SPEECH STEGANALYSIS USING CHAOTIC-TYPE FEATURES

Osman Hilmi Koçal, Emrah Yürüklü, İsmail Avcıbaş

Uludag University, Department of Electronics Engineering, 16059, Bursa, Turkey
Tel: +90 224 442 8179
Fax: +90 224 442 8021
{kocal, yuruklu, avcibas}@uludag.edu.tr

ABSTRACT

In this paper, we investigate chaotic-type features for universal speech steganalysis. These features are used in the design of linear regression classifier. The steganalyzer is tested on watermarking and steganographic methods. Experimental results show the potential of chaotic-type features for their discriminatory power to be used in steganalysis. We believe that the integration of chaotic-type features with linear ones may capture both linear and non-linear aspects of speech signals leading to robust and efficient hybrid steganalyzers.

1. INTRODUCTION

Steganography is the art and science of hiding the very presence of communication by embedding secret messages into innocent looking electronic signals such as digital images, video and audio. To achieve covert communication, *stego-signals*, signals containing a secret message, should be indistinguishable from *cover-signals,* signals not containing any secret message. In this respect, s*teganalysis* is the set of techniques that aim to distinguish between cover-signals and stego-signals. In this paper, we address the steganalysis of digital speech signals.

Steganographic algorithms differ in hiding the messages. Least Significant Bit (LSB) method embeds the message by flipping the LSBs of audio samples [1, 2] or, alternatively, transforms coefficients [3, 4]. Spread-spectrum techniques add scaled and spreaded version of the message into the cover signal in the time or frequency domain, possibly with perceptual weighting to guaranty inaudibility [5].

While there has been quite some effort in the steganalysis of digital images, see good survey papers [6, 7, 8], steganalysis of digital audio is relatively unexplored, only two approaches are reported to date [9, 10]. The potential of distortion metrics in predicting the presence of steganographic content within one and two dimensional signals i.e., audio and image signals are shown in [9, 11]. In [9, 11], the basic idea rests on the evidence that the distortion measures computed between signals and their denoised versions have statistically distinguishable distributions for cover-signals and stego-signals. These statistically distinguishable features are then used to build a steganalyzer to discriminate cover-signals from stego-signals. The efficacy of this approach is shown within both steganographic and watermarking contexts.

In [10, 12], the basic approach works by finding predictable higher-order statistics of "natural" signals within a multi-scale decomposition, and then showing that embedded messages alter these statistics. Their statistical model begins by building a linear basis that captures certain statistical properties of audio signals. A low-dimensional statistical feature vector is extracted from this basis representation and used by a non-linear support vector machine for classification. The efficacy of this approach for speech is shown on LSB embedding and Hide4PGP [10].

Common to these approaches are the observation that data hiding changes the underlying statistical structure of the cover-signal and features that capture these statistical differences can be used for the classification of cover and stego-signals.

The traditional approach to speech modelling is the linear model which leads to the well known linear prediction model. This model has been used in speech coding, synthesis and recognition. However there is theoretical and experimental evidence for the existence of non-linear phenomena in speech signals that can not be accounted by the linear model [13, 14, 15].

In this paper we explore the potential of nonlinear features based on chaos theory and measure how chaotic characteristics of cover and stego-signals change after data hiding. The parameters of chaotic processes like Lyapunov exponents [16] and false nearest neighbours (FNN) [17] are used as features to build a universal steganalyzer. A universal steganalyzer is not targeted for a known data hiding method and works for different data hiding methods and supposed to work even for unknown methods. Experimental results with well known data hiding methods show the potential of the proposed method for speech signal steganalysis.

In Section 2 we describe the construction of the feature vector with Lyapunov exponents and false nearest neighbours. Steganalyzer design and results are given in Sections 3 and 4. Conclusions are drawn in Section 5.

2. CHAOTIC-TYPE FEATURES

In this section we present false nearest neighbours (FNN) and Lyapunov exponents as chaotic-type features and give the scatter plots of these features to show how discriminative they are over cover and stego speech signals. Actually, Lyapunov exponents and FNN parameters are described on non-linear chaotic dynamical systems. A discrete-time dynamical system can be modelled as $\boldsymbol{Y}(n+1)=\mathrm{F}[\boldsymbol{Y}(n)]$, where

