x(n+k)\varepsilon(m+k) \; + \varepsilon(n+k)x(m+k) + \varepsilon(n+k)\varepsilon(m+k)\right]\}
\end{aligned}
\tag{11}
$$

If we put $d_C^2 = \sum\limits_{k=1}^{D_E} x^2(n+k) + 2x(n+k)x(m+k) + x^2(m+k)$ into Eq. (11), then:

$$
\begin{aligned}
d_S^2 \quad = \quad & d_C^2 + \sum_{k=1}^{D_E} \left\{ 2x(n+k)\varepsilon(n+k) + \varepsilon^2(n+k) + 2x(m+k)\varepsilon(m+k) + \varepsilon^2(m+k) \right. \\
& \left. -2\left[x(n+k)\varepsilon(m+k) + \varepsilon(n+k)x(m+k) + \varepsilon(n+k)\varepsilon(m+k) \right] \right\}
\end{aligned}
\tag{12}
$$

As long as all the pairs of $\langle \varepsilon(m+k), \varepsilon(n+k) \rangle$, $\langle \varepsilon(m+k), \varepsilon(m+k) \rangle$, $\langle \varepsilon(m+k), x(m+k) \rangle$, $\langle \varepsilon(m+k)x(n+k) \rangle$, $\langle \varepsilon(n+k), x(m+k) \rangle$ ve $\langle \varepsilon(n+k), x(n+k) \rangle$, are independent and identically distributed, the mean values of Eq. (12) can be simplified as:

$$
\underset{x(n)\in\Omega_n}{E\left[d_S^2\right]} = \underset{x(n)\in\Omega_n}{E\left[d_C^2\right]} + D_E . E\left[\varepsilon^2(n+k)\right] + D_E . E[\varepsilon^2(m+k)]
\tag{13}
$$

Because $\varepsilon(n)$ is zero-mean, and σ_ε^2 is varianced, white noise is calculated as:

$$
\underset{x(n)\in\Omega_n}{E\left[d_S^2\right]} = \underset{x(n)\in\Omega_n}{E\left[d_C^2\right]} + D_E . 2\sigma_\varepsilon^2
\tag{14}
$$

The variance value used in DVV analysis and seen in Eq. (8) can be defined for cover and stego signals as follows:

$$
(\sigma_d^2)_C \underset{x(m)\in\Omega_n}{=} E\left[d_C^2\right] - (\mu_d)_C^2
\tag{15}
$$

$$
(\sigma_d^2)_S \underset{x(m)\in\Omega_n}{=} E\left[d_S^2\right] - (\mu_d)_S^2
\tag{16}
$$

Using Eqs. 414) and (15), the term $(\sigma_d^2)_S$ can be written as:

$$
(\sigma_d^2)_S = (\sigma_d^2)_C + (\mu_d)_C^2 + D_E . 2\sigma_\varepsilon^2 - (\mu_d)_S^2
\tag{17}
$$

With the assumption of $(\mu_d)_C^2 \approx (\mu_d)_S^2$, Eq. (17) can be simplified as:

$$
(\sigma_d^2)_S = (\sigma_d^2)_C + D_E . 2\sigma_\varepsilon^2
\tag{18}
$$

Such results show that the variance values of stego signals used in DVV analysis are always higher than those of the cover signal.

The first three elements of the proposed $\mathbf{F}_{surr}$ feature vector can be seen in Figure 3 and are calculated over 2000 stego and cover signals using DSSS [16] and stochastic modulation [17] steganography techniques. Here the n_surr value is 20, and the difference between the feature values of the cover and stego signals is demonstrated in Figure 3.
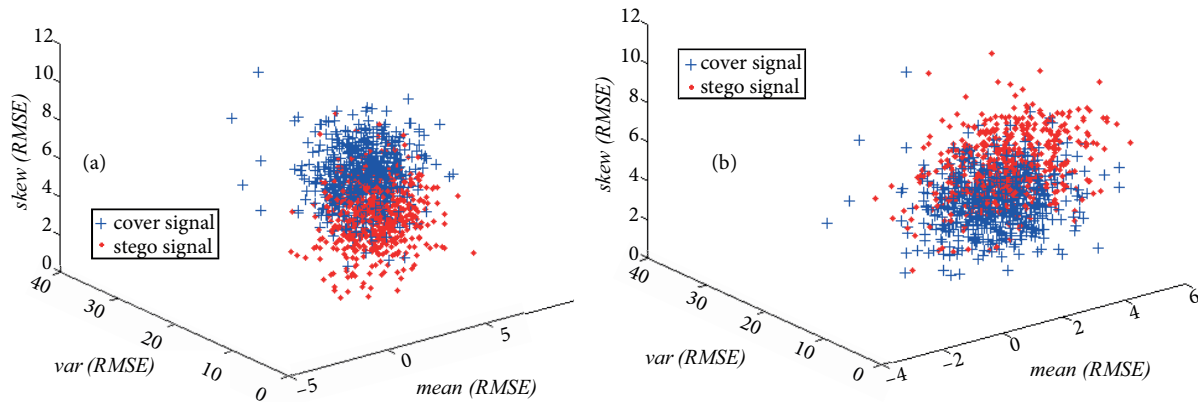
Figure 3. The first three elements of F_{surr} of 2000 cover and stego signals, which are constituted with (a) steganographic techniques DSSS and (b) stochastic modulation. n_{surr} is 20 for F_{surr} .

4. Experimental results

Tests were performed using nine different methods of hiding data. Five of these methods used watermarking techniques, whereas the other four used steganographic techniques. Watermarking methods were used to extend the case to the widest range of possibility, although a stego signal can be deliberately changed before the signal is received [18]. The watermarking techniques used were direct-sequence spread spectrum (DSSS), frequency hopping with spread spectrum (FHSS), echo hiding (ECHO) [16], stochastic modulation (STOMOD) [17], and DCT-based watermarking (COX) [19]. The four steganographic methods used were Steganos (www.steganos.com), MP3Stego (www.petitcolas.net/fabien/steganography/mp3stego), Steghide (http://steghide.sourceforge.net), and Hide4Pgp (www.heinz-repp.onlinehome.de/Hide4PGP.htm). Note that stochastic modulation can be used for steganography; in this study, however, using STOMOD for watermarking is taken into account.

These methods were selected due to their popularity, free availability, and wide usage in related works. The individual performance of steganography and watermarking methods is used to determine whether the proposed feature set is useful for audio steganalysis or not.

4.1. Dataset

The database used for the test scenario is a subset of the TIMIT speech database (https://catalog.ldc.upenn.edu/LDC93S1), which is widely known in the evaluation of automatic speech recognition systems. The TIMIT database comprises over 6300 utterances from 630 male and female speakers, sampled at 16 kHz. Two thousand speech segments, ignoring dialects and male–female differences, were randomly selected from the TIMIT database for experiments in this study. For every data-hiding method studied, a stego signal subset was constituted that contained hidden data in utterances using the data-hiding method.

4.2. Data hiding

We concealed messages into excerpts using nine data-hiding methods. The procedure for concealing messages was randomly selected; half of the set used stego and cover signals for training, while the other half was used for testing. An objective distortion measure, the signal-to-watermark ratio (SWR), was used to define the payload level in the watermarking methods:

$$SWR = \frac{\sum x(n)^2}{\sum (x(n) - y(n))^2},$$ (19)

where $x(n)$ and $y(n)$ signify cover and watermarked signals, respectively. For steganographic methods, the performances of three different data-hiding rates of each individual method were evaluated and found to be at 100% and 50% of maximum allowed capacity.

4.3. Feature set

The feature set proposed in this study has four elements, all of which are based on DVV values, as described in Eq. (6). This new feature set was used for all steganalysis trials, unless otherwise stated. In producing the DVV-based feature vector, the DVV MATLAB Toolbox was utilized (http://www.commsp.ee.ic.ac.uk/~mandic/dvv.htm).