$\boldsymbol{Y(n)}$ is state-vector of the system. Also, a speech production system can be viewed as non-linear dynamical system [18, 19]. A speech signal segment $x(n)$, n=1,….,N, can be considered as a one-dimensional projection of a vector function applied to the unknown multidimensional dynamic vector variable $\boldsymbol{Y(n)}$. According to the embedding theorem [20], the vector $\boldsymbol{X(n)}=[x(n),x(n+T_D),...,x(n+(D_E-1)T_D)]$ formed by samples of the original delayed by multiples of a constant time delay T_D defines a motion in a reconstructed D_E-dimensional space that has many common aspects with the original state- space of $\boldsymbol{Y(n)}$. Many quantities of the original dynamical system in the original state-space $\boldsymbol{Y(n)}$, like Lyapunov exponents, are conserved in the reconstructed space traced by $\boldsymbol{X(n)}$. The useful information about the original unknown dynamical system $\boldsymbol{Y(n)} \rightarrow \boldsymbol{Y(n+1)}$ can be uncovered if the embedding dimension D_E is large enough. The embedding theorem does not specify a method to determine the required parameters, time delay T_D and embedding dimension D_E. The embedding dimension for human speech signals is commonly selected between 3 and 7 for speech signals [21, 22].

2.1 False Nearest Neighbours

One of the foundations of the analysis of chaotic time series is Takens embedding theorem [20]. The theorem guarantees that if we perform a correct time-delay embedding of scalar signal, we may construct a D_E-dimensional space that inherits the dynamics of the original signal even we don't know the real dimension of the signal. The False Nearest Neighbours is the most accepted method for finding minimum embedding dimension which was proposed by Kennel et al. [17]. This method gives us a false nearest neighbours fraction for a given embedding dimension and time delay. The algorithm first determines all nearest neighbours for all the points in phase space. The nearest neighbour vector of $\boldsymbol{X(n)}$ can be given as

$$\boldsymbol{X}^{NN}(n)=[x^{NN}(n),\ x^{NN}(n+T_D),\ ...,\ x^{NN}(n+(D_E-1)T_D)] \qquad (2)$$

If the point is within the neighbourhood of the reference point the neighbourhood is a true neighbourhood. These two points will preserve the neighbourhood for all embedding dimensions. But if the points have become neighbours by the projection of small embedding dimension, the neighbourhood is a false neighbourhood. The next step is to control all of the neighbourhoods if the nearest neighbour is still a neighbour for the new higher embedding dimension. The algorithm labels the k. neighbourhood as true neighbourhood if

$$\frac{\left\| x(k+D_E T_D) - x^{NN}(k+D_E T_D) \right\|}{R_A} \qquad (3)$$

is greater than a number of order two, where R_A is the nominal radius of the phase space defined as the root-mean-square value of the data about its mean [16]. After labelling all neighbourhoods as true or false neighbours, the algorithm calculates the percentage of the false nearest neighbours to all neighbours as follows

$$FNN=\frac{\text{The Number of False Nearest Neighbours}}{\text{The Number of Nearest Neighbours}}\text{x}100 \qquad (4)$$

In Fig. 1 the difference between the cover and stego objects' false nearest neighbours fractions of 3 embedding dimensions (D_E=3, 4, 5) over 100 samples is shown.

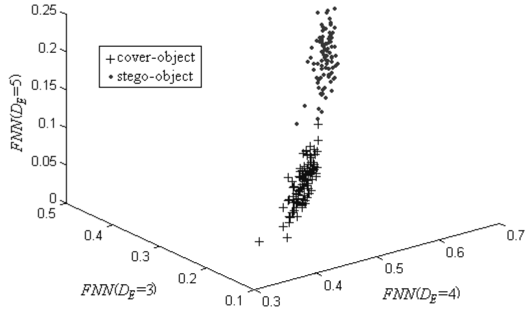

Fig. 1 False Nearest Neighbours Fractions of 3 embedding dimensions (D_E=3, 4, 5) of cover and stego-objects over 100 samples for DSSS data hiding method.

Fig. 1 clearly shows that FNN values of stego and cover-speech-signals are discriminant enough to be used as features in the steganalyzer design.

2.2 Lyapunov Exponents

Lyapunov exponents are used to quantify the predictability of chaotic systems. Chaotic systems are said to be 'sensitive to initial conditions'. 'Sensitive to initial conditions' means nearby trajectories, the set of points in phase space visited by a signal trajectory after transients are gone, will diverge at an exponential rate. Of course in reality, initial conditions can only be specified with some finite precision. Two close trajectories will move apart at an exponential rate which is described by the *Lyapunov exponent.* There are a number of exponents equal to the embedding dimension (D_E) of the phase space $\lambda_1,..,\lambda_{DE}$. *ith* Lyapunov exponent, λ_i can be calculated for *ith* principal axis [23].

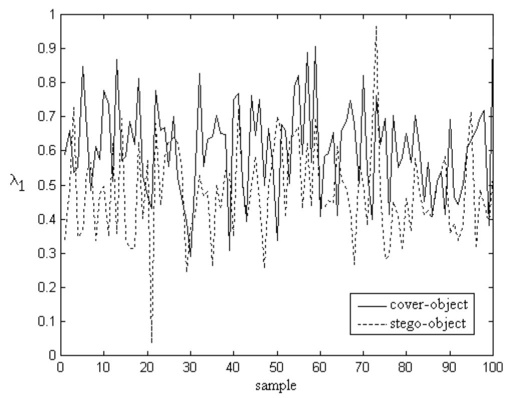

Fig. 2 Largest Lyapunov exponent (λ_1) of 100 samples of cover and stego-objects for Steganos data hiding method.