To determine if DVV-based features can be used to improve the performance of other chaotic-based steganalyzer feature sets already investigated in the literature, a unified feature set was constituted and tested. The proposed chaotic-based feature set [8] is used for this purpose, which consists of 22 elements and is defined below:

$$\mathbf{F_{chaotic}}(D_E) = \{\mathbf{F}_{FNF}(D_E)|\, D_E = 3, 4, 5, 6, 7\} \cup \lambda\{_i|\, i = 1, 2, 3, 4, 5, 6, 7\} \tag{20}$$

Here, $\mathbf{F}_{FNF}(D_E)$ is the feature set of the false neighbor fraction (FNF), which is based on the statistical values of the FNF method [20], while λ_i is i. The Lyapunov exponent (LE) value of the signal is as shown in [21]. Using these chaotic features in audio steganalysis is very advantageous, as described in [8]. $\mathbf{F}_{FNF}(D_E)$ is described as follows:

$$F_{FNF}(D_E) = [FNF, mean\,(d_{D_E}(s(n), s(m))), RMS\,(d_{D_E}(s(n), s(m)))] \tag{21}$$

Here, the three elements of the feature vector are the fraction of false neighbors, the average size of the neighborhood, and the root mean-squared size of the neighborhood. The FNF and LE values of the signals are calculated with TISEAN [22].

By assembling chaotic features with proposed DVV-based features, a unified feature set, which comprises 26 elements, can be defined as follows:

$$\mathbf{F}\,(D_E) = \mathbf{F}_{surr} \cup \{\mathbf{F}_{FNF}(D_E)|\, D_E = 3, 4, 5, 6, 7\} \cup \lambda\{_i|\, i = 1, 2, 3, 4, 5, 6, 7\} \tag{22}$$

4.4. Feature selection and classifiers

To achieve the highest detection rate in order to evaluate redundant features, a sequential forward floating search method (SFFS) [23] was coupled with a support vector machine (SVM) classification method (http://sourceforge.net/projects/svm/). A radial basis function with gamma equal to 4 was used in the SVM as a SVM function.

4.5. Simulation results

To assess the validity of the proposed feature set, several tests were carried out that considered the effect of payload. The performance of the DVV-based steganalyzer is provided in Table 1 for the TIMIT dataset. As a performance metric, results are given as percent of the performance, missed detection count (MISS), and false alarm count (FA). In all simulations, half of the dataset was used for training the SVM classifier and the remaining half was used to test the designed steganalyzer with a trained classifier.

In the TIMIT dataset, it is observed that SWR values greater than 38 dB for DSSS, 34 dB for FHSS, 20 dB for Cox, and 18 dB for ECHO become noticeable, namely audible. For this reason, 20 dB, 30 dB, and 40 dB were selected as payloads. It is shown that a 30-dB payload causes performance to drop to 55% with the

Table 1. SVM-based steganalysis performance for different data-hiding capacities with DVV features.

Watermarking methods				
Method	SWR	MISS	FA	%
DSSS	20 dB	0/1000	0/1000	100.0
	30 dB	1/1000	2/1000	99.9
	40 dB	5/1000	6/1000	99.5
FHSS	20 dB	2/1000	0/1000	99.9
	30 dB	5/1000	6/1000	99.5
	40 dB	16/1000	13/1000	98.0
ECHO	20 dB	98/1000	236/1000	83.3
	30 dB	176/1000	395/1000	71.5
	40 dB	293/100	489/1000	60.9
COX	20 dB	285/1000	212/1000	75.2
	30 dB	452/1000	457/1000	54.6
	40 dB	478/1000	485/1000	51.9
STOMOD	20 dB	13/1000	8/1000	99.0
	30 dB	53/1000	34/1000	95.7
	40 dB	171/1000	98/1000	89.8

Steganography methods				
Method	Capacity usage	MISS	FA	%
STEGA	50%	142/1000	166/1000	84.5
	100%	64/1000	73/1000	93.2
HIDE4PGP	50%	42/1000	176/1000	89.1
	100%	33/1000	84/1000	94.2
STEGHIDE	50%	121/1000	154/1000	86.3
	100%	48/1000	83/1000	93.5
MP3	50%	287/1000	167/1000	77.3
	100%	203/1000	124/1000	83.7

Cox method; note, however, that this level of payload is audible. For steganographic methods, the performance of each method was evaluated individually for two different data-hiding rates: 100% and 50% of maximum allowed capacity. There was no audibility concern for steganographic methods, as the maximum data-hiding rate used in the tests is equal to the maximum allowed capacity—the audibility threshold. The performance of the steganalyzer decreases as payload (in terms of SWR or capacity usage) also decreases. This is because a higher payload causes more corruption to the audio data, which makes detecting hidden data much simpler.

Curves of the receiver-operating characteristics (ROCs) for watermarking and steganography methods are illustrated in Figure 4. The calculated conditions for producing ROC curves hold the same configuration used in Table 1—namely, the same database and the same SVM classifier. ROC curves provide performance
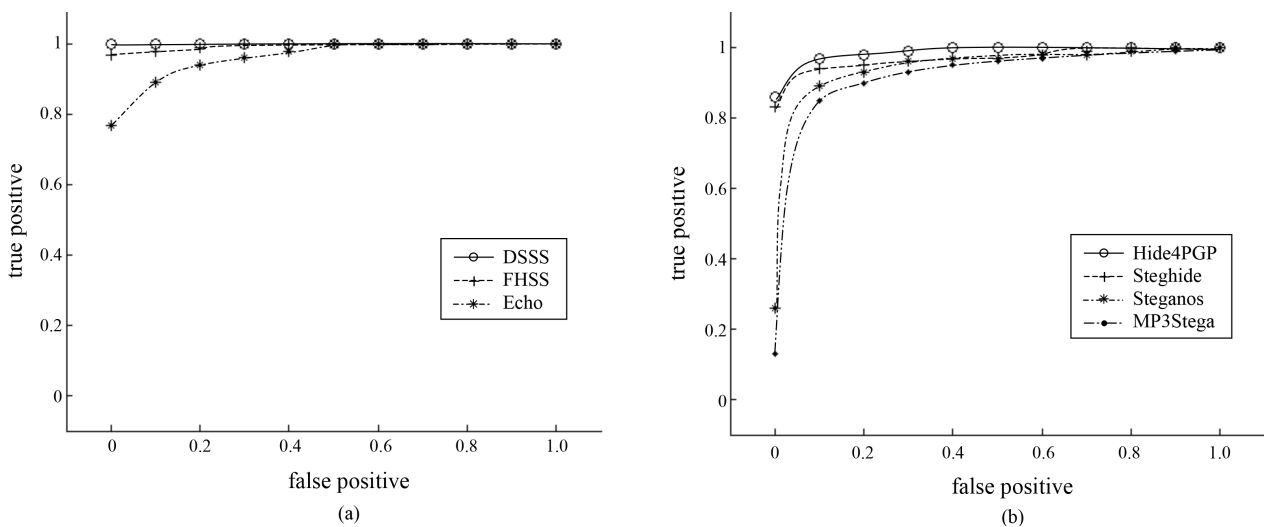

Figure 4. ROC curves for (a) watermarking and (b) steganographic methods obtained using the SVM classifier. The dataset is a random sample of 2000 speeches from the TIMIT database. The payload is 20 dB for watermarking methods and 100% capacity usage for steganographic methods.

Table 2. SVM-based steganalysis performance for different data-hiding capacities with chaotic features.