In Fig. 2 we can see the difference between the cover and stego objects' largest Lyapunov exponent (λ_1) over 100 samples where embedding dimension was selected as 7.
Figs 1 and 2 show that these chaotic-type features are discriminant enough to be used as features in the steganalyzer design.

3. STEGANALYZER DESIGN

The chaotic features in Section 2 are used as feature vectors for the steganalyzer design. We use a training set of cover speech signals and stego-speech signals. A linear regression classifier was designed using the statistics collected with the database of audio signals. Computed chaotic-type features are regressed to, respectively, -1 and 1, depending upon whether the speech signal did not or did contain a hidden message. In the regression model [24], we expressed each decision label $g_i \in [-1,1], i = 1,..,N$ as a linear combination of chaotic-type features, $g_i \in \beta_1 f_{1i} + \beta_2 f_{2i} + ... + \beta_q f_{qi}$, where $F = (f_{1i}, f_{2i},...f_{qi})$ is the vector of q chaotic-type features computed from i^{th} speech sample and $\beta_1, \beta_2,...\beta_q$ are the regression coefficients. The regression coefficients are predicted in the training phase, and then they are used in testing phase. In the testing phase, chaotic-type features are measured for the incoming speech signal, then the decision value is obtained by using the predicted regression coefficients. If the output exceeds the threshold 0, then the decision is that the speech contains message, otherwise the decision is that the speech does not contain any message.

3.1 The Feature Vector

The feature vector is obtained from Lyapunov exponents and false nearest neighbour rates of stego and cover-speech-signals. TISEAN is used for calculating false nearest neighbour rates of the signal [25]. The program explores false nearest neighbours by changing neighbourhood size for specified embedding dimension, D_E, [17]. The feature vector consists of 3 components: the fraction of false nearest neighbours, the average size of the neighbourhood, and the average of the root-mean-squared (RMS) size of the neighbourhood.

$$F_1=[\,FNN,\ mean(R_A),\ RMS(R_A)] \tag{7}$$

For a given embedding dimension, D_E, we calculate the Lyapunov Exponents of the given signal by using the TSToolbox [26]. The program generates D_E dimensional Lyapunov exponents in increasing order as follows

$$F_2=[\,\lambda_1, \lambda_2, ..., \lambda_{d_E}\,] \tag{8}$$

The embedding dimension of the stego and cover-speech signals is selected between 3 and 7 for FNN and 7 for Lyapunov exponents. The complete feature vector is consisted of 22 elements as follows

Table 1. The feature vector with 22 elements. F_1 is the false nearest neighbours feature vector as given in (7).

k	1, 2, 3	4, 5, 6	7, 8, 9	10, 11, 12	13, 14, 15
f_k	$F_1(D_E=3)$	$F_1(D_E=4)$	$F_1(D_E=5)$	$F_1(D_E=6)$	$F_1(D_E=7)$

k	16	17	18	19	20	21	22
f_k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7

4. EXPERIMENTAL RESULTS

We have performed steganalysis experiments over five different algorithms, of which three are watermarking techniques and two are steganographic techniques. The watermarking techniques are direct-sequence spread spectrum (DSSS) and frequency hopping with spread spectrum (FHSS) and echo hiding (ECHO) [5]. The steganographic methods are Steganos [2] and MP3Stego [4]. The rationale of using these tools was their popularity, freely availability and wide usage.
The speech segments have durations of three to four seconds, are sampled at 16 kHz and recorded in acoustically shielded medium. The procedure consists of embedding messages to all available cover signals, randomly selecting half of the set of the stego and cover signals for training, leaving the other 50% for testing phase. The embedded message size was 10% of the audio size, which is usually the maximum allowed capacity for LSB embedding. The embedding algorithms were tested on 100 speech utterances. The detection results are given in Table2. The results show the potential of the proposed chaotic-type features.

Table 2. The performance of the classifiers

Data Hiding Methods	False Positive	False Negative
DSSS	0/50	0/50
FHSS	0/50	0/50
ECHO	2/50	9/50
STEGANOS	4/50	5/50
MP3STEGO	6/50	13/50

5. CONCLUSIONS

In this paper, chaotic-type features are investigated for universal speech steganalysis. These features were used in the design of linear regression classifier. The steganalyzer was tested on watermarking and steganographic methods. Experimental results showed the potential of chaotic-type features for their discriminatory power to be used in steganalysis. We believe that the integration of chaotic-type features with linear ones may capture both linear and non-linear aspects of speech signals leading to robust and efficient hybrid steganalyzers.