Watermarking methods				
Method	SWR	MISS	FA	%
DSSS	20dB	0/1000	0/1000	100.0
	30 dB	0/1000	0/1000	100.0
	40 dB	2/1000	2/1000	99.8
FHSS	20 dB	0/1000	0/1000	100.0
	30 dB	1/1000	3/1000	99.8
	40 dB	13/1000	9/1000	99.0
ECHO	20 dB	74/1000	148/1000	88.9
	30 dB	132/1000	295/1000	78.7
	40 dB	212/100	437/1000	67.6
COX	20 dB	227/1000	245/1000	76.4
	30 dB	435/1000	442/1000	56.2
	40 dB	471/1000	451/1000	53.9
STOMOD	20 dB	7/1000	5/1000	99.4
	30 dB	40/1000	22/1000	96.9
	40 dB	108/1000	61/1000	91.5

Steganography methods				
Method	Capacity usage	MISS	FA	%
STEGA	50%	105/1000	117/1000	88.9
	100%	32/1000	84/1000	94.2
HIDE4PGP	50%	32/1000	143/1000	91.3
	100%	29/1000	55/1000	95.8
STEGHIDE	50%	44/1000	65/1000	94.6
	100%	19/1000	32/1000	97.5
MP3	50%	239/1000	127/1000	81.7
	100%	154/1000	114/1000	86.6

results if the threshold of the decision is changed. Figure 4 illustrates that DVV-based features are powerful even with biased thresholds.

The unified feature set, as described in Eq. (22), is also evaluated to investigate the unification of DVV-based features with other proposed, chaotic-type steganalyzers. The performance results of the steganalyzer with unified features can be seen in Table 2. The steganalyzer used DVV-based features only, showing that using a unified set of features provides higher performance results.

Table 3 gives the comparative results of using chaotic and DVV-based features individually and together. Using the support of DVV-based features, chaotic-based features are more discriminative than when they are alone. Comparing DVV-based and chaotic-type features shows that even though DVV-based features use very limited information from the audio data (namely, *RMSE* values of the original and surrogate DVV curves), the performance results are very promising. Since chaotic-type features were selected from a variety of chaos measurement methods [8], high performance results should be expected. However, the DVV-based steganalyzer achieved slightly higher performance results for some steganalysis conditions. Table 2 shows that

Table 3. Comparative performance results of different feature-based steganalyzers for different data-hiding capacities.

Watermarking methods				
Method	SWR	CHAOTIC	DVV	CHAOTIC + DVV
DSSS	20 dB	100.0	100.0	100.0
	30 dB	100.0	99.9	100.0
	40 dB	99.8	99.5	99.8
FHSS	20 dB	100.0	99.9	100.0
	30 dB	99.8	99.5	99.8
	40 dB	98.8	98.0	99.0
ECHO	20 dB	87.8	83.3	88.9
	30 dB	78.5	71.5	78.7
	40 dB	66.4	60.9	67.6
COX	20 dB	72.2	75.2	76.4
	30 dB	55.4	54.6	56.2
	40 dB	52.3	51.9	53.9
STOMOD	20 dB	99.3	99.0	99.4
	30 dB	96.6	95.7	96.9
	40 dB	90.3	89.8	91.5

Steganography methods				
Method	Capacity usage	CHAOTIC	DVV	CHAOTIC + DVV
STEGA	50%	88.2	84.5	88.9
	100%	92.6	93.2	94.2
HIDE4PGP	50%	90.8	89.1	91.3
	100%	93.8	94.2	95.8
STEGHIDE	50%	88.8	86.3	94.6
	100%	94.4	93.5	97.5
MP3	50%	81.0	77.3	81.7
	100%	85.5	83.7	86.6

with the support of DVV-based features, the performance of a chaotic-based steganalyzer provides even better performance.

When comparing recently proposed audio steganalysis approaches, the proposal of Geetha et al. [23], which uses Hausdorff distance and higher-order statistics, is analyzed. In Geetha et al.'s study, the selected classifier is the J48 decision tree algorithm, and a database of 200 audio samples was used for steganalysis. For our study, the SVM is selected as a classifier, and 2000 randomly selected audio samples from the TIMIT database were used. Even though the database and classifier used in this study differ from those of Geetha et al.'s study, information about the performance of the proposed steganalyzers is still provided. For a fair comparison, tests should be performed using the same database and detection algorithms; however, Table 4 shows that the proposed DVV-based steganalyzer provides better performance for the DSSS and Steghide methods and poorer performance for the ECHO method.

Table 4. Comparative performance results of different feature-based steganalyzers with different database and classifiers.

Method	Capacity usage	Hausdorff	DVV	Chaotic + DVV
DSSS	10% (SWR: 33 dB)	78.6	99.7	99.9
Echo	10% (SWR: 26 dB)	88.4	78.8	83.4
Steghide	10%	83.2	78.7	86.4

5. Conclusion

We have proposed a new feature vector based on DVV analysis of nonlinear data. We have constituted a new numerical feature vector by using DVV analysis, which is normally used for detecting the nonlinearity level of signals. Tests for the proposed steganalyzer were carried out by using the TIMIT database. Compared to recently proposed similar steganalyzer studies [3,5,8,24–26], DVV-based features used as speech steganalyzers show very promising results, specifically for the DSSS, FHSS, STOMOD, Hide4PGP, and Steghide methods. Moreover, tests were repeated for a new steganalyzer that is composed of a combination of DVV-based and chaotic-based features. A number of test results for different embedding capacities and different decision thresholds (ROC curves) are also presented. Use of DVV-based features in cooperation with other chaotic features improves the performance further. This combined steganalyzer provides the advantages of both DVV-based and chaotic features, and it can be used for almost all speech steganography methods proposed in the literature.

Moreover, comparative results are obtained from our proposed steganalyzers and those of Geetha et al. [24] for the same usage capacities used in [24].

Test results show that for the Steghide and Hide4pgp steganography methods, our combined DVV-based and chaotic-based steganalyzer has the most distinctive feature set among the published high-performance steganalyzers [3,5,8,9,24–26]. Given the performance results of this study, it can easily be said that DVV-based features can be used for speech steganalysis as well as nonlinearity detection. In fact, the combined DVV- and chaotic-based steganalyzer is the best tool for distinguishing speech data from secret data hidden by the Steghide and Hide4pgp steganography methods.

For the special case of mp3 audio files, recent studies have shown that with specific features, a more than 95% detection rate is achievable with SVM classifiers, even at low hiding capacity rates such as 20% [27,28]. The performance of the proposed steganalyzer can be increased further by using side information specific to the mp3 files. As long as chaotic-type and DVV-based steganalyzers fall short of this level of achievement,

the proposed DVV and chaotic-type features should be preferred for non-mp3 samples, namely coded but not compressed audio files such as .wav files.

Acknowledgments

<div align="center">References</div>

[1] Westfeld AP. Detecting low embedding rates. In: IH 2002 International Workshop; 7–9 October 2002; Noordwijk-erhout, the Netherlands. Berlin, Germany: Springer-Verlag, 2003. pp. 324-339.

[2] Westfeld AP, Pfitzmann A. Attacks on steganographic systems. In: IH 1999 International Workshop; 29 September–1 October 1999; Dresden, Germany. Heidelberg, Germany: Springer-Verlag, 1999. pp. 61-66.

[3] Johnson MK, Lyu S, Farid H. Steganalysis of recorded speech. In: SPIE 2005 Symposium on Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents; 16–20 January 2005; San Jose, CA, USA. Bellingham, WA, USA: SPIE. pp. 664-672.

[4] Altun O, Sharma G, Çelik M, Sterling M, Titlebaum E, Bocko M. Morphological steganalysis of audio signals and the principle of diminishing marginal distortions. In: IEEE 2005 International Conference on Acoustics, Speech, and Signal Processing; 18–23 March 2005; Philadelphia, PA, USA. New York, NY, USA: IEEE. pp. 21-24.

[5] Özer H, Avcıbaş İ, Sankur B, Memon N. Steganalysis of audio based on audio quality metrics. In: SPIE 2003 Conference on Security and Watermarking of Multimedia Contents; 20 January 2003; Santa Clara, CA, USA. Bellingham, WA, USA: SPIE. pp. 55-66.