Acknowledgement: The authors would like to thank Hamza Ozer of TUBITAK UEKAE Speech Group for providing the database of digital audio and speech tracks used in the experiments.

REFERENCES

[1] A. Brown, S-Tools version 4.0, Copyright C. 1996, http://members.tripod.com/steganography/stego/s-tools4.html.

[2] Steganos, www.steganos.com.

[3] I. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia", *IEEE Trans. on Image Process.,* vol. 6, no: 12, pp. 1673-1687, December 1997.

[4] www.petitcolas.net/fabien/steganography/mp3stego

[5] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", *IBM Systems Journal,* vol. 35, no: 3&4, pp. 313-336, 1996.

[6] N. F. Johnson, S. Katzenbeisser, "A survey of steganographic techniques", in S. Katzenbeisser and F. Petitcolas (Eds.): *Information Hiding,* pp. 43-78, Artech House, Norwood, MA, 2000.

[7] J. Fridrich, M. Goljan "Practical Steganalysis of Digital Images - State of the Art", *Proc. SPIE Photonics West*, Vol. 4675, 2002, 1--13

[8] R. Chandramouli, M. Kharrazi, N. D. Memon, "Image Steganography and Steganalysis: Concepts and Practice", IWDW 2003: 35-49.

[9] H. Ozer, I. Avcibas, B. Sankur, N. Memon, "Steganalysis of Audio Based on Audio Quality Metrics", *Security and Watermarking of Multimedia Contents V*, Proceedings of the SPIE, Vol. 5020, pp. 55-66, 2003.

[10] M.K. Johnson, S. Lyu and H. Farid , "Steganalysis of Recorded Speech",SPIE Symposium on Electronic Imaging, San Jose, CA, 2005.

[11] I. Avcibas, N. Memon, B. Sankur, "Steganalysis using image quality metrics", *IEEE Trans. on Image Process.*, vol. 12, pp. 221-229, Feb. 2003.

[12] S. Lyu and H. Farid, "Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines", 5th International Workshop on Information Hiding, Noordwijkerhout, The Netherlands, 2002.

[13] P. Maragos, a. Potamianos, "Fractal Dimensions of Speech Sound: Computation and Application to Automatic Speech Recognition", J. Acust. Soc. Amer., 105(3), pp. 1925-1932, March, 1999.

[14] V. Pitsikalis, P. Maragos*, "*Speech analysis and feature extraction using chaotic models", Acoustics, Speech, and Signal Processing, 2002. Proc. IEEE, ICASSP '02 vol.1, pp. 533-536, 2002.

[15] G. Kubin, "Synthesis and Coding of Continuous Speech with the Non-linear Oscillator Model", Proc. IEEE, ICASP'96, pp. 267-270, 1996.

[16] H. D. I. Abarbanel, T. W. Frison, L. S. Tsimring, "Obtaining Order in a World of Chaos", *IEEE Signal Processing Magazine*, May 1998.

[17] M. B. Kennel, R. Brown, H. D. I. Abarbanel, "Determining Minimum Embedding Dimension Using a Geometrical Construction", *Physical Review A* 45:3403-3411, 1992.

[18] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics,* Springer-Verlag, Applied Mathematical Sciences, vol.68, 1993.

[19] X. Liu, R. J. Povinelli, M. T. Johnson, "Detecting Determinisim in Speech Phonemes", *IEEE Digital Signal Processing Workshop*, 2002.

[20] F. Takens, "Dynamical Systems and Turbulance", *Lecture Notes in Mathematics 898,* Warwick 1980. Berlin: Springer-Verlag, 1981.

[21] M. Banbrook, S. McLaughlin, "Is Speech Chaotic?: Invariant Geometrical Measures for Speech Data", *IEE Colloquium on "Exploiting Chaos in Signal Processing",* Digest No: 1994/193, pp8/1-8/10, June 1994.

[22] S. McLaughlin, M. Banbrook, I. Mann, "Speech Characterization and synthesis by nonlinear methods", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no.1, pp. 1-17, January 1999.

[23] G. Elert, "The Chaos Hypertext Book", www.hypertextbook.com/chaos

[24] A. C. Rencher, *Methods of Multivariate Analysis*, New York, John Wiley, 1995, ch. 6,10.

[25] R. Hegger, H. Kantz, T. Schreiber, "Practical Implementation of Nonlinear Time Series Methods: The TISEAN Package", *Chaos*, vol 9:413-435, 1999.

[26] C. Merkwirth, U. Parlitz, I. Wedekind, W. Lauterborn, "TSTOOL User Manual Version 1.11", http://www.physik3.gwdg.de/tstool/HTML/index.html.