[6] Djebbar F, Ayad B, Meraim KA, Hamam H. Comparative study of digital audio steganography techniques. EURASIP J Audio SPEE 2012; 25: 1-16.

[7] Meghanathan N, Nayak L. Steganalysis algorithms for detecting the hidden information in image, audio and video cover media. International Journal of Network Security & Its Application 2010; 2: 43-55.

[8] Koçal OH, Yürüklü E, Avcıbaş İ. Chaotic-type features for speech steganalysis. IEEE T Inf Foren Sec 2008; 3: 651-661.

[9] Koçal OH, Yürüklü E, Dilaveroğlu E. A new approach for speech audio steganalysis using delay vector variance method. J Fac Eng Arch 2014; 19: 27-36.

[10] Kokkinos I, Maragos P. Nonlinear speech analysis using models for chaotic systems. IEEE T Speech Audi P 2005; 13: 1098-1109.

[11] Banbrook M, McLaughlin S. Speech characterization and synthesis by nonlinear methods. IEEE T Speech Audi P 1999; 7: 1-17.

[12] Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. Physica D 1992; 58: 77-94.

[13] Schreiber T, Schmitz A. Discrimination power of measures for nonlinearity in a time series. Phys Rev E 1997; 55: 5443-5447.

[14] Gautama T, Mandic DP, Van Hulle MM. Indications of nonlinear structures in brain electrical activity. Phys Rev E 2003; 67: 046204.1-046204.5.

[15] Gautama T, Mandic DP, Van Hulle MM. A novel method for determining the nature of time series. IEEE T Bio-Med Eng 2004; 51: 728-736.

[16] Bender W, Gruhl D, Morimoto N, Lu A. Techniques for data hiding. IBM Syst J 1996; 35: 313-336.

[17] Fridrich J, Goljan M. Digital image steganography using stochastic modulation. In: SPIE 2003 Conference on Security and Watermarking of Multimedia Contents; 20 January 2003; Santa Clara, CA, USA. Bellingham, WA, USA: SPIE. pp. 191-202.

[18] Simmons GJ. Prisoners' problem and the subliminal channel. In: Chaum D, editor. Advances in Cryptology. New York, NY, USA: Plenum Press, 1984. pp. 51-67.

[19] Cox IJ, Kilian J, Leighton FT, Shamoon T. Secure spread spectrum watermarking for multimedia. IEEE T Image Process 1997; 6: 1673-1687.

[20] Kennel MB, Abarbanel HDI. False neighbors and false strands: a reliable minimum embedding dimension algorithm. Phys Rev E 2002; 66: 026209.

[21] Hilborn R. Chaos and Nonlinear Dynamics. 2nd ed. Oxford, UK: Oxford University Press, 2000.

[22] Hegger R, Kantz H, Schreiber T. Practical implementation of nonlinear time series methods: the TISEAN package. Chaos 1999; 9: 413-435.

[23] Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. Pattern Recogn Lett 1994; 15: 1119-1125.

[24] Geetha S, Ishwarya N, Kamaraj N. Audio steganalysis with Hausdorff distance higher order statistics using a rule based decision tree paradigm. Expert Syst Appl 2010; 37: 7469-7482.

[25] Avcıbaş İ. Audio steganalysis with content-independent distortion measures. IEEE Signal Proc Let 2006; 13: 92-95.

[26] Özer H, Sankur B, Memon N, Avcıbaş İ. Detection of audio covert channels using statistical footprints of hidden messages. Digit Signal Process 2006; 16: 389-401.

[27] Qiao M, Sung AH, Liu Q. MP3 audio steganalysis. Inform Sciences 2013; 31: 123-134.

[28] Yu X, Wang R, Yan D, Zhu J. MP3 Audio steganalysis using calibrated side information feature. J Comput Inf Syst 2012; 8: 4241-4248